Essays on production risk management in smallholder agriculture: An economic investigation with observational microdata

A Thesis submitted by

Sumedha Shukla

PhD21402

Under the supervision of

Prof. Gaurav Arora

For the degree of

Doctor of Philosophy

in the

Econometrics Lab, Department of Social Science and Humanities

Indraprastha Institute of Information Technology Delhi

July, 2026

Essays on production risk management in smallholder agriculture: An economic investigation with observational microdata

Submitted by

Sumedha Shukla

PhD21402

For the degree of

Doctor of Philosophy

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

July, 2026

Essays on production risk management in smallholder agriculture: An economic investigation with observational microdata

A Thesis submitted by

Sumedha Shukla

PhD21402

Under the supervision of

Prof. Gaurav Arora

For the degree of

Doctor of Philosophy

in the

Econometrics Lab, Department of Social Science and Humanities

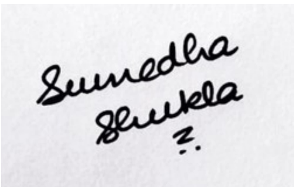
Indraprastha Institute of Information Technology Delhi

July, 2026

Declaration

This is to certify that the thesis titled "***Essays on production risk management in smallholder agriculture: An economic investigation with observational microdata***" has been authored by me. It presents the research conducted by me under the supervision of ***Prof. Gaurav Arora***.

To the best of my knowledge, it is an original work, both in terms of research content and narrative, and has not been submitted elsewhere, in part or in full, for a degree. Further, due credit has been attributed to the relevant state-of-the-art and collaborations (if any) with appropriate citations and acknowledgments, in line with established norms and practices.



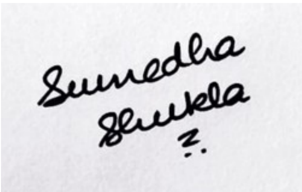
Sumedha Shukla (PhD21402)

Econometrics Lab
Department of Social Science and Humanities
Indraprastha Institute of Information Technology Delhi
New Delhi 110020
July, 2026

Declaration Regarding Use of AI

I acknowledge that I am fully responsible for the entire content of my thesis, including any sections assisted by online tools, including Artificial Intelligence-based tools.

I accept full accountability for any violations of ethical standards in publications arising from the use of such tools.

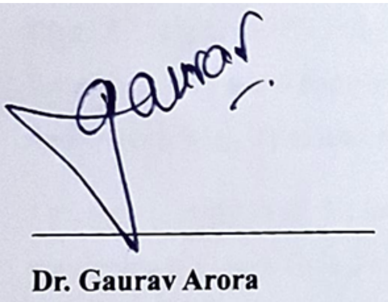


Sumedha Shukla (PhD21402)

Certificate

This is to certify that the thesis titled "**Essays on production risk management in smallholder agriculture: An economic investigation with observational microdata**" being submitted by **Sumedha Shukla** to the Indraprastha Institute of Information Technology Delhi, for the award of the degree of Doctor of Philosophy, is an original research work carried out by her under my supervision.

The results contained in this thesis, to the best of my knowledge, have not been submitted in part or in full to any other university or institute for the award of any degree/diploma.



Associate Professor (Economics)
Department of Social Science and Humanities
Indraprastha Institute of Information Technology Delhi
New Delhi 110020
July, 2026

Acknowledgements

This thesis would not have been possible without the unwavering support of my **family** who have been endlessly patient, motivating, loving and encouraging throughout this journey. I also owe my **friends** an immense amount of gratitude for their belief in my abilities and their willingness to support me at all times.

I am deeply grateful to my advisor **Prof. Gaurav Arora**, for his kindness and steady support which have carried me through the PhD program. His capable and patient mentorship, insightful questions, emphasis on rigor and meticulous comments have shaped me into a better researcher.

I am sincerely thankful to **Prof. J.V. Meenakshi** for always entertaining my doubts and encouraging me to ask questions. Her invaluable guidance has significantly improved the quality of my work.

I am incredibly grateful for the detailed and constructive comments I received from members of my examination committee: **Prof. Bharat Ramaswami**, **Prof. Sheetal Sekhri** and **Prof. Ziran Li**, which significantly improved the quality of my thesis. I also thank the members of my annual review and comprehensive examination committees for generously investing their time in reviewing my research and offering constructive feedback.

I am very grateful to all the faculty members whose teaching has been instrumental in developing my research skills. Some of my most challenging courses with **Prof. Rakesh Chaturvedi** have also helped me grasp the fundamentals much better.

The origins of this journey lie in the research mentorship provided by faculty at the TERI School of Advanced Studies, my alma mater, and especially **Prof. Soumendu Sarkar** whose encouragement inspired me to pursue the PhD I had long aspired to undertake.

I would also like to express gratitude to the **academic administrative staff** and the **IT support staff** at IIIT-Delhi for all their support in this long journey.

Abstract

This thesis presents applied microeconomic analyses pertaining to production risk management on small farms, with a focus on distributional aspects of farmer decision-making and farm productivity. In the **first paper**, I study the heterogeneous impact of agricultural credit on crop yield quantiles for households operating millet-based cropping systems. Equilibrium credit rationing under asymmetric information is the economic channel underlying heterogeneous credit-yield linkage. Therefore, I exploit within-sample, quantile-based variation in credit-yield linkage to detect credit rationing incidence among sample households. Specifically, I propose a new metric for household-specific ex-ante odds of credit rationing, beyond identifying whether the average-household was rationed. I interpret these odds within a state-contingent technology framework, which reveals how credit rationing influences farm-level adaptation under uncertainty. In the **second paper**, I employ a beta regression model in conjunction with a non-linear instrumental variable approach, i.e., Beta-IV framework to estimate input-conditioned crop yield densities from plot-level observational data for wheat, rice, cotton and pigeonpea. The key inputs considered are nitrogen, phosphorus and irrigation levels. My work reveals novel evidence on how crop inputs influence production risk across major crops. I use these density estimates to evaluate the role of mixed-cropping to manage production risk among risk-averse, smallholder farmers in semi-arid India (Di Falco and Chavas 2006). My **third paper** is a data contribution to algorithmically reconcile identifier inconsistencies in land quality information in the Village Dynamics Studies in South Asia (VDSA)-plot-level panel dataset comprising 1,689 farm households across multiple survey waves during 1975-2014. My algorithm uses VDSA's plot nomenclature and phonetic/vernacular similarities to reduce data loss extent from 52% to <2%. In the **fourth paper**, I examine farmers' subjective soil quality perceptions and assess how these perceptions diverge from data-based evidence. I construct and model a plot-level measure of soil quality misperception as the gap between survey-based soil quality metrics and corresponding data-based measures from high-resolution gridded maps. Further, I examine previously-unexplored two-way causal relations between soil misperceptions and input application decisions by using appropriate IVs for each directional relationship. Overall, my essays advance academic and policy research on smallholder agriculture in the global South.

Keywords: credit, crop yields, production risk management, matching, quantile regressions, beta regressions, input risk properties, instrumental variables, soil perceptions, microdata
JEL Codes: C14, C21, C36, Q12, Q15, Y1

सारांश

यह शोध-प्रबंध लघु कृषकों के बीच उत्पादन जोखिम प्रबंधन से संबंधित सूक्ष्म आर्थिक विश्लेषण प्रस्तुत करता है, जिसमें किसानों की निर्णय-प्रक्रिया तथा कृषि उत्पादकता के वितरणात्मक पहलुओं पर विशेष ध्यान दिया गया है। **पहले शोधपत्र में**, मैं बाजरा-आधारित कृषि प्रणालियों में कार्यरत परिवारों की फसल उपज पर कृषि ऋण के असमरूप (क्वांटाइल-विशिष्ट) प्रभाव का अध्ययन करती हूँ। असमान जानकारी (asymmetric information) की उपस्थिति में संतुलन ऋण-सीमा निर्धारण (equilibrium credit rationing) इस असमरूप ऋण-उपज संबंध के पीछे की एक आर्थिक कड़ी है। अतः मैं ऋण-उपज संबंध में क्वांटाइल-विशिष्ट भिन्नता का उपयोग करते हुए सैंपल परिवारों के लिए ऋणदाता-निर्धारित ऋण-सीमा की बंधनकारिता का आकलन करती हूँ। विशेष रूप से, मैं यह अनुमान लगाती हूँ कि किसी कृषक के लिए ऋणदाता-निर्धारित सीमा के बंधनकारी होने की क्या प्रायिकता है। यह प्रश्न मौजूदा साहित्य द्वारा पूछे गए इस प्रश्न से भिन्न है कि औसत परिवार के लिए ऋणदाता-निर्धारित सीमा बंधनकारी है या नहीं। मैं इस पूर्व-प्रायिकता (ex-ante probability) को ऋण-सीमा बंधनकारिता के एक मापदंड के रूप में प्रस्तावित करती हूँ तथा इसकी व्याख्या state-contingent technology ढाँचे के भीतर करती हूँ। इसके माध्यम से मैं यह आकलन करती हूँ कि बंधनकारी ऋण-सीमा अनिश्चितता की स्थिति में किसानों की अनुकूलन-क्षमता (adaptive capacity) को किस प्रकार प्रभावित करती है। **दूसरे शोधपत्र में**, मैं भूखंड-स्तरीय अवलोकनात्मक आंकड़ों (observational data) के आधार पर गेहूं, चावल, कपास और अरहर की इनपुट-सापेक्ष फसल उपज घनत्वों (input-conditioned yield densities) का अनुमान लगाने के लिए बीटा रिग्रेशन मॉडल (beta regression model) तथा नॉन-लीनियर इंस्ट्रुमेंटल वेरिएबल (non-linear instrumental variable) अप्रोच, अर्थात् बीटा-IV फ्रेमवर्क, का प्रयोग करती हूँ। इस पद्धति से भारत के अर्ध-शुष्क क्षेत्रों में प्रमुख फसलों के मुख्य इनपुटों (नाइट्रोजन, फास्फोरस और सिंचाई) के उपज-जोखिम गुणों के संबंध में नए अनुभवजन्य आकलन प्राप्त होते हैं। इन अनुमानित घनत्वों के आधार पर मैं अर्ध-शुष्क क्षेत्रों में जोखिम-विमुख (risk-averse) लघु कृषकों के बीच उत्पादन जोखिम प्रबंधन में मिश्रित-खेती (mixed-cropping) की भूमिका का मूल्यांकन करती हूँ (Di Falco and Chavas 2006)। **तीसरे शोधपत्र में**, मैं Village Dynamics Studies in South Asia (VDSA) के भूखंड-स्तरीय पैनल डेटासेट (1975–2014 में एकाधिक सर्वेक्षण तरंगों में शामिल 1,689 कृषक परिवार) में पहचान-सूचक (identifier) असंगतियों को दूर करने के लिए एक एल्गोरिथ्मिक समाधान प्रस्तुत करती हूँ। यह एल्गोरिथ्म VDSA की नामकरण पद्धति तथा ध्वन्यात्मक/स्थानीय (vernacular) भाषा समानताओं का उपयोग करके मृदा-गुणवत्ता (soil quality) संबंधी डेटा हानि को 52% से घटाकर <2% कर देता है। **चौथे शोधपत्र में**, मैं किसानों की मिट्टी की गुणवत्ता संबंधी व्यक्तिपरक धारणाओं का अध्ययन करती हूँ और यह आकलन करती हूँ कि ये धारणाएं डेटा-आधारित प्रमाणों से किस प्रकार भिन्न हैं। जियोकोडेड सर्वेक्षणों से प्राप्त किसानों की मिट्टी-गुणवत्ता संबंधी धारणाओं तथा उच्च-विभेदन ग्रिड-आधारित मानचित्रों (high-resolution gridded maps) से प्राप्त मिट्टी मापों के बीच के अंतर के आधार पर मैं एक भूखंड-स्तरीय मापदंड विकसित करती हूँ। इसके बाद, मैं इस अंतर और किसानों के इनपुट-प्रयोग संबंधी निर्णयों के बीच के कारणात्मक संबंध (causal linkage) का विश्लेषण करती हूँ। विशेष रूप से, मैं इस संबंध की दोनों दिशाओं में जांच करती हूँ — अर्थात् किस प्रकार धारणाएं इनपुट-प्रयोग संबंधी निर्णयों को आकार देती हैं, तथा किस प्रकार इनपुट-प्रयोग संबंधी निर्णय धारणाओं का आधार बन सकते हैं। कुल मिलाकर, मेरे निबंध वैश्विक दक्षिण (global south) में लघु कृषकों की कृषि से संबंधित शैक्षणिक और नीतिगत शोध को आगे बढ़ाते हैं।

Keywords: ऋण, फसल उपज, उपज-जोखिम प्रबंधन, मैचिंग, क्वांटाइल रिग्रेशन, बीटा रिग्रेशन, इनपुट के उपज-जोखिम गुण, इंस्ट्रुमेंटल वेरिएबल, मिट्टी गुणवत्ता संबंधी धारणाएं, भूखंड-स्तरीय डेटा

JEL Codes: C14, C21, C36, Q12, Q15, Y1

Contents

Table of Contents

List of Figures

List of Tables

List of Papers Based on Thesis

Working Papers (Presented at A* Equivalent Conferences with DOI):**

**: Full list of conferences is included in the CV.

- Shukla, S., Arora, G., "Heterogeneous crop yield impacts of agricultural credit: A quantile regression approach embedded in the state-contingent framework".
  - **DOI:** https://doi.org/10.22004/ag.econ.314027
- Shukla, S., Arora, G., Agarwal S.K., "Estimation of crop yield density conditional on input choices for smallholder farms using a BetaIV framework."
  - **DOI:** https://doi.org/10.22004/ag.econ.322580
- Shukla, S., Arora, G., Agarwal S.K., "Role of mixed-cropping in production risk management on small farms (An application of the BetaIV framework for yield density estimation)"
  - **DOI:** https://doi.org/10.22004/AG.ECON.361060
- Shukla, S., Arora, G. "Soil quality perceptions: Characterizing bias and linkage with farming decisions for rice growers in India."
  - **DOI:** https://ageconsearch.umn.edu/record/336014?v=pdf
- Shukla, S., Arora, G. "A rank similarity test for quantile treatment effects in conjunction with propensity score matching: An application to heterogeneous crop yield impacts of agricultural credit."
  - **DOI:** http://d.repec.org/n?u=RePEc:ags:aaea22:322569&r=

Chapter 1: Introduction

Enhancing crop yields under increasing land scarcity represents a pressing global challenge, compounded by climate-induced increases in yield variability that heighten food and nutrition insecurity risks (Ray et al. 2015). The challenge is particularly acute in arid and semi-arid regions (accounting for >33% of global land area (Wickens 1998)), where low yield potential and heavy reliance on rainfed agriculture intensify livelihood threats from erratic weather. Indian agriculture is vulnerable with over 85% farmers operating marginal (<1 hectare) and small (<2 hectare) landholdings and rainfed farming accounting for over 51% of net sown area and 40% of total food production (Ministry of Agriculture and Farmer's Welfare (MoAFW) 2022). Therefore, designing effective tools and policies for farm risk management is essential. This thesis aims to assess the effectiveness of different farm-level crop yield management strategies in India. My focus is on *distributional* assessments using farm-level data, in a manner that moves beyond conventional average production outcomes (Bedoya et al. 2018).

In the first paper, I study heterogeneous impact of agricultural credit on crop yields. Agricultural credit can relax working-capital constraints for lesser-endowed farmers. Yet its access remains skewed reflecting underlying socio-economic disparities. Equilibrium credit rationing in the presence of asymmetric information provides an economic rationale for disproportionate credit access and favourable production outcomes for better-endowed farmers. This study employs quantile regression models in conjunction with propensity score matching to estimate the heterogeneous impact of agricultural credit on crop yield quantiles for a representative sample of farmers operating millet-based cropping systems in semi-arid tropics of India. I then exploit the within-sample, quantile-based variation in the credit-yield linkage to augment sample-wide detection of binding credit constraints. The proposed approach moves beyond rationing detection based on average *realized* outcomes to derive *ex-ante credit rationing odds* (i.e., at the time of borrowing a loan).

I then interpret credit rationing odds within a state-contingent technology framework revealing previously unexplored structural implications of rationing for farmers' adaptive capacity under uncertainty. I show theoretically that rationing-induced credit constraints lower farmer welfare and lead to *riskier* net revenue bundles in the second-order stochastic dominance sense. A decomposition of input adjustments driving this shift draws on a stochastic version of Shepherd's Lemma (Chambers and Quiggin 2000; 2001) and reveals that a binding credit constraint *may increase or decrease input use depending on the risk-influencing properties of a specific input*. I further propose formal statistical hypotheses to empirically uncover the rationing mechanism through which equally productive farmers may face differential credit terms on the basis of group-based affiliations that influence lender perception about repayment ability.

I find that credit attainment is skewed toward upper-caste, asset-rich, land-owning households with higher-quality soils. Borrowers experience higher yields relative to a no-credit scenario, but gains are concentrated in the upper yield quantiles, indicating higher yield dispersion due to credit attainment. The average net revenue gained per rupee of credit exceeds borrowing costs by slightly less than a rupee (per acre), consistent with weakly binding credit constraints for the average farmer. However, given the skewed (but positive) credit impact across yield quantiles, median impact of credit is likely lower, suggesting binding constraints for only a *subset* of borrowers. Returns from crops inter-cropped with millets drive *ex-ante* rationing among ~48% of borrowers. Rationed households are found to systematically employ irrigation and superior seed varieties less frequently, but increase reliance on low-cost labour and fertilizer. Evidence suggests wealth-based disparities in credit allocation and rationing which are only partially mitigated by observed productivity indicators. Overall, these findings raise important questions for agricultural credit targeting in semi-arid regions.

In the second paper I estimate crop yield densities for wheat, rice, cotton and pigeonpea conditional on key inputs (nitrogen, phosphorus and irrigation) for smallholder farms in the semi-arid regions of India. In contrast to existing studies that rely largely on agronomic experiments or coarse spatial data from resource-rich settings, my study utilizes plot-level, observational data which has the potential to provide insights into farmers' decision-making processes in semi-arid regions - a distinctive, policy-relevant empirical setting marked by pervasive abiotic stress and limited access to institutionalized risk management strategies (e.g., crop insurance) where inputs play a critical role in production risk management.

My empirical framework (based on Arora et al. 2021) utilizes a reparametrized version of Beta regressions with two shape parameters: a location parameter and a dispersion parameter (Nelson and Preckel 1989; Smithson and Verkuilen 2006) which are modelled as a function of input application, land quality, proxy variables for farmer management ability, cropping history and weather outcomes. A key challenge when using observational data is the endogeneity between observed input choices and yields whereby higher yields (both potential and realized) can facilitate better input application and changes in input application can also influence both potential and realized yields. Yield potential accounts for mean yield as well as the likelihood of lower/higher yields, i.e., the higher order moments of the yield distribution. Hence, both the location and dispersion model suffer from endogeneity. To address endogeneity, I employ a non-linear instrumental variable framework with Tobit regressions to handle corner solutions in input application choices. Instruments include input prices, output prices, indirect input costs and plot-level cropping history. Following Terza et al. (2008), I utilize the two-stage residual inclusion (2SRI) method for endogeneity resolution over the standard two-stage predictor substitution method to account for non-linearity in the estimation framework.

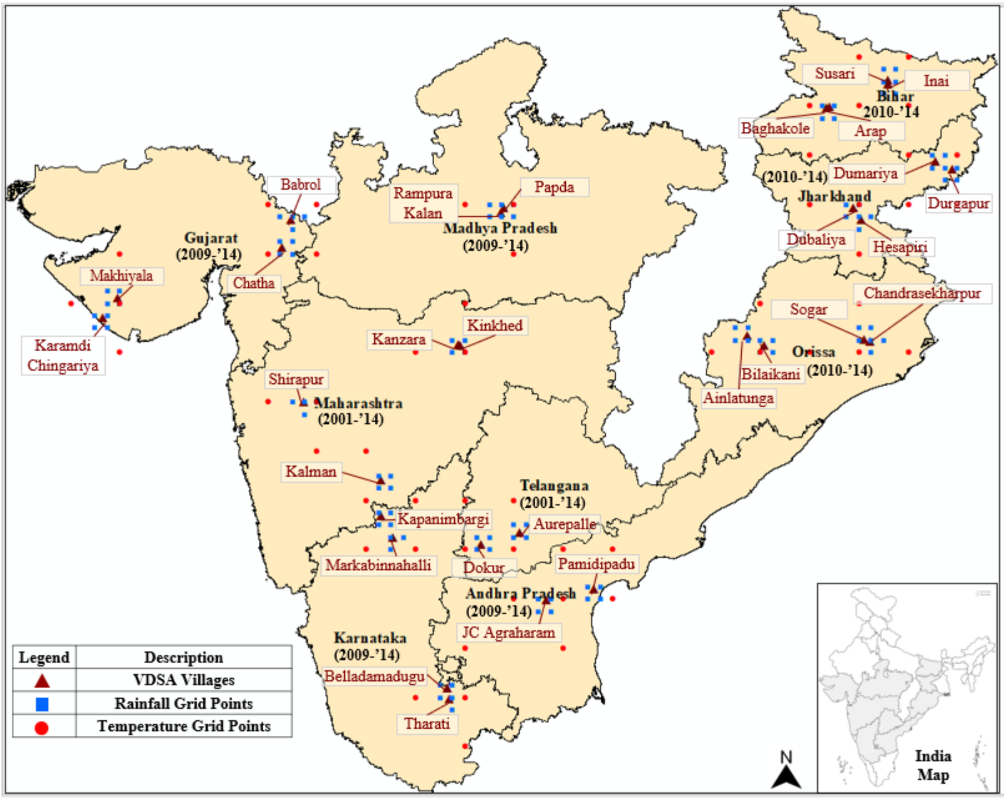
The density estimates provide novel evidence on risk properties of crop inputs for Indian agriculture and offer a realistic stepping stone for agricultural policy design. As an illustration, in the final section of this paper I present a case-study where I examine the welfare effects of cotton-pigeonpea mixed-cropping[1] (relative to monocropping) by incorporating the share of land allocated to mixed cropping as a farmer's decision to manage production risk. While mixed cropping has been shown to enhance farm incomes, improve dietary diversity and lower production costs (DiFalco et al. 2007; Donfouet et al. 2017; Tesfaye and Tirivayi 2020), the production risk effects mediating these outcomes remain ambiguous. Existing empirical work on cropping diversity and production risk has focused primarily on crop-specific (intra-species) varietal diversification. In contrast, evidence on the risk-reducing potential of inter-species mixed cropping systems (e.g., cereal-legume, rice-cereal, rice-fodder), which are widely practiced in semi-arid regions (Waha et al. 2020), remains limited. In this case-study, welfare implications are evaluated using certainty equivalent defined as expected income minus the cost of risk exposure where risk exposure is quantified using the variance and skewness of yield distributions, which is an adaptation of the Di Falco and Chavas (2006; 2009) model to my inter-species multiple-cropping context. I find that mixed cropping reduces the variability of returns and increases skewness, leading to an overall reduction in downside risk exposure for the representative farmer.

The first two papers in this thesis rely on the VDSA dataset provided by the International Crop Research Institute for the Semi-Arid Tropics (ICRI(SAT)). The semi-arid regions, home to a majority of agricultural poor (Abel at al. 2021) are vulnerable to increasing uncertainty in agricultural incomes due to climate change. Developing effective agricultural risk management strategies for these regions requires an in-depth understanding of the farmers'

[1] Mixed cropping refers to simultaneous cultivation of multiple crops on a single plot.

decision-making processes including the underlying production environment and economic constraints. A comprehensive snapshot of farming decisions in this environment relies critically on longitudinal microdata, i.e., data at the plot or farm household level collected over time (Mullen 2016). However, the costly nature of such data collection has proven prohibitive for much of the developing world (You et al. 2009; Ravisankar, et al. 2021).

In this backdrop, the VDSA program provides a unique, plot-level, panel dataset on farming operations for 1,689 farming households located across thirty Indian SAT villages, spanning multiple waves in the period 1975-2014 (ICRISAT 2014). Figure 1.1 displays the VDSA village locations and survey-years for each state for 2001-2014[2]. The data are organized across twelve modules, capturing farming households' endowment status, landholding information, non-farm activities, consumption, transactions, and cropping patterns. Widely recognized as a global public good, the data has been employed in over 300 studies and cited more than 41,000 times between January 1975 and June 2024 (see Chapter 4 for details). Beyond individual applications, the VDSA dataset has shaped stylized facts in the development economics literature and spurred innovations in panel data methodologies that extend well beyond the SAT regions (Mullen 2016).



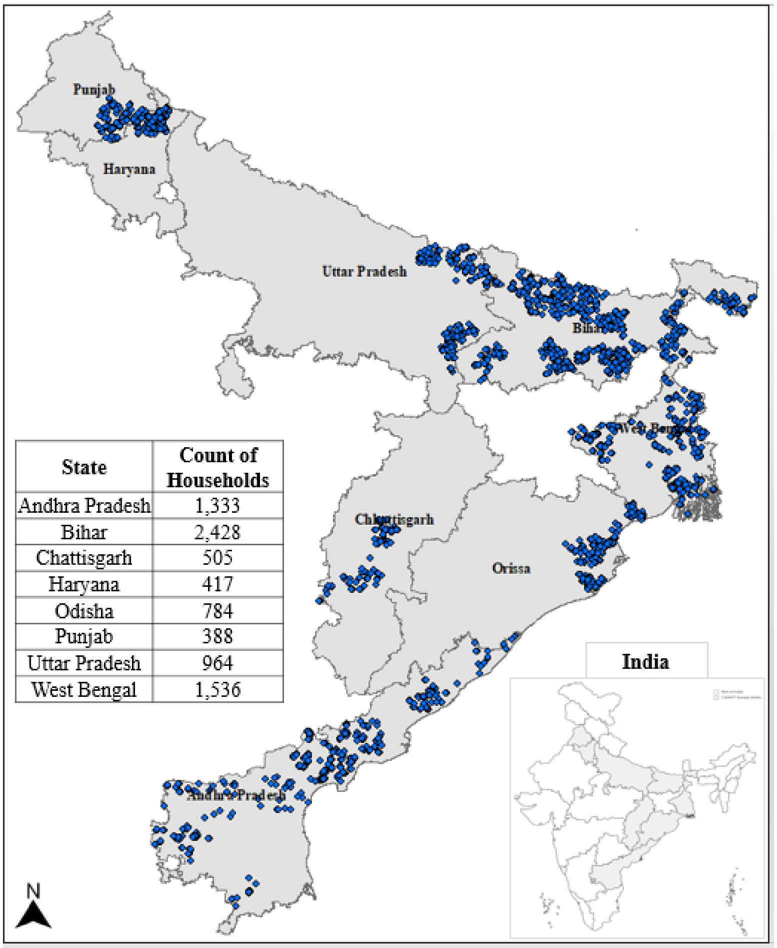
[2] Figure 1.1 also displays the nearest weather grid points for each village. Gridded data on temperature and rainfall is sourced from Indian Meteorological Department. Details are provided in Supplementary Information of Chapter 2.

Figure 1.1: Spatiotemporal span of the VDSA Microdata (2001-14)

One challenge in using the VDSA dataset to understand farming decisions stems from information losses when merging the cropping module which contains plot-level, season-wise information on crop choice, input use, and output with the landholding module, which records critical biophysical constraints i.e., plot-level soil depth, slope, texture, colour, and quality, due to inconsistencies in plot identifiers. Attempting to merge the modules at the plot level results in the loss of 52% of observations across all major crops, undermining attempts to analyse production choices constrained by agroecology and potentially inducing non-random sample selection. The third paper addresses this issue by designing and implementing an algorithmic merging procedure, supplemented by a manual merging criterion based on phonetic similarities and vernacular identifiers, reducing data losses by 50 percentage points. This entails a general contribution to the extensive economic literature using the VDSA microdata to study farm production decisions.

The fourth paper in this thesis examines the determinants of farmers' perceptions about soil quality. Maintaining soil quality amid increasing human pressure on lands is a major challenge. Policy efforts to improve soil fertility tend to inform farm-level decisions. But *adoption* of policy mandates depends upon farmer *perceptions* about soil quality which drive their *subjective* cost benefit calculations. Perceptions can be inconsistent with data-based evidence because expectations about unobserved soil quality are typically formed based on observable (but) imperfect proxies like crop yields. I characterize soil quality perceptions of Indian rice-growers addressing potential deviations between subjective soil quality perceptions (recorded in primary surveys) and corresponding data-based evidence (recorded in digital soil maps). Further, I evaluate potential drivers of soil quality *mis*perceptions, including growers' socio-economic status and farming history. Finally, I explore spatial patterns in soil quality misperceptions and the linkages with farm-level decisions.

Divergences between subjective soil perceptions and data-based measurements have been a focal point of enquiry for decades (Liebig and Doran 1999; Corbeels et. al. 2000; Andrews et al. 2003; Desbiez et. al. 2004; Okoba and Graff 2005; Moges and Holden 2007; Gruver and Weil 2007; Asafu-Adjaye 2008; Berazneva et al. 2018; Delgado and Stoorvogel 2022). However, the geographical scope of this literature remains limited (largely to Sub-Saharan Africa) and linkages with economic decision-making are seldom examined. I add to this literature by presenting evidence from diverse agroecological contexts across the Indo-Gangetic Plains (see Figure 1.2 for CIMMYT plot locations) and by analysing spatial patterns in soil quality *mis*perceptions and linkages with input application decisions.



State	Count of Households
Andhra Pradesh	1,333
Bihar	2,428
Chattisgarh	505
Haryana	417
Odisha	784
Punjab	388
Uttar Pradesh	964
West Bengal	1,536

Figure 1.2: Plot locations in CIMMYT data.

The study employs geo-referenced, plot-level data on soil quality perceptions recorded in an ordered manner along two dimensions: soil fertility (low, medium, high) and soil texture (light, medium, heavy) for 8,327 rice-growing farmers during the 2018 *kharif* (rainy) season. Data-based soil quality measurements were obtained by matching the CIMMYT plot locations to spatially-delineated (1km x 1km resolution) global digital soil maps provided ISRIC. Soil fertility is measured by organic carbon stock (tonnes/hectare) and nitrogen content (centigram/kilogram) while soil texture is measured by the clay content (gram/kilogram). A snapshot of data construction is provided in Figure 1.3.

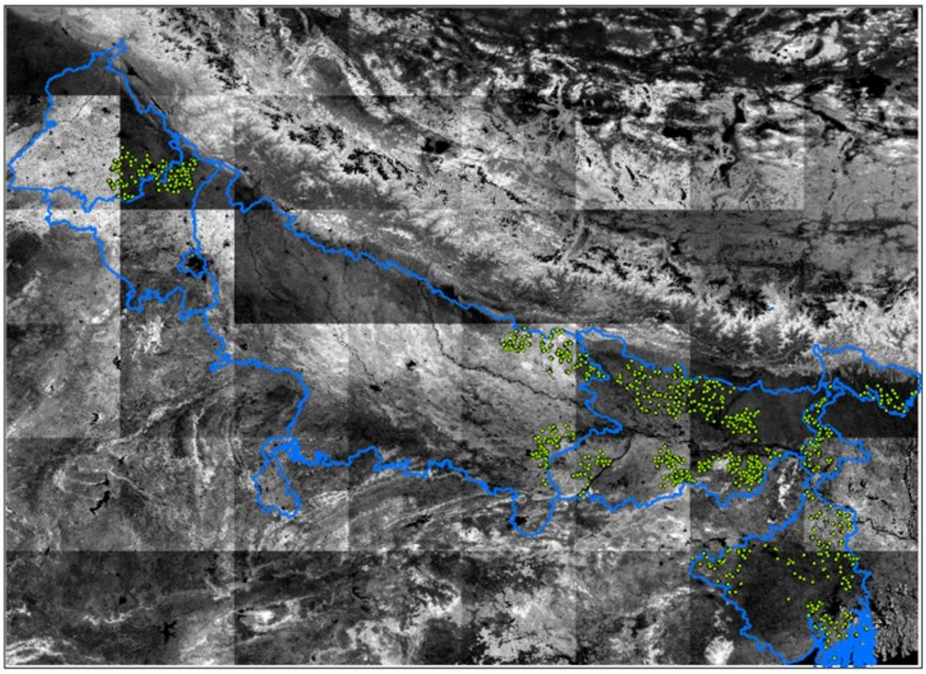


Figure 1.3: CIMMYT plot locations in the Indo-Gangetic plains superimposed on gridded soil organic carbon stock data from ISRIC.

I find that a majority of farmers in the sample consistently perceived soil texture but inconsistently perceived soil fertility on their farms relative to data-based measurements. Higher yields, irrigation access, and higher socio-economic status (measured by grower's wealth, education, land ownership, and caste) were associated with better soil quality perceptions. Both *consistent* and *biased* perceptions of soil quality were found to spatially clustered (Moran's-I value = 0.43 (p<0.001)) and exhibited significant spatial heterogeneity. Lastly, a Chi-squared test of independence revealed that over-perceptions of soil quality were associated with lower nutrient supplementation and lower incidence of crop residue burning, whereas under-perceptions of soil quality were less likely to affect (and be affected by) farm-level decisions. These findings suggest that land management policy can be improved by incorporating farmers' *subjective* perceptions about their soil quality.

The next chapters will provide details for each paper and the final chapter concludes.

1.1 References

Abel, Christin, Stéphanie Horion, Torbern Tagesson, Wanda De Keersmaecker, Alistair W. R. Seddon, Abdulhakim M. Abdi, and Rasmus Fensholt. 2020. "The Human–environment Nexus and Vegetation–rainfall Sensitivity in Tropical Drylands." *Nature Sustainability* 4 (1): 25–32. https://doi.org/10.1038/s41893-020-00597-z.

Andrews, S.S., C.B. Flora, J.P. Mitchell, and D.L. Karlen. 2003. "Growers' Perceptions and Acceptance of Soil Quality Indices." *Geoderma* 114 (3–4): 187–213. https://doi.org/10.1016/s0016-7061(03)00041-7.

Arora, G., Agarwal, S. K., Shukla, S. 2021. Agricultural input use and index insurance adoption: Concept and evidence. Working Paper. https://doi.org/10.22004/AG.ECON.304508

Asafu-Adjaye, John. 2008. "Factors Affecting the Adoption of Soil Conservation Measures: A Case Study of Fijian Cane Farmers." *Journal of Agricultural and Resource Economics* 33 (1): 99–117. https://www.jstor.org/stable/41220616.

Berazneva, Julia, Linden McBride, Megan Sheahan, and David Güereña. 2018. "Empirical Assessment of Subjective and Objective Soil Fertility Metrics in East Africa: Implications for Researchers and Policy Makers." *World Development* 105 (January): 367–82. https://doi.org/10.1016/j.worlddev.2017.12.009.

Chambers, RG & Quiggin, J 2000, *Uncertainty, Production, Choice and Agency: The State-Contingent Approach*. vol. 1, First edn, Cambridge University Press, New York.

Chambers, Robert G., and John Quiggin. 2001. "Decomposing Input Adjustments Under Price and Production Uncertainty." *American Journal of Agricultural Economics* 83 (1): 20–34. https://doi.org/10.1111/0002-9092.00134.

Corbeels, Marc, Abebe Shiferaw Dakka, and Mitiku Haile. 2000. "Farmers' Knowledge of Soil Fertility and Local Management Strategies in Tigray, Ethiopia." *Managing Africa's Soils (No. 10),* Accessed here.

Delgado, Luciana, and Jetse J. Stoorvogel. 2022. "Role of Soil Perception and Soil Variability by Smallholder Farmers in the Low Adoption Rates of Extension Packages in Central America." *Journal of Rural Studies* 93 (May): 92–103. https://doi.org/10.1016/j.jrurstud.2022.05.009.

Desbiez, Arnaud, Robin Matthews, Bhaba Tripathi, and Jim Ellis-Jones. 2004. "Perceptions and Assessment of Soil Fertility by Farmers in the Mid-hills of Nepal." *Agriculture Ecosystems & Environment* 103 (1): 191–206. https://doi.org/10.1016/j.agee.2003.10.003.

Di Falco, S. 2006. "Crop Genetic Diversity, Farm Productivity and the Management of Environmental Risk in Rainfed Agriculture." *European Review of Agricultural Economics* 33 (3): 289–314. https://doi.org/10.1093/eurrag/jbl016.

Di Falco, Salvatore, and Jean-Paul Chavas. 2009. "On Crop Biodiversity, Risk Exposure, and Food Security in the Highlands of Ethiopia." *American Journal of Agricultural Economics* 91 (3): 599–611. https://doi.org/10.1111/j.1467-8276.2009.01265.x.

Di Falco, Salvatore, Jean-Paul Chavas, and Melinda Smale. 2007. "Farmer Management of Production Risk on Degraded Lands: The Role of Wheat Variety Diversity in the Tigray Region, Ethiopia." *Agricultural Economics* 36 (2): 147–56. https://doi.org/10.1111/j.1574-0862.2007.00194.x.

Donfouet, Hermann Pythagore Pierre, Aleksandra Barczak, Cécile Détang-Dessendre, and Elise Maigné. 2017. "Crop Production and Crop Diversity in France: A Spatial Analysis." *Ecological Economics* 134 (January): 29–39. https://doi.org/10.1016/j.ecolecon.2016.11.016.

Gruver, J.B., and R.R. Weil. 2007. "Farmer Perceptions of Soil Quality and Their Relationship to Management-sensitive Soil Parameters." *Renewable Agriculture and Food Systems* 22 (4): 271–81. https://doi.org/10.1017/s1742170507001834.

Bedoya, Guadalupe, Arguelles, Luca Bittarello, Jon Davis, Nikolas Mittag, Jonathan Martin Villars Davis, and Nikolas Karl Mittag. 2018. “Distributional Impact Analysis: Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3006213.

ICRISAT. 2014. VDSA Timeline Flyer. Accessed here.

Liebig, Mark A., John W. Doran, J. B. Gruver, R. R. Weil, Jeffrey E. Herrick, José Sarukhán, Samantha J. Langley-Turnbaugh et al. "Evaluation of Farmers' Perceptions of Soil Quality Indicators." *American Journal of Alternative Agriculture 14*, no. 1 11-21. https://doi.org/44503080.

Ministry of Agriculture and Farmers’ Welfare (MoAFW). 2022. Report of the Rainfed Farming Division. Accessed here.

Moges, Awdenegest, and N. M. Holden. 2007. “Farmers’ Perceptions of Soil Erosion and Soil Fertility Loss in Southern Ethiopia.” *Land Degradation and Development* 18 (5): 543–54. https://doi.org/10.1002/ldr.795.

Mullen, J. D. 2016. Impact assessment of ICRISAT Village Level Studies: 1975 to 2013. Tech. Rep., An independent external review commissioned by ICRISAT. Accessed here.

Nelson, Carl H., and Paul V. Preckel. 1989. “The Conditional Beta Distribution as a Stochastic Production Function.” *American Journal of Agricultural Economics* 71 (2): 370–78. https://doi.org/10.2307/1241595.

Okoba, B. O., and J. De Graaff. 2005. “Farmers’ Knowledge and Perceptions of Soil Erosion and Conservation Measures in the Central Highlands, Kenya.” *Land Degradation and Development* 16 (5): 475–87. https://doi.org/10.1002/ldr.678.

Ravisankar, Natesan, Meraj A. Ansari, Mohammad Shamim, Ashisa K. Prusty, Raghuveer Singh, Azad S. Panwar, Debashis Dutta, et al. 2022. “Sustainable Livelihood Security of Small Farmers Improved Through a Resilient Farming System in Semi arid region of India.” *Land Degradation and Development* 33 (15): 2830–42. https://doi.org/10.1002/ldr.4358.

Ray, Deepak K., James S. Gerber, Graham K. MacDonald, and Paul C. West. 2015. “Climate Variation Explains a Third of Global Crop Yield Variability.” *Nature Communications* 6 (1): 5989. https://doi.org/10.1038/ncomms6989.

Smithson, Michael, and Jay Verkuilen. 2006. “A Better Lemon Squeezer? Maximum-likelihood Regression With Beta-distributed Dependent Variables.” *Psychological Methods* 11 (1): 54–71. https://doi.org/10.1037/1082-989x.11.1.54.

Terza, Joseph V., Anirban Basu, and Paul J. Rathouz. 2007. “Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling.” *Journal of Health Economics* 27 (3): 531–43. https://doi.org/10.1016/j.jhealeco.2007.09.009.

Tesfaye, Wondimagegn, and Nyasha Tirivayi. 2019. “Crop Diversity, Household Welfare and Consumption Smoothing Under Risk: Evidence From Rural Uganda.” *World Development* 125 (September): 104686. https://doi.org/10.1016/j.worlddev.2019.104686.

Waha, Katharina, Jan Philipp Dietrich, Felix T. Portmann, Stefan Siebert, Philip K. Thornton, Alberte Bondeau, and Mario Herrero. 2020. “Multiple Cropping Systems of the World and the Potential for Increasing Cropping Intensity.” *Global Environmental Change* 64 (August): 102131. https://doi.org/10.1016/j.gloenvcha.2020.102131.

Wickens, Gerald E. 1998. “Arid and Semi-arid Environments of the World.” In *Adaptations of Desert Organisms*, 5–15. https://doi.org/10.1007/978-3-662-03700-3_2.

You, Liangzhi, Stanley Wood, and Ulrike Wood-Sichra. 2009. “Generating Plausible Crop Distribution Maps for Sub-Saharan Africa Using a Spatially Disaggregated Data Fusion and Optimization Approach.” *Agricultural Systems* 99 (2–3): 126–40. https://doi.org/10.1016/j.agsy.2008.11.003.

Chapter 2: A quantile regression approach to estimate heterogeneous credit-yield linkage and elicit binding credit constraints: Evidence from semi-arid villages in India

Sumedha Shukla

(IIIT-Delhi)

sumedhas@iiitd.ac.in

Gaurav Arora

(IIIT-Delhi)

gaurav@iiitd.ac.in

Working Paper Link: https://ageconsearch.umn.edu/record/314027?ln=en&v=pdf

DOI: https://doi.org/10.22004/AG.ECON.314027

A quantile regression approach to estimate heterogeneous credit-yield linkage and elicit binding credit constraints: Evidence from semi-arid villages in India

Abstract:

Agricultural credit can relax working-capital constraints for lesser-endowed farmers, yet access remains skewed, reflecting underlying socio-economic disparities. Equilibrium credit rationing in the presence of asymmetric information is an economic channel to articulate disproportionate credit access and favourable production outcomes for better-endowed farmers. This study employs quantile regression models in conjunction with propensity score matching to estimate the heterogeneous impact of agricultural credit on crop yield quantiles for a representative sample of farmers operating millet-based cropping systems in semi-arid tropics of India. We exploit the within-sample variation in the credit-yield linkage to augment sample-wide detection of binding credit constraints. The proposed approach moves beyond average *realized* outcomes, to derive *ex-ante* odds of credit rationing. We interpret these odds in a state-contingent technology framework revealing previously unexplored structural implications of rationing for farmers' adaptive capacity under uncertainty. We further propose formal statistical hypotheses to empirically uncover the rationing mechanism through which equally productive farmers may face differential credit terms on the basis of group-based affiliations that influence lender perception about repayment ability. We find that credit attainment is skewed toward upper-caste, asset-rich, land-owning households with higher-quality soils. Borrowers experience higher yields relative to a no-credit scenario, but gains are concentrated in the upper yield quantiles, indicating higher yield dispersion due to credit attainment. The average net revenue gained per rupee of credit exceeds borrowing costs by slightly less than a rupee (per acre), consistent with weakly binding credit constraints for the average farmer. However, given the skewed (but positive) credit impact across yield quantiles, median impact of credit is likely lower, suggesting binding constraints for only a *subset* of borrowers. Returns from crops inter-cropped with millets drive *ex-ante* rationing among ~48% of borrowers. Rationed households are found to systematically employ irrigation and superior seed varieties less frequently, but increase reliance on low-cost labour and fertilizer. Evidence suggests wealth-based disparities in credit allocation and rationing which are only partially mitigated by observed productivity indicators. Overall, these findings raise important questions for agricultural credit targeting in semi-arid regions.

Keywords: agricultural credit, crop yields, propensity score matching, quantile regression, state contingent approach

JEL Codes: Q12, O13, C14

A quantile regression approach to estimate heterogeneous credit-yield linkage and elicit binding credit constraints: Evidence from semi-arid villages in India

2.1 Introduction

Agricultural credit is argued to be an important instrument for improving rural incomes and reducing poverty. At the core of this argument is the thesis that credit can enable small and marginal farmers to exit low capital constraints and pursue farming with the level of resources that is similar to their better-endowed counterparts. However, the evidence points to unequal credit impacts that reflect pre-existing socio-economic gaps among producers (Carter 1989; Liverpool and Winter-Nelson 2010; Abate et al. 2016; Jimi et al. 2019).

Credit production linkages depend on attributes of the production environment (Diagne and Zeller 2001), farmers' endowments (Liverpool and Winter-Nelson 2010), and the structure of local lending institutions (Siamwalla et al. 1990; Kochhar 1997; Abate et al. 2016). Farm credit utilization depends on the expected per-acre returns from cropland management options such as adopting a new technology (e.g., hybrid seeds) or investing in long-term assets (e.g., groundwater wells) (Carter and Wiebe 1990; Ciaian et al. 2012; Regassa et al. 2023). Relaxing liquidity constraints through credit provision has been shown to enhance farm output supply (Sial and Carter 1996; Regassa et al. 2023), yet credit-led productivity improvements can be muted when input or labour use is distorted by adverse production conditions, such as dry soils or inadequate infrastructure (Diagne and Zeller 2001; Ali et al. 2014). For example, Nguyen et al. (2023) found that credit access enables the mitigation of adverse weather impacts in Vietnam, whereas Ali and Deininger (2012) reported such favourable outcomes to be restricted to low drought-risk regions of Ethiopia.

The impact heterogeneity outlined above arises, in part, from the endogeneity of the credit-production relationship i.e., credit can positively impact production while access to credit itself depends on farm productivity potential (Braverman and Guasch 1986; Carter 1988; Feder et al. 1990). The endogeneity problem can be articulated through borrowing and lending behaviors in rural credit markets as follows. An incentive to borrow is derived from farm-level capital constraints where production costs must be incurred prior to realizing end-of-season returns. Lenders provide loans in lieu of an interest premium and a collateral security. Enforcing rural credit contracts by ensuring timely repayment is costly because farm production is uncertain and lenders can only partially observe farmers' cropping potential and farm output i.e., there is information asymmetry.

Asymmetric information gives rise to adverse selection and moral hazard issues, which cause a truncation in lender returns with an increase in interest rate (Stiglitz and Weiss 1981), collateral (Stiglitz and Weiss 1981) and loan amount (Jaffee and Russell 1976), effectively limiting the *size* and *quantity* of loans supplied by the lender, i.e., *credit rationing* in equilibrium. At the household level, credit rationing translates into unmet credit demand i.e., binding credit constraints (Feder et al. 1990; Sial and Carter 1996; Boucher et al. 2009). The distribution of these constraints across the population however, is not equal and is typically a function of the pre-existing socioeconomic status of farmers (Braverman and Guasch 1986; Ali and Deninger 2012) that can also drive the credit impacts (Regassa et al. 2023; Beaman et al. 2023) aften through mediating channels such as input availability, technology adoption and land ownership (Kochar 1997; Deninger and Olinto 2000; Abate et al. 2016).

In this paper we estimate heterogeneous credit impacts across crop yield quantiles for a sample of small farms in India. Through this approach we seek to clarify that differential credit impacts are driven by complex socioeconomic channels that also embed in credit access restrictions. We exploit the within-sample variation in the credit-yield linkage to augment sample-wide detection of binding credit constraints (i.e., rationing). The proposed approach moves beyond detecting rationing *occurrence* based on average *realized* outcomes, to derive *ex-ante odds* of credit rationing (i.e., at the time of borrowing). We then interpret these odds in a state-contingent technology framework revealing previously unexplored structural implications of rationing for farmers' adaptive capacity under uncertainty. We further use these odds to propose formal statistical hypotheses to empirically uncover the rationing mechanism through which equally productive farmers may face differential credit terms on the basis of group-based affiliations that influence lender perception about repayment ability.

We will now articulate equilibrium credit rationing in the presence of asymmetric information as an economic channel for both disproportionate credit access and favourable production outcomes. Default risk of agricultural credit arises from uncertainty in farmer returns due to production environment attributes and risk management behaviour (e.g., input use). Loan contracts are designed to satisfy the lender's participation constraint (expected returns exceed opportunity cost of capital) and borrower's incentive compatibility (inducing "less-risky" production plans and repayment). Increments in interest rate or loan amount worsen adverse selection as honest borrowers and safer projects drop out of the market, and intensify moral hazard as remaining borrowers shift toward riskier production plans. Consequently, expected lender returns are a non-monotonic function of the interest rate such that the expected return-maximizing interest rate lies below the market-clearing level (Stiglitz and Weiss 1981). Similarly, the lender's zero-profit loan offer curve in loan size-interest rate space implies an equilibrium loan size cap below the market-clearing quantity (Jaffee and Russell 1976). Some borrowers who are observationally similar to those receiving their desired loan size are therefore denied credit fully (quantity rationing) or receive smaller loans than desired (loan-size rationing). Importantly, these rationing equilibria describe the existence of rationing but not its distribution, i.e., they do not specify *which borrowers* within or across observable classes (e.g., based on land size groups) are rationed.

Carter (1988) extends this framework to smallholder agriculture and shows that equilibrium rationing can take the form of *statistical discrimination*, whereby lesser-endowed farmers are selectively rationed out of the credit market. Lenders use observable characteristics such as land ownership and land size as proxies for repayment capacity, assign borrowers to groups, and impute upon each borrower the repayment ability of the group's representative farmer. As a result, farmers from groups perceived as *riskier* or *less productive* face worse contract terms than equally productive farmers from 'safer' groups.

In a similar vein Kjenstad et al. (2015) extend the argument in Jaffee and Russell (1976) to show that equilibrium loan size rationing can arise even without asymmetric information. Specifically, if default risk increases with loan size borrowers always demand more credit than lenders are willing to supply and ex-post agency costs (e.g., bankruptcy costs) further widen this gap. It can therefore be reasoned that if the lender expects ex-post agency costs to be higher for one group of farmers (e.g., small) relative to others (say large), loan size rationing is likely to be more severe for the former relative to the latter. In this manner group-based imputation of repayment ability leads to credit supply rules that prioritise pre-existing endowments over farming ability, and provides a structural rationale for endogenous credit impacts and non-uniform farm outcomes.

At the household level, rationing presents as binding credit constraints and underpins the rationale for subsidized rural credit programs worldwide (David and Meyer 1979; Braverman and Guasch 1986; Boucher et al. 2009). Detecting credit rationing and assessing its impact is therefore a central policy concern, typically pursued through two approaches. The first directly elicits rationing through end-of-season survey-based measures of unmet credit demand (e.g., Feder et al. (1990); Petrick (2004)) while the second infers rationing from the shadow value of capital in farmers' optimization problem, either estimated directly (e.g., Sial and Carter (1996)) or inferred indirectly through efficiency of input use or expenditure (e.g., Deininger and Olinto (2000), Kumbhakar and Bokusheva (2009))[3]. Both approaches however rely on realized outcomes with the former extensively employed in recent years to study structural differences in agricultural outcomes driven by credit rationing status (Diagne and Zeller 2001; Foltz 2004; Petrick 2004; Ali et al. 2014; Regassa et al. 2023). In fact, most evidence on heterogenous credit impacts documents differences across specific socioeconomic parameters such as land rights (Jimi et al. 2019), lender type (Regassa et al. 2023), farm size (Abate et al. 2016) or wealth status (Liverpool and Winter-Nelson 2010) but does not link them structurally to the nature of credit access restrictions faced by farmers.

Our study presents a general quasi-experimental design that employs quantile regressions in conjunction with PSM to estimate credit impacts across *millet* yield quantiles using the VDSA dataset for 2001-2014. Matching is appropriate in our context because we encounter a short panel whereby the median household grows millets only thrice during 2001-2014[4]. We focus on a crop-specific indicator since agricultural credit provision is intrinsically tied to the crop(s) that farmer intends to grow (Besley 1994; Reserve Bank of India (RBI) 2017). Whereas past studies focus on farm-level outcomes (e.g., output supply) in documenting the credit impacts, our study adds to the limited evidence on credit-yield linkages which mediate (oft-reported) ambiguous credit impacts on farm outcomes[5] (please see Table 1, Supplementary Information (SI) Section 1(S1)). Importantly, unless yields are consistently ordered between the treated and counterfactual states, quantile treatment effects may not correspond to the quantiles of the treatment effect distribution. This is an identification condition known as *rank invariance*, which is typically assumed to be true (Angrist and Pischke 2009). While rank invariance cannot be tested in practice, we discuss an alternative *rank similarity condition* and propose a strategy to test it when PSM is employed for endogeneity resolution, interpreting our credit-yield impacts considering these results[6].

[3]Positive shadow indicates a binding capital constraint, meaning that farmers have an incentive to borrow more credit for production. When the shadow value is zero farmers are considered to be credit-*un*constrained (Sial and Carter 1996). Suboptimal input use (despite positive marginal returns) suggests binding credit constraints.
[4]Sorghum yields are rarely analyzed by economists, despite the crop's critical role as a nutrition source in semi-arid regions globally (Isom and Worker 1979; Hossain et al. 2022). Our crop choice is also driven by data availability for a causal analysis (discussed hereafter).
[5]Foltz (2004) found a positive credit impact on farm profitability, while Carter (1989) and Ciaian et al. (2012)) found its impact on farm output was muted.
[6] Our statistical tests are motivated by Dong & Shen (2018), & Frandsen & Lefgren (2018). A detailed discussion of the test is outside of the scope of this paper. However we leverage the results from this exercise in interpreting our results (please see Shukla and Arora (2022)).

Since unequal credit access partly explains heterogeneous credit impacts, we utilize the estimated quantiles-based credit-yield linkage for indirect elicitation[7] of credit rationing by parametrizing *ex-ante* (i.e., at the time of borrowing) credit rationing odds across sample households which moves beyond detecting rationing *occurrence* based on average *realized* outcomes. We characterize these odds by articulating the impact of rationing induced credit constraints on farmer adaptation to cropping risks through input use decisions in a state contingent technology framework (Chambers and Quiggin 2000; Rasmussen 2003). We show theoretically that rationing-induced credit constraints lower farmer welfare and lead to *riskier* net revenue bundles in the second-order stochastic dominance sense. A decomposition of input adjustments driving this shift draws on a stochastic version of Shepherd's Lemma (Chambers and Quiggin 2001) and reveals that a binding credit constraint *may increase or decrease input use depending on the risk-influencing properties of a specific input*. We further propose formal statistical hypotheses to empirically uncover the rationing mechanism through which equally productive farmers may face differential credit terms on the basis of group-based affiliations that influence lender perception about repayment ability.

Overall, our paper will provide a framework to evaluate the utility of credit for informing efficient credit targeting policy. The rest of this essay is organized as follows. Section 2.2 provides contextual details of the Indian agricultural credit sector and outlines main features of our data. Section 2.3 outlines the methodology, followed by Results and Discussions in Section 2.4 and Conclusions and Future Work in Section 2.5.

2.2. Context and Data

2.2.1 Agricultural Credit in India

"Agricultural credit falls short of the right quantity, is not of the right type … and often fails to go to right people." --- All India Rural Credit Survey (RBI), 1954

A central theme of India's agricultural policy is to expand institutional credit access in an inclusive fashion. Reflecting this spirit, institutional credit disbursal (as a percentage of agricultural gross value added (GVA)) grew on-average 4.7% per year during 1970-2022 (Figure 2.1, Panel A). Formal credit expansion relied on a multi-agency architecture comprising commercial banks, agricultural cooperatives (introduced in 1904), and regional rural banks (RRBs) established since 1986. Despite the institutional diversity, since early 2000s, credit expansion is led by commercial banks while cooperatives and RRBs struggle

[7] We employ the indirect approach for rationing detection because credit constraints were not explicitly recorded in VDSA data.

with large repayment overdue (Mohan 2006) - heightening credit rationing concerns given higher information asymmetry faced by commercial banks relative to cooperatives & RRBs.

In 1969, agriculture was granted 'priority sector lending' status under which commercial banks were required to allocate 18% of total credit to agriculture. This was combined with a socially coercive branch expansion rule (rolled out in 1977) that mandated three rural branches in unbanked locations for every new branch in an already banked area (Gulati and Juneja 2019). Unmet disbursal targets motivated further measures such as the Special Agriculture Credit Plan (SACP) in 1994 which required annual growth of 20-25% in amount of agricultural credit disbursed. The SACP was extended to private banks in 2004 with additional directives to channel 40% of credit to "small and marginal" farmers (<2 hectares), who comprised over 85% of agricultural households (Agricultural Census 2016) but historically received less than 30% of credit (reference). Complementary initiatives included differential interest rate scheme (DRIS) which offered loans up to ₹15,000 at 4% interest to economically weaker and socially disadvantaged groups (RBI 2009a), and the introduction of Kisan Credit Cards (KCC) in 1998 to provide collateral free short-term crop loans i.e., loans of duration <12 months taken with a stated purpose of funding cropping requirements.

Despite this emphasis on commercial lending for agriculture, Figure 2.1(Panel A) shows that cooperatives held a majority share in the formal credit market until 2005. Interestingly, commercial banks surpassing cooperative loan shares coincides with the introduction of interest subvention scheme (ISS) in 2006. Under ISS, short-term crop loans of up to ₹300,000 were issued at a fixed interest rate of 7%, with additional interest subventions of up to 3% for prompt repayment announced in subsequent years (Press Information Bureau 2014). This represented a reduction in interest rates ceiling relative to the earlier benchmark prime lending rate (BPLR) system, under which the ceiling for loans up to ₹200,000 was tied to each bank's BPLR which was typically higher because it incorporated cost of funds, operating expenses, and a profit margin (RBI 2009a). Figure 2.1, Panel B displays the BPLR interest rate bands until 2006[8] and the interest subvention ceiling bands thereafter.

Part of the rationale for the ISS shift was the persistence of sub-BPLR lending whereby banks offered concessional interest rates below BPLR to borrowers they deemed "creditworthy," while poorer or riskier farmers typically received loans only at the ceiling. This pattern disproportionately disadvantaged farmers with fewer assets or weaker social

[8] Bands refer to the BPLR rates announced by five major public sector banks until 2003-04 and five major banks thereafter. For e.g., State Bank of India's declared BPLR rate in 2001 was 11.5% (see the full list here)

networks (RBI 2009a; RBI 2009b). Such disparities in credit access and loan amounts have been documented by caste (Kumar 2013; Karthick and Madheswaran 2018; Zeeshan et al. 2025), wealth (Bhende 1986; Swain 2002; Gulati and Juneja 2019) and land ownership (Mohan 2006; All India Rural Financial Inclusion Survey 2016) in India. A closer examination of loan disbursements in semi-arid districts reveals a similar pattern. Average loan amount received by medium and large landholders is more than double that of small and marginal landholders (Figure 2.1, Panel C). Moreover, within the VDSA data, loan disbursements increase by wealth quartiles although the pattern diminishes over time (see Figure 2.1, Panel D). Overall, although formal agricultural credit has grown substantially in volume, persistent disparities in access and loan amounts imply that binding credit constraints arising from credit rationing remain a credible concern.

2.2.2 Study Region, Data Structure and Crop Choice

We employ the VDSA data which constitutes plot-level farm production information of 520 *sorghum* (*great millet*) growing households located across eighteen semi-arid and tropical villages in eight Indian states during 2001-2014 period. Figure 2.2 provides a representation of our study region, i.e., village locations and survey-years for each state[9]. The VDSA villages located in major semi-arid and tropical agro-climatic zones of India were selected to represent key cropping and irrigation patterns in the region, where survey households were selected randomly to represent the distribution of land size in each village (Rao et al. 2011). Our sample VDSA states accounted for ~85% national sorghum acreage and output (by weight) during the study period (please see SI Section 2 (S2), Figure 1).

We present a representativeness assessment of sorghum-growing farmers in our sample with reference to the district-level statistics for farm operations and farmer demographics obtained from the 2001 and 2011 waves of Agricultural Census. Specifically, the sorghum-growing farmer in our sample is on average 50 years old with 5 years of education; operates 4.5-acre farms from which about 2-acre land is allocated for sorghum cultivation. On the contrary, the district-level farmers have similar age and education profile; and operate 6-acre farms with an average plot size of 3 acres. Further, the average sorghum yield is 345 kg/acre in our sample, similar to the national average of 352 kg/acre (please see SI, Figure 2(S2)).

[9]Surveys were conducted through Indian crop-years, which are different from calendar years. For example, crop-year 2001-'02 means July 2001 to June 2002 period that constitutes three growing seasons: *Kharif* (July-October), *Rabi* (November-March) and *Zaid* (April-June).

Each household may operate one or more land plots and grow multiple crops across the growing seasons of each agricultural year. Data on credit attainment for a household were recorded once every year around the month of July[10]. Households borrow for consumptive, agriculturally productive, or non-farm investment purposes, comprising 41%, 41%, and 9% of total borrowing instances, respectively. Our focus is on agriculturally productive borrowing and we encounter a highly unbalanced panel of sorghum-growing households such that the median household grows sorghum only three times in fourteen years. The short length of this panel dataset is driven partly by the frequency of VDSA surveys whereby majority (~50%) households were surveyed only during the 2009-2014 period (see Figure 2). Diversified cropping is utilized as a strategy to increase cropping resilience (Isom & Worker 1979), such that our sample households on-average grow six crops annually. Sorghum is typically intercropped with three kinds of crops: (a) pulses including pigeonpea, chickpea, cowpea, blackgram and greengram; (b) cereals, i.e., rice, wheat and maize; (c) commercial crops including sugarcane, cotton, oilseeds (castor, groundnut, soybean and sunflower) and vegetables (onions and chillies).

Agricultural loans for sorghum production are mostly short-term (i.e., duration is less than 12 months) with an average interest rate of ~10%. Credit attained for sorghum production is typically utilized for fulfilling working capital requirements. A breakup of the cost of production for sorghum as per national comprehensive analysis of cost of production (CACP) surveys reveals that human labour accounts for majority of the working capital costs across the sorghum-producing VDSA states while fertilizer and irrigation account for only 15 percent (please see SI, Figure 3(S2)). Notably, the combined cost of animal and human labour exceeds 70 per cent of total production costs, reflecting the low level of mechanization in sorghum production across semi-arid regions (Bhattarai et al. 2020).

Most borrowers in our sample are also *first-time-growers* (FTG) for whom the first year of credit attainment coincides with the first year of sorghum cultivation (see Figure 2.3(Panel A)). This means that we do not have data for *pre*-borrowing periods for such households, i.e., 70% in formal credit and 62% in informal credit categories. Sorghum-growers that were observed for at least one year of pre-borrowing and *post*-borrowing periods amount to just 25% and 33% of formal and informal borrowers, respectively, and systematically differ from our full sample of borrowers (see Figure 2.3(Panel B)). We implement a quasi-experimental

[10] All relevant details of data construction are outlined in SI Section 2.

design that relies on the PSM strategy to achieve causal inference from this restrictive data structure of our sample.

We observe a similar data structure is for all major crops in our study region[11]. We estimate the credit impacts for sorghum because our sample constitutes *both* borrowing and non-borrowing households that grow sorghum concurrently (i.e., in the same year). For other cases, credit attainment tends to be clustered by crop-type such that most households growing a particular crop obtain credit, and so the counterfactual data is insufficient to estimate credit impacts. In fact, farmers' choice of crop might itself reflect the ease of obtaining credit (Appau et al. 2019). For example, 97% of rice-growing households in our sample obtained credit (see SI, Section 3 for details).

2.3 Methodological Framework

Our methodology comprises of three main components. We first implement a quasi-experimental framework by specifying a quantile regression model in conjunction with PSM to estimate agricultural credit impacts across sorghum yield quantiles. Second, we utilize a state-contingent production technology framework to examine how credit rationing influences farmers' adaptive capacity under uncertainty through input use decisions in crop production adapting the framework presented in Chambers and Quiggin (2000; 2001) and Rasmussen (2004) for our credit constraint setup. Third, we propose a strategy for ex-ante estimation of credit rationing odds relying on the estimated credit-yield impacts across quantiles and specify empirical hypotheses to examine the economic mechanisms underlying credit rationing impacts in our sample.

2.3.1 Empirical Framework: Quantile Regressions with PSM

We designate the first-instance of obtaining agricultural credit as "treatment" for sorghum-growers during 2001-2014. Consequently, untreated households are drawn from the pool of non-borrowers in each time-period. We evaluate *first-instance* credit uptake because we encounter too-few non-borrowers when evaluating impact of multiple credit instances (i.e., most borrowers availed multiple loans in our sample). Our focus on *agricultural credit* implies that credit was utilized for cropping and farm investment purposes only. Note that we designate a binary treatment even though credit amounts were reported by borrowing households. This is because loan amounts may be systematically misreported in primary surveys (Zinman 2009; Wollburg et al. 2021). Moreover, we observe a high variance in loan

[11]Major crops include sorghum, cotton, paddy, wheat, pigeonpea and soybean.

amounts for distinct purposes (e.g., fertilizer purchase or well construction), which appeals to a context-specific impact attribution of marginal credit allocation on production outcomes. A binary treatment variable further facilitates estimation of the PSM-based treatment effect on yield distributions (Frölich 2007; Firpo 2007). Last but not the least, when estimating credit impacts from formal lenders, we are also cognizant of lending from informal sources such as family and friends and make the necessary exclusions in our estimation framework[12].

We record the credit borrowing activity of sorghum-growers denoted as $i \in I = \{1, 2, ..., N\}$, for time-periods, $t \in \{t_1, t_2, ..., t_{14}\}$ where t_1 denotes the year 2001, t_2 denotes 2002 and so on. Define subgroups, $\{g_k\}_{k=1}^{14}$, that sort households by the calendar year of first obtaining credit. Households in group g_1 obtained agricultural credit for the first time in period t_1; households in group g_2 first obtained credit in t_2; and so on. Households that never borrow until the end of our study are collected in group g_∞. The control group for set g_k is denoted as g_l such that $l > k$, meaning g_l comprises of households that did not borrow until t_k. Accordingly, we define $D_{g_k t} = 1(t \geq t_k)$ as the binary treatment variable that indicates whether credit had been borrowed in group g_k in period t[13]. Trivially, $D_{g_\infty t} = 0$ for all t. Treatment status for household $i \in g_k$ is denoted as $D_{i_k t} = 1(t \geq t_k)$. $D_{i_k t} = 0$ refers to households $j_k \in g_{l:l>k}$ that are potential counterfactuals for first-time borrowers in period t_k.

We denote sorghum yields for household i in period t as random variable Y_{it}. The impact of credit can be evaluated by comparing the distribution of potential yields under two counterfactual states: with-credit, i.e., $Y_{1i_k} \sim F_1(Y)$ and without-credit, i.e., $Y_{0i_k} \sim F_0(Y)$. Here index i_k specifies treatment timing (t_k) in Y_{1i_k}, while Y_{0i_k} refers to *i*'s counterfactual yield had it not borrowed until t_k. A quantile yield function specifies τ^{th}-yield quantile *with credit*, $Q_\tau(Y_{1i_k}) = \inf\{y_{1i_k} : F_1(Y) \geq \tau\} \equiv F_1^{-1}(\tau)$ [14], and *without credit*:

[12]About 10% households reported first-instance credit uptake from both lender-types in the same year. We excluded such households from our analysis. Only households that access credit from formal sources *prior to* informal sources are part of our sample.

[13]Our notation refers to a "staggered treatment" status in event studies (please see Callaway and Sant'Anna 2021 for details).

[14]Angrist and Pischke (2008) clarify that the equivalence (≡) applies for continuous random variables with well-behaved density functions that do not have gaps or spikes.

$Q_\tau(Y_{0i_k}) = \inf\{y_{0i_k} : F_0(Y) \geq \tau\} \equiv F_0^{-1}(\tau)$ where $\tau \in (0,1)$, and y_{1i_k} and y_{0i_k} refer to the random sequences drawn from $F_1(Y)$ and $F_0(Y)$, respectively. The quantile treatment effect (QTE) is defined as $Q_\tau(Y_{1i_k}) - Q_\tau(Y_{0i_k})$.

A classical issue in evaluating QTE is that in the real-world borrowers are not simultaneously observed in the counterfactual states. To estimate QTE we rely on sample yields $y_{it} = D_{i_k t} y_{1it} + (1 - D_{i_k t}) y_{0it}$ to represent Y_{it}. For $i \in g_k$, yields observed in period t_k and onwards represent treated yields, $y_{1it} \equiv y_{it:t \geq t_k}$ whereas yields observed prior to t_k represent control yields, $y_{0it} \equiv y_{it:t<t_k}$. Recall that our binary treatment definition $D_{i_k t_k} = 1$ denotes first-time borrowers in t_k whereas $D_{i_k t_k} = 0$ denotes potential counterfactuals for these first-time borrowers (i.e., $j_k \in g_{l:l>k}$). A conditional quantile function (CQF) specifies τ^{th}-yield quantile in each treatment year t_k, i.e., $Q_\tau(Y_{1ik} \mid D_{i_k t_k} = 1) = \inf\{y_{i_k t_k} : F_1(Y) \geq \tau\}$ for borrowers (*with-credit*), and $Q_\tau(Y_{0ik} \mid D_{i_k t_k} = 0) = \inf\{y_{i_k t_k} : F_0(Y) \geq \tau\}$ for non-borrowers (*without-credit*), where $\tau \in (0,1)$ and $k \in \{1, 2, ..., 14\}$. The quantile treatment effect for the treated (QTT) is defined as $Q_\tau(Y_{1i_k} \mid D_{i_k t_k} = 1) - Q_\tau(Y_{0i_k} \mid D_{i_k t_k} = 1)$.

We can further summarize the quantile impact estimator through a quantile regression model for yields conditional on the treatment status ($D_{i_k t_k}$) and control vector ($\vec{X}_{i_k t_k}$). That is,

$$Q_\tau(Y_{i_k t_k} \mid \vec{X}_{i_k t_k}, D_{i_k t_k}) = \alpha_{0\tau} + \alpha_\tau D_{i_k t_k} + \vec{X}'_{i_k t_k} \vec{\beta}_\tau ; Q_\tau(u_{i_k t_k} \mid \vec{X}_{i_k t_k}) = 0; \tau \in (0,1); k \in \{1, 2, ..., 14\} \quad (1)$$

Equation (1)[15] stacks τ-quantile yields for treated (g_k) and control ($g_{l:l>k}$) groups on the left-hand side (L.H.S.) as a function of $D_{i_k t_k}$ and $\vec{X}_{i_k t_k}$. Assume $Y \perp D \mid X$, then model (1) provides *sample-based estimate* of conditional quantile treatment effect for the treated (CQTT), i.e., $\hat{\alpha}_\tau = Q_\tau(Y_{1i_k} \mid \vec{X}'_{i_k t_k}, D_{i_k t_k} = 1) - Q_\tau(Y_{0i_k} \mid \vec{X}'_{j_k t_k}, D_{i_k t_k} = 0)$. $\hat{\alpha}_\tau$ estimates sample-based difference in yield quantiles *with* and *without-credit* conditional on $\vec{X}$. For example, $\hat{\alpha}_{\tau=0.1}$ provides an estimate of change in borrowers' lower decile of yields conditional-on-$\vec{X}$.

---

[15] The τ^{th} conditional quantile regression estimator $\hat{\beta}_\tau$ is obtained as a solution to the following optimization problem: $\min_{\beta_\tau} \sum_{i:y_i \geq x_i'\beta_\tau}^{N} \tau \mid y_i - x_i'\beta_\tau \mid + \sum_{i:y_i < x_i'\beta_\tau}^{N} (1-\tau) \mid y_i - x_i'\beta_\tau \mid$.

Comparing $\hat{\alpha}_\tau$ across $\tau \in (0,1)$ can reveal whether credit causes higher or lower dispersion in yields distribution.

We clarify that $\hat{\alpha}_\tau$ quantifies the difference in conditional quantiles of yield *distribution*, and *not* changes in individual/household yields under treatment and control states. For example, we can infer whether credit raises (or decreases) the lower decile of yields (i.e., $\tau = 0.1$) but cannot claim whether individual borrowers in $\tau = 0.1$ quantile experienced higher or lower yields. Mathematically, $\hat{\alpha}_\tau \neq \hat{Q}_\tau(Y_{1,it} - Y_{0,it} \mid \vec{X}_{it})$ *unless* the *rank invariance assumption* holds, i.e., each household's yield-rank is same irrespective of treatment status. It turns out that rank invariance is not testable in practice. We discuss an alternative *rank similarity condition* and provide a statistical test for this condition (please see SI Section 4). Second, $\hat{\alpha}_\tau$ does not provide an inference on the marginal yield quantiles, i.e., $Q_\tau(Y_{it} \mid \vec{X}_{it}) = \vec{X}_{it}^{'}\beta_\tau$ does not imply $Q_\tau(Y_{it}) = Q_\tau(\vec{X}_{it}^{'})\beta_\tau$ (Angrist and Pischke 2009).

Note that $\hat{\alpha}_\tau$ in model (1) naively posits yield reports of non-borrowing households to represent the counterfactual yields of borrowing households. However, yields are likely to be endogenous to farmers' credit access. Omitted factors in model (1) that drive differential credit access among farmers, i.e., non-zero correlation with D, will lead to biased estimation of α_τ. To reconcile this issue we employ a propensity score matching approach whereby non-borrowers are denoted as counterfactuals whenever their likelihood of obtaining credit is similar to that of the borrowers in our sample.

2.3.1.1 Steps in propensity score matching

To construct a counterfactual for borrowers in our sample we first specify a logistic regression model for the first-instance of obtaining credit as a function of the factors that determine farmers' credit demand and credit supply rules of lenders in the area. That is,

$$\log\left(\frac{\Pr(g_k)}{\Pr(g_{l:l>k})}\right) \equiv \log\left(\frac{\Pr(D_{i_k t_k}=1)}{\Pr(D_{i_k t_k}=0)}\right) = \delta_0 + \vec{V}_{i_k t_k}^{'}\vec{\delta} + \varepsilon_{i_k t_k}; k = \{1,2,...,14\} \qquad (2)$$

In equation (2) we model a representative farmer's odds of obtaining credit in treatment-year t_k relative to the odds that she obtained credit in a later year (i.e., $t_l > t_k$). Consequently, vector $\vec{V}_{i_k t_k}$ comprises of variables that would explain the differential credit access between treated and control households in t_k. We define $\vec{V} \in \{\vec{V}^{eco}, \vec{V}^{fe}, \vec{V}^{ins}\}$ where $\vec{V}^{eco}$ captures economic situation of households (i.e., asset position, land ownership and non-farm income);

$\vec{V}^{fe}$ denotes time-invariant controls (i.e., household demographics and soil quality); and $\vec{V}^{ins}$ indicates credit access (i.e., macro-level credit supply and productivity potential).

The association between $\vec{V}^{eco}$ and credit access is expected to be ambiguous. Larger operational holdings can enhance credit demand and reduce barriers to credit markets (Feder et al. 1990). However, better asset position may also indicate lower credit reliance (Sial and Carter 1996; Petrick 2004). We account for institutions that facilitate credit access ($\vec{V}^{ins}$) through trends in national agricultural loan disbursal; number of bank branches in rural and semi-urban areas; and an indicator variable for differential credit activity during each year's Monsoon season (June-September) relative to dry and summer months (Kochar 1997; Kumar et al. 2020). $\vec{V}^{fe}$ represents household-specific, time-invariant factors that reflect an innate borrowing capacity; measured through prior borrowing experience (i.e., for purposes other than sorghum production); poor soils that suppress credit demand (Boucher et al. 2009; Ali et al. 2014) and provide a negative signal to lenders. Further, factors such as age, experience, caste and education status might be associated with farming ability and borrowing potential (Feder et al. 1990; Regassa et al. 2023). For example, older farmers tend to borrow less often (Sial and Carter 1996). Caste status provides a reflection on social network and access to farming resources (including informal credit) in India (Karthick and Madheswaran 2018; Rao 2017). Please see Table 2.1 for descriptions and summary statistics for variables included in model (2).

As the next step we assign counterfactuals to treated households, $i \in \{g_k\}_{k=1}^{14}$, based on the matching rule that non-borrowers, $j \in I \setminus \{g_k\}_{k=1}^{14}$, must exhibit a similar likelihood of obtaining credit (conditional on $\vec{V}$) as the borrowers in our sample. That is, household i is assigned counterfactual j if $|\hat{\Pr}(D_{it}=1)-\hat{\Pr}(D_{jt}=1)| \leq 0.05$ where $\hat{\Pr}(D_{it}=1)=\hat{\delta}_0+\vec{V}_{it}'\hat{\vec{\delta}}$ from model (2). Matching ensures a balanced distribution of confounding observables, $\vec{V}$, across treated and control cohorts (Rosenbaum and Rubin 1983), which implies conditional independence of potential yields, Y_0, from treatment assignment, i.e., $Y_0 \perp D \mid \vec{V}$.

PSM reconciles systematic treatment selection by balancing the distribution of covariates $\vec{V}$ across borrowers and non-borrowers. This implies that $Y_0 \perp D \mid \hat{\Pr}(D=1) \mid \vec{V}$ (Rosenbaum and Rubin 1983). It is possible, however, that household and/or time-specific determinants of farm outcomes (e.g., technology or market access) are also associated with the borrowing activity. Consequently, we introduce two constraints to our search rule for counterfactuals.

First, treated households are matched to non-borrowers in *the same* treatment-year i.e., $j(i_k):|\hat{\Pr}(D_{it_k}=1|i\in g_k)-\hat{\Pr}(D_{jt_k}=0|j\in g_{l:l>k})|\leq 0.05$ [16]. Second, besides matching by year, we search for counterfactuals from the same village as the treated households. Therefore, we define counterfactual j for household i treated in period k through the search rule that: $j(i_k):|\hat{\Pr}(D_{it_k}=1|i\in g_k)-\hat{\Pr}(D_{jt_k}=0|j\in g_{l:l>k},\nu_j=\nu_i)|\leq 0.10$ where ν_j represents the village of household *j* as recorded in VDSA data. The spatial-temporal dynamics of our matching strategy control for systematic production trends across treated and control households through time (e.g., see Liu and Lynch 2011) and space (e.g., see Binswanger and Rosenzweig 1986; Bhattarai et al. 2020)[17].

Overall, our matching strategy controls for village-year cluster effects, implying that observed and unobserved factors will be balanced across treated and counterfactual groups (Arpino and Cannas 2016; Huntington-Klein 2021; Chang and Stuart 2022). We further show that our matching strategy is analytically equivalent to including fixed effects in linear panel models to eliminate cluster-specific unobserved heterogeneity (see SI Section 5). Below we describe the quantile treatment estimation procedure based on the *matched* sample.

2.3.1.2 Estimating CQTT based on PSM-based counterfactual

PSM ensures a similar distribution of covariates or confounders of credit access across treatment and control households, thereby establishing conditional independence between treatment assignment and potential yields in the matched sample. It is well established that a difference in mean yields by treatment status provides a consistent estimate of how access to credit would change the average crop yields of non-borrowers. Frölich (2007) showed that PSM can be used not only to estimate the change in mean yields of non-borrowers if they enjoyed a similar credit access as the borrowers, but also to estimate the impact of credit on entire yield distribution of the non-borrowers.

Let $f_{V|D=1}$ and $f_{V|D=0}$ denote the density of a covariate that determines selection into the treated and control households, respectively. We are interested in evaluating credit-led change in yield distribution among the borrowers, i.e., $F_1(Y|D=1)-F_0(Y|D=1)$. Given that

[16] In practice it is possible that households in $g_{l:l>k}$ may not grow sorghum in k .Such households are excluded from our set of potential counterfactuals.

[17] For example, farm machinery rental services are typically localized within villages, leading to systematic variations in equipment availability across villages (Bhattarai et al. 2020).

$F_0(Y \mid D=1)$ is unobserved, an alternative näive comparison of yield distribution by treatment status, i.e., $F_1(Y \mid D=1) - F_0(Y \mid D=0)$, is biased. The consistent credit impact estimator would substitute $F_0(Y \mid D=0)$ with $\int F_0(Y \mid V=v, D=0) f_{V|D=1} dv$, i.e., yield distribution of non-borrowers *where* covariate (V) is distributed as in borrowers' population. PSM ensures that $f_{V|D=1} = f_{V|D=0}$ in the matched sample given that the support of covariate V is the same across cohorts (Frölich 2007). Moreover, when V is a multidimensional vector, PSM facilitates dimensionality reduction for matching, i.e., instead of matching on all covariates in $\vec{V}$ we can match on one-dimensional treatment probability $p_{\vec{V}} = \Pr(D=1 \mid \vec{V})$. So, the credit impact on a region of the yield distribution (say, lower decile) is estimated to be $F_1(\tau_{0.1} \mid D=1) - \int F_0(\tau_{0.1} \mid p_{\vec{V}}, D=0) f_{p_{\vec{V}}|D=1} dz$. Clearly, inverting constituent distributions in this impact estimator will provide credit-based conditional quantile changes.

With a theoretical grounding of PSM-based distributional impacts we specify a regression model to estimate conditional yield quantiles among *matched* borrowers and non-borrowers.

$$Q_\tau(Y_{i_k t_k} \mid \vec{X}_{i_k t_k}, D_{i_k t_k}) = \alpha_{0\tau} + \alpha_{1\tau} D_{i_k t_k} + \vec{X}'_{i_k t_k} \vec{\beta}_\tau ; Q_\tau(u_{i_k t_k} \mid \vec{X}_{i_k t_k}) = 0; \tau \in (0,1); k \in \{1,2,...,14\} \quad (3)$$

where $\tau \in (0,1)$, $u_{i_k t_k}$ is the random error, $\vec{X}_{i_k t_k} = \{\vec{X}'^{inp}_{i_k t_k}, \vec{X}'^{we}_{i_k t_k}, \vec{X}'^{land}_{i_k}, \vec{X}'^{farmer}_{i_k}\}$ and $k \in \{1,2,...,14\}$. Here $\vec{X}'^{inp}_{i_k t_k}$ accounts for input use including fertilizer (i.e., nitrogen and potassium[18]), irrigation, and high-yielding variety seeds which are expected to increase yields (Pray and Nagarajan 2009; Reyes-Cabrera et al. 2023). $\vec{X}'^{land}_{i_k}$ controls for household-specific land quality as measured by farmer reported soil quality and soil depth. Deeper, higher quality soils are associated with better yields as compared with shallow soils and soils with reported salinity, alkalinity or erosivity issues (Basak et al. 2021; Rajakaruna and Boyd 2019). $\vec{X}'^{farmer}_{i_k}$ includes farmer's age and education levels as a proxy for the household's farming experience which has an ambiguous relation with crop yields (Fuwa et al. 2007), and

[18] Phosphorus is also an important nutrient that is typically applied jointly with nitrogen. We observe a high degree of correlation (correlation coefficient = 0.56, p-value<0.001) between nitrogen and phosphorous application such that controlling for nitrogen effectively controls for phosphorus application as well.

$\vec{X}'^{we}_{i_k t_k}$ controls for season-wise village-level weather (i.e., temperature and rainfall)[19]. Please see Table 2.2 for descriptions and summary statistics for variables included in model (3).

2.3.2 Conceptual Framework: Characterising the impact of credit rationing on agricultural production with State-Contingent Technology

In this paper we employ a state-contingent production framework to examine how credit rationing influences input use decisions in crop production across borrower-groups whom the lenders may distinguish statistically through regional indices such as land quality, wealth status, or social indicators (e.g., gender, caste, etc.).

Previous studies employed a stochastic production technology to evaluate the impact of credit rationing on farmer welfare (e.g., Carter 1989, Feder et al. 1990, Ciaian et al. 2012). We aim to evaluate how rationing-led credit constraints may restrict farmers' adaptive capacity through input application decisions under production uncertainty. We argue that the prior approaches based on stochastic production function are limiting to analyse how rationing impacts farmer adaptation to cropping risks through input use decisions. The stochastic production technology is restrictive as it models output uncertainty through random 'disembodied' technical change (i.e., with exact same input-output relationship) across the nature states (Chambers and Quiggin 2000; Rasmussen 2003). On the other hand, the state-contingent approach specifies a flexible, state-specific input-output relationship which allows for ex-ante adaptation through input use that is informed by the marginal rate of output transformation (MRT) between nature states (Chambers and Quiggin 2000; Rasmussen 2003). Exogenous state probabilities in conjunction with the grower's risk attitude and preferences drive the MRT between nature states.

Through this framework we aim to quantify the extent to which credit rationing may be an impediment for production risk management leading to sub-optimal net returns driven by riskier state-contingent output among the farmers facing excess credit demand.

2.3.2.1 Articulating how credit constraint impacts production outcomes under state-contingent technology

In this section we will formalize how a rationing-led credit constraint inhibits farmers' decision-making and their ability to manage cropping risks. Let random states $\boldsymbol{\Omega} = \{1, 2, ..., M\}$ and associated probabilities $\Theta = \{\theta_1, \theta_2, ..., \theta_M\}$ specify the production

[19] The village-level weather information is constructed using gridded data provided by the Indian Meteorological Department (IMD). For details, please see SI Section 3(B).

uncertainty facing each farmer, where $\theta_m \in (0,1)$ denotes ex-ante probability of realizing end-of-season state of nature *m* such that $\sum_m \theta_m = 1$. Farmer *i* is assumed to borrow working capital *B* rupees per-acre in lieu of collateral (*c* rupees per-acre) and interest rate (*r*). Farmer decisions under uncertainty are examined by employing a state-contingent production technology that links ex-ante, *non-stochastic* input application to random output in each state, and subsequently net returns per-acre given market prices, denoted as $\vec{s} = \{s_m\}_{m=1}^{M}$. Borrower decisions do not alter θ_m, i.e., state *m* occurs exogenously, and we assume, without loss of generality, that higher values of index *m* correspond to greater net returns, i.e., $s_1 < s_2 < ... < s_M$.

Following Chambers and Quiggin (2000), we specify production technology as $T(\vec{y},\vec{x}) \leq 0$ that maps *ex-ante* input application $\vec{x} \in \mathbb{R}_+^d$ to state-contingent output vector $\vec{y} \in \mathbb{R}_+^M$ such that $y_m \leq q_m(\vec{x})$, $q'_{m,x_j} \equiv \partial q_m(\cdot)/\partial x_j > 0$ and $q''_{m,x_j} \equiv \partial^2 q_m(\cdot)/\partial x_j^2 < 0 \;\; \forall x_j \in \vec{x}$ and $m \in \Omega$. That is, each input exhbits positive marginal product across all states of nature, which decreases in the level of input applied. Further, we assume an 'output-cubical' technology such that state-contingent output, which is bounded by $q_m(\vec{x})$, does not directly affect (or depend upon) output in other states. Rasmussen (2004) clarifies that the output-cubical technology generally specifies production under risk when the same ex-ante input application leads to distinct end-of-season output through exogenous realization of *m*.

Denote $\vec{p} = (p_y, \vec{p}_x)$ comprising output (crop) price $p_y \in \mathbb{R}_{++}$ and an input price vector $\vec{p}_x \in \mathbb{R}_{++}^d$. We assume away price-risk, meaning that vector $\vec{p}$ remains the same across all nature states. Farmer revenue is given as $z_m = p_y y_m$ and net-revenue as $s_m = z_m - \sum_{j=1}^{d} p_j x_j$. To further this analysis we define a revenue-cost function based on Chambers and Quiggin (2000), which is defined as

$$C(\vec{p},\vec{z}) = \min\left\{\sum_{j=1}^{d} p_j x_j : y_m \leq q_m(\vec{x}) \text{ and } z_m \leq p_y y_m, m \in \Omega\right\} \tag{4}$$

if there exist input vectors that yield state-contingent revenue stream $\vec{z}$ through output array $\vec{y}$. If such an input vector does not exist then $C(\vec{p},\vec{z}) = \infty$. We further assume that $C(\vec{p},\vec{z}) \geq 0$ where equality holds *iff* $\vec{z} = 0$; $C(\vec{p},\vec{z})$ is linearly homogenous, non-decreasing and concave in $\vec{p}_x$; non-increasing in p_y; and increasing and convex in $\vec{z}$. Further, if $\vec{z}$ and

p_y are scaled uniformly then the revenue-cost function remains unchanged. Finally, $C(\vec{p},\vec{z})$ satisfies the Shepherd's Lemma (Chambers and Quiggin 2001) such that the optimal input demand under production uncertainty is given as $x_j^*(\vec{p}) = \partial C(\vec{p}, z^*(\vec{p}))/\partial p_j$ where $z^*(\vec{p})$ represents the producer's optimal (or utility-maximising) state-contingent revenue-vector.

Chambers and Quiggin (1999) show that the revenue-cost function is dual to the state-contingent production technology in terms of specifying the optimization problem of an expected utility maximiser (please see Corollary 2, pp. 293). Our interest in employing the optimal revenue-cost function to investigate farmer decisions is to be able to measure how a credit-constraints might impact the ex-ante input-mix employed by farmers through the Shepherd's Lemma, and hence their ability to mitigate cropping risk.

We now specify farmer preferences and the utility maximization problem to analyze the distinction between optimal state-contingent revenue streams with and without the credit constraint. Following Chambers and Quiggin (1999; 2000) we define farmer preferences over state-contingent net-revenues as a continuous and strictly increasing function $W:\mathbb{R}^M \to \mathbb{R}$, i.e., $W(s_1,s_2,...,s_M)$ such that $s_m = z_m - C(\vec{p},\vec{z})$ where revenue-cost $C(.)$ is dual to state-contingent production strucutre. $W(s_1,s_2,...,s_M)$ exhibits preferences of a risk-averse producer *with respect to subjective probabilities* $\Psi = \{\psi_1,..,\psi_M\}$ *if* $W(\bar{s}\vec{1}_M) \geq W(s_1,s_2,...,s_M)$ with $\bar{s} = \sum_m \psi_m s_m = \bar{z} - C(\vec{p},\bar{z}\vec{1}_M)$ where $\vec{1}_M$ is the *M*-dimensional unit vector and $W(.)$ is a generalized Schurr-concave function for Ψ. Subjective probabilities can be recovered from smoothly differentiable W as follows:

$$\psi_m(\vec{s}) = \frac{W_m(\bar{s}\vec{1})}{\sum_{t\in\Omega} W_t(\bar{s}\vec{1})} \text{ where } W_m = \frac{\partial W}{\partial s_m} > 0 \;\forall m \in \Omega \tag{5}$$

Trivially, $\sum_{m\in\Omega} \psi_m = 1$. Subjective probabilities $\{\psi_m\}$ are unique and proportional to the marginal rate of substitution between state-contingent incomes. As shown in equation (5), each ψ_m is given as the slope of the indifference curve where it intersects the equal-incomes vector. O'Donnell et al. (2010) refer to ψ_m as 'decision-weights of a risk-neutral farmer' who would select the same production plan as risk-averse farmer with preferences $W(.)$. Menapace et al. (2013) established that risk-averse farmers will overweigh the odds of the ‘bad’ states relative to the ‘good’ states when compared to risk-neutral (or less risk-averse) farmers. That is, $\psi_m > (<) \theta_m$ whenever $m < (>) m^*$ (defined in the previous section).

Farmers’ utility maximization problem without a credit constraint:

$\max_{\vec{z}} L = W(s_1, s_2, ..., s_M)$ such that $s_m = z_m - C(\vec{p}, \vec{z})\ \forall m \in \Omega$. That is, farmer is concerned with the choice of state-contingent reveue stream, $\vec{z}$, and assuming an interior solution, we can write the first-order conditions (F.O.C.) of optimization as follows:

$$W_m - C_m(\vec{p}, \vec{z}) \times \sum_{t \in \Omega} W_t \leq 0\ \forall \mathrm{m} \in \Omega \qquad (6).$$

Equation (6) indicates that under Schur-concave preferences the marginal cost of producing state-contingent revenue in state *m* (with fixed p_y) is always as large as the marginal utility of that state divided by marginal utility of a unit increase in revenue in each state (Chambers and Quiggin 1997)[20]. Equation (6) further simplifies by using (5):

$$\psi_m - C_m(\vec{p}, \vec{z}) \leq 0\ \forall \mathrm{m} \in \Omega \qquad (7)$$

which, upon summing the *M* F.O.C.s in (7), obtains efficient revenue-set $\mathrm{Z}(\vec{p}) = \{\vec{z} : \sum_{t \in \Omega} C_t(\vec{p}, \vec{z}) \geq 1\}$ where marginal cost of a sure unit-increase of revenue to be greater than one. The boundary of $\mathrm{Z}(\vec{p}) \equiv \{\vec{z}^* : \sum_{t \in \Omega} C_t(\vec{p}, \vec{z}^*(\vec{p})) = 1\}$ is denoted as efficient frontier.

Under a credit constraint the farmer solves $\max_{\vec{z}, \delta} L = W(s_1, ..., s_M) + \delta(B - C(\vec{p}, \vec{z}))$ where δ is the Lagrange multiplier and $B < C(\vec{p}, \vec{z}^*(\vec{p}))$. The F.O.Cs. are now written as

$$W_m - C_m(\vec{p}, \vec{z}) \times \left(\delta + \sum_{t \in \Omega} W_t\right) \leq 0\ \forall \mathrm{m} \in \Omega, \text{ and}$$

$$B - C(\vec{p}, \vec{z}) = 0 \qquad (8)$$

For state-pairs (t, m) the following condition holds irrespective of the credit constraint:

$$\frac{W_t}{W_m} = \frac{C_t}{C_m}\ \forall (t, m) \in \Omega^2, t \neq m. \qquad (9)$$

System of $M - 1$ equations in (9) in conjunction with credit constrained revenue-cost condition $C(\vec{p}, \vec{z}) = B$ yields optimal revenue-vector $\vec{z}'(\vec{p}, B)$ and langrage multiplier $\delta'(\vec{p}, B)$>0. We now clarify how optimal revenues compare under constrained and

[20]Total differentiation of the multivariate function $W(\vec{s})$ is given as $dW = \sum_{m \in \Omega} \left(\partial W / \partial s_m\right) ds_m$

unconstrained credit access, i.e., $\vec{z}'(\vec{p},B)$ and $\vec{z}^*(\vec{p})$, respectively, in terms of change in farmer welfare and risk exposure.

Result 1 *(Credit constraint and farmer welfare)* Given the state-contingent revenue-cost function $C(\vec{p},\vec{z})$ and farmer preferences $W(.)$ we find that marginally higher credit access is welfare-increasing for a credit-constrained farmer.

By the Envelope Theorem and a farmer's value function under credit constraint defined as $V(\vec{p},K) \triangleq W(s'_1,...,s'_M)$ such that $\vec{s}'_m = \vec{z}'_m(\vec{p},B) - C(\vec{p},\vec{z}'(\vec{p},B))\ \forall m \in \Omega$, we can write that $dV/dB = \partial L/\partial B\big|_{\vec{z}=\vec{z}'(\vec{p},K)} = \delta'(\vec{p},B) > 0$, meaning that marginally relaxing the credit constraint improves farmer welfare (Carter 2001). Wolfsetter (1970, p. 141) showed through their Lemma 4.2 and Proposition 4.4 that $W(\vec{s}^*) \geq W(\vec{s}')$ implies (and is implied by) $\vec{s}^*$ second order stochastically dominates $\vec{s}'$ [21]. Below we formalize the risk-properties of $\vec{s}^*$ and $\vec{s}'$ through this stochastic ordering concept.

Result 2: (Credit constraint leads to riskier net returns) Given $C(\vec{p},\vec{z})$, $W(.)$ and a common support for $\vec{s}^*$ and $\vec{s}'$, we can infer that $\vec{s}'$ is riskier than $\vec{s}^*$ in the second order stochastic dominance (S.O.S.D.) sense, i.e., $\int_0^{\hat{m}\leq M} F_{\vec{s}^*}(m)dm \leq \int_0^{\hat{m}\leq M} F_{\vec{s}'}(m)dm$ where $F_{\vec{s}}(m)$ denotes the cumulative distribution function over net returns ordered across states $m \in \Omega$. This implies,

(A) $E_{\psi}(\vec{s}') \leq E_{\psi}(\vec{s}^*)$, and

(B) $\exists\, \vec{\tilde{s}} = \vec{s}^* - [E_{\psi}(\vec{s}^*) - E_{\psi}(\vec{s}')]$ of which $\vec{s}'$ is a mean-preserving spread. That is, we say $\vec{s}' = \vec{\tilde{s}} + \vec{\varepsilon}$ where random vector $\vec{\varepsilon}$ has zero mean and finite variance[22].

We will now assess how farmer's adjust input use in response to constrained credit access by employing the Shepherd's Lemma. Specifically, we can express optimal demand for input $x_j \in \vec{x}$ as $x_j(\vec{p},\vec{z}^*) = \partial C(\vec{p},\vec{z}^*)/\partial p_j$ (Chambers and Quiggin 2001). Given prices we can decompose input shift associated with change in optimal revenue from $\vec{z}^*$ to $\vec{z}'$, i.e., under unconstrained to constrained credit access, respectively, as follows:

[21] Notice that state-contingent net returns vector, $\vec{s}$, represents a random variable with potential realizations contained in the vector with assigned subjective probabilities in $\vec{\psi}$.

[22] Given $\vec{\tilde{s}} = \vec{s}^* - \vec{c}$ (with $\vec{c}$ being a constant vector) it is trivial to show second (and higher) central moment equalities between $\vec{s}^*$ and $\vec{\tilde{s}}$.

$$\Delta x_j^{\text{total}} = x_j(\vec{p},\vec{z}') - x_j(\vec{p},\vec{z}^*) = \underbrace{(x_j(\vec{p},\vec{\tilde{z}}) - x_j(\vec{p},\vec{z}^*))}_{\Delta x_j^{\text{contraction}}} + \underbrace{(x_j(\vec{p},\vec{z}') - x_j(\vec{p},\vec{\tilde{z}}))}_{\Delta x_j^{\text{risk}}} \tag{10}$$

where $\vec{\tilde{z}}$ is defined as an intermediate revenue vector having $E_\pi(\vec{z}') = E_\pi(\vec{\tilde{z}}) \equiv \mu' \leq \mu^* \equiv E_\pi(\vec{z}^*)$ and the same revenue-mix as $\vec{z}^*$ (meaning $z_s^*/z_t^* = \tilde{z}_s/\tilde{z}_t\ \forall(s,t) \in \Omega^2, s \neq t$). So the shift from $\vec{z}^*$ to $\vec{\tilde{z}}$ corresponds to a *pure-mean contraction effect* that lowers the expected revenue holding the risk-properties (i.e., higher-order moments of $\vec{z}^*$ and $\vec{\tilde{z}}$) constant (Scott and Horvath 1980)[23]. On the other hand, the movement from $\vec{\tilde{z}}$ to $\vec{z}'$ corresponds to a *pure-risk effect* through risk-adjustment at lower expected revenue (due to constrained credit access).

Holding risk-properties constant, state-wise reduction in revenue under the contraction effect is such that $E_\pi(\vec{z}^*) \geq E_\pi(\vec{\tilde{z}})$. This implies $C(\vec{p},\vec{z}^*) \geq C(\vec{p},\vec{\tilde{z}})$ and given prices and technology, $x_j(\vec{p},\vec{z}^*) \geq x_j(\vec{p},\vec{\tilde{z}})$ i.e., $\Delta x_j^{\text{contraction}} \leq 0\ \forall j$. Next with regards to the pure-risk effect, the sign of the input shift associated with a movement towards a riskier revenue vector will depend on the risk-property of each input. Chambers and Quiggin (2001) define inputs as either being *risk-complements* or *risk-substitutes*. In particular, $x_j(\vec{p},\vec{\tilde{z}})$ is a risk-complementary input if $x_j(\vec{p},\vec{\tilde{z}}) \leq x_j(\vec{p},\vec{z}')$ i.e., *more* x_j will be used when producing a *riskier* revenue vector ($\Delta x_j^{\text{risk}} \geq 0$). Similarly $x_j(\vec{p},\vec{\tilde{z}})$ is said to be a risk-substituting input if $x_j(\vec{p},\vec{\tilde{z}}) \geq x_j(\vec{p},\vec{z}')$ i.e., *more* x_j will be used when producing a *less-risky* revenue vector ($\Delta x_j^{\text{risk}} \leq 0$)[24].

---

[23] Note that $z_s^*/z_t^* = \tilde{z}_s/\tilde{z}_t\ \forall(s,t) \in \Omega^2, s \neq t$ represent ${}^{M}C_2$ conditions implying that all standardized moments of $\vec{z}^*$ and $\vec{\tilde{z}}$ representing the shape and risk-properties (i.e., dispersion relative to the mean) of the associated revenue distribution(s) are equal. To see this, fix $t = 1$ and consider a subset of the conditions for $s \in \{2,3,...,M-1\}$.
$z_s^*/z_1^* = \tilde{z}_s/\tilde{z}_1 \Rightarrow z_s^* = \left(z_1^*/\tilde{z}_1\right) \times \tilde{z}_s = \omega\tilde{z}_s$ where $\omega = \left(z_1^*/\tilde{z}_1\right)$. Trivially, the subset of conditions ($z_m^* = \omega\tilde{z}_m \forall m \in \Omega$) imply all others. The γ^{th} standardized moment of $\vec{z}^*$ is $m_\gamma(\vec{z}^*) = E(z_m^* - \mu^*)^\gamma/(\sigma^*)^\gamma$ where $\sigma^* = \sqrt{E(z_m^* - \mu^*)^2}$ is standard deviation of $\vec{z}^*$.
Substituting $z_m^* = \omega\tilde{z}_m$, recognizing $\mu^* = \omega E(\vec{\tilde{z}}) = \omega\mu'$ and $\sigma^* = \omega\tilde{\sigma}$ gives:
$m_\gamma(\vec{z}^*) = E(\omega\tilde{z} - \omega\mu')^\gamma/\omega^\gamma \times \tilde{\sigma}^\gamma = E(\tilde{z} - \mu')^\gamma/\tilde{\sigma}^\gamma = m_\gamma(\vec{\tilde{z}})\ \forall\gamma > 2$.

[24] This definition yields a characterization of input risk properties in terms of partial derivatives of input-demand functions: "an input is a risk-substitute if its responsiveness to

Result 3: (<u>*Credit constraint and input use*</u>): Given $C(\vec{p},\vec{z})$ and *W*(.), the total shift in input demand depends on relative size of $\Delta x_j^{\text{contraction}}$ and Δx_j^{risk} i.e.,

(1) $\Delta x_j^{\text{total}} \le 0$ when the input is a risk-substitute i.e., $\Delta x_j^{\text{contraction}} \le 0$ and $\Delta x_j^{\text{risk}} \le 0$.

(2) $\Delta x_j^{\text{total}} \le 0$ when the input is a risk-complement and the contraction effect is stronger i.e., $\underbrace{\Delta x_j^{\text{contraction}}}_{\le 0} \ge \underset{\ge 0}{\Delta x_j^{\text{risk}}}$.

(3) $\Delta x_j^{\text{total}} \ge 0$ when the input is a risk-complement and the risk effect is stronger i.e., $\underbrace{\Delta x_j^{\text{contraction}}}_{\le 0} \le \underset{\ge 0}{\Delta x_j^{\text{risk}}}$.

Our results show that a binding credit constraint is welfare-reducing and limits the farmer to a riskier revenue distribution and sub-optimal input application. However sub-optimal input levels are not necessarily lower than the optimal input application and the total effect on input demand depends on the risk-properties of inputs. Below we will provide an illustration for the two-dimensional case (i.e., $\Omega = \{1,2\}$). We also showcase the sequence of results in our applied setting through a numerical example (please see SI Section 6 (S6)).

<u>*2.3.2.2 Illustration of credit constraint impacts on production outcomes*</u>

Figure 2.4 (panel (A)) presents for *M* = 2 where 1 □'good state' and 2 □'bad state', a comparison between optimal revenue-vector for a risk-neutral farmer, (point *N*), a risk-averse farmer, (point *A*); and a credit constrained risk-averse farmer (point *B*). Following subjective probability definition in equation (2) we denote the tangent to risk-averter's indifference curve at the bisector as the *fair-odds line* (Chambers and Quiggin 2001) which has the slope equal to *negative* ratio of state-specific subjective probabilities. Movement along the fair-odds line (keeping expected returns fixed) and away from bisector (equal returns in each state) corresponds to a 'riskier' shift (i.e., higher $|(z_1/z_2)-1|$). On the other hand, moving radially outward along a ray from the origin (*revenue-mix ray*) keeps riskiness ($|(z_1/z_2)-1|$) constant while *increasing* the expected revenue.

---

state-contingent revenue variation is large and positive for the lowest revenue states and small, and possibly negative, for the highest income states i.e., $\sum_m \left(\frac{\partial x_d(\vec{p},\vec{z},\gamma)}{\partial z_m}\right)\left(z_m - \sum_m \pi_m z_m\right) \le 0$ with a reversed inequality for risk-complements." (Chambers and Quiggin 2001, p. 27).

Pictorially, at risk-averter's optimal, the fair-odds line cuts the isocost frontier from below, while at risk-neutral farmer's optimal, the slope of fair-odds line is exactly equal to the slope of isocost curve. Mathematically, this implies that the risk-neutral farmer's optimal occurs where marginal revenue of radial expansion is zero. In contrast at point *A* the expected returns are lower than at *N* but the riskiness is also lower i.e., the risk-averse farmer trades-off higher expected returns (or scale) at point *N* in lieu of a lower-risk (or smoother) bundle and stops at a point where a small radial expansion of state-contingent revenues would increase their expected returns. Under a binding credit constraint, the risk-averse farmer's optimal choice implies that the farmer trades off lower expected returns (under a binding constraint) for a relatively higher level of risk and hence point *B* lies on a *lower fair odds line and a revenue-mix ray further away from the bisector* relative to point *A*. Consider point *C* on the same revenue-mix ray as point *A and* on the fair-odds line containing point *B*. The shift from point *A* to *B* can be decomposed into a pure-mean contraction effect (i.e., from *A* to *C* along the same revenue-mix ray) and a pure-risk effect (i.e., *C* to *B* along the same fair-odds line) (see Panel (B), Figure (4)).

In summary, rural credit markets are characterized equilibrium rationing which, at the household-level, would cause constrained credit access, and hence lower farmer welfare, riskier returns and sub-optimal input use (Results 1-3). An important nuance from our analysis is that while a binding credit constraint consistently reduces the use of risk-substituting inputs, it may simultaneously increase the use of risk-complementing inputs.

2.3.3 Estimating ex-ante odds of credit rationing via heterogeneous credit-yield linkage

Here we outline an analytical strategy to employ the conditional quantile functions (as estimated through Section 2.3.1) to evaluate credit rationing odds among borrowing households in our sample. Two existing methods to elicit credit rationing include: (1) "direct elicitation" by employing end-of-season primary surveys to inquire upon farmers' unmet credit demand (Feder et al. 1990; Boucher et al. 2009); and (2) "indirect elicitation" by comparing *estimated* shadow price of credit with interest costs facing an average farmer (Sial and Carter 1996). We advance this strand of literature by using 'estimated' conditional yield quantile functions to evaluate *ex-ante* odds of rationing for sample households. The sample-wide rationing incidence reported by Sial and Carter (1996) may be inferred as a central tendency statistic from our household-specific odds of credit rationing.

Consider lender's participation constraint or LPC, i.e., $E_{m\in\Omega}(\rho) \geq B(1+r)$ that informs credit allocation. Given $\rho(z_m; B, r, c) = \min\{z_m + c,\ B(1+r)\}$, LPC states that $\sum_{m\leq m^*} \theta_m (z_m + c) + \tilde{\theta}_m(m^*) B(1+r) \geq B\left(1+r\right)$ where m^* was earlier defined such that $z_m \geq B(1+r)\ \forall m \geq m^*$ and $\tilde{\theta}_m(m^*) = \Pr(m \geq m^*)$. We re-write $E_m(\rho)$ in expected-revenue terms:

$$E_m(z) - E_{m>m^*}(z) \geq \Pr(m < m^*) \times (B(1+r) - c) \tag{11}$$

where $E_{m>m^*}(z) = \sum_{m>m^*} \theta_m z_m$. Notice that equation (9) implies a threshold-criteria for state-contingent revenues to exceed lender's opportunity cost of capital given $z_{m>m^*} > B(1+r)$, which yields a strict inequality $E_m(z) + [1 - \tilde{\theta}_m(m^*)]c > B(1+r)$. Further, the lender may not directly observe abstract nature-states in Ω in evaluating $E_m(\cdot)$ but construct an expectation, $E_{\tilde{X}}(z)$, conditional on population (e.g., district-wise) covariates ($\tilde{X}$) of uncertainty in farm-level productivity. In particular, the lender may assess credit allocation depending on crop-type, wealth and property rights, social networks (e.g., gender/caste), and/or agro-ecology (e.g., soil/weather)[25].

We now evaluate the nature state(s) where which credit-led revenue premium (estimated through equation (3) will be large enough to exceed liability $1+r$ (given crop price) by re-express $E_m(z)$ in terms of quantile function $Q_\tau(z) = \inf\{z : F(z) \geq \tau\}$. In line with the arguments presented in Sial and Carter (1996) if credit-led revenue premium were to exceed $1+r$ prior to the topmost yield quantile, the farmer will be deemed 'credit rationed'.

It is useful to define a Riemann-Stieltjes sum over L partitions of the revenue interval $[z_1, z_M]$: $S(L, F_{\tilde{X}}(z), z_{l+1}) = \sum_{l=1}^{L-1} z_{l+1}(F_{\tilde{X}}(z_{l+1}) - F_{\tilde{X}}(z_l))$, evaluated at right-boundary of each interval $z_{l+1} - z_l = (z_M - z_1)/L$[26]. As $L \to \infty$ or the partition mesh becomes finer, i.e.,

[25]Section 7 (S7) in the SI provides a sample agricultural loan application form from State Bank of India (link) that is shown to collect information on loan purpose, farmer income, assets, liabilities and demography (e.g., caste) alongside farming experience, land size, cropping pattern access to irrigation and markets. Lenders also match farm location to public land records repository (a sample record shown here. Last accessed Oct. 11, 2025) which provides information on land value and ownership structure. Spatial location of farms can be matched to local (e.g., village-level) weather and soils data from auxiliary sources.

[26]Equivalently, one might consider partitioning the state-space 'bad state' $\equiv m_1 < m_2 < \ldots < m_L \equiv$ 'good state'. Either way, more importantly, we note that partitioning on

$\max_l(z_{l+1}-z_l) \to 0$, the Riemann-Stieltjes sum approximates the conditional expectation of revenue, meaning that $\lim_{L\to\infty} S(L, F_{\tilde{X}}(z), z_{l+1}) = E_{\tilde{X}}(z) = \int_z z dF_{\tilde{X}}(z)$ (Resnick 1999). We can express $S(L, F_{\tilde{X}}(z), z_{l+1})$ in terms of the quantile function. For L partitions on the quantile-space $\tau \in [0,1]$ such that $\tau_l = F_{\tilde{X}}(z_l)$, $\tau_1 = F_{\tilde{X}}(z_1) = 0$ and $\tau_L = F_{\tilde{X}}(z_L) = 1$, we have $S(L, F_{\tilde{X}}(z), z_{l+1}) \equiv S(L, F_{\tilde{X}}^{-1}(\tau), \tau_{l+1}) = \sum_{l=1}^{L-1} F_{\tilde{X}}^{-1}(\tau_{l+1})(\tau_{l+1}-\tau_l)$. As $L \to \infty$, $\lim_{L\to\infty} S(L, F_{\tilde{X}}^{-1}(\tau), \tau_{l+1}) = \int_0^1 Q_{\tilde{X}}(\tau)d\tau$. So,

$$E_{\tilde{X}}(z) \simeq \sum_{l=1}^{L-1} F_{\tilde{X}}^{-1}(\tau_{l+1})(\tau_{l+1}-\tau_l) = \int_0^1 Q_{\tilde{X}}(\tau)d\tau \tag{12}$$

Therefore, the LPC in equation (11) can be written as function of the conditional quantiles where $\tau_{m^*} = F_{Z|\tilde{X}}^{-1}(B(1+r))$:

$$\int_0^{\tau_{m^*}} Q_{\tilde{X}}(\tau)d\tau \geq \Pr(m < m^*) \times [B(1+r) - c] \tag{13}$$

We now link the credit-led premium specific to each yield quantile, i.e., $\hat{\alpha}_{1\tau}$ estimated from equation (3), to credit allocation decisions governed by the LPC in equation (13). The idea is to assess revenue premium from marginal credit for a representative borrower against the interest cost. In effect we differentiate both sides of inequality (13) with respect to $dB > 0$ with a note that $\partial m^*/\partial B \neq 0$ given Result#2 (S.O.S.D.: $F_{\tilde{X}}(z(B)) \geq F_{\tilde{X}}(z(B+dB))\ \forall z$). We evaluate credit-led shift in LPC by using the Leibniz rule of differentiation, that is[27]:

$$\int_0^{\tau_{m^*}} \frac{\partial}{\partial B}\left(Q_{\tilde{X}}(\tau \mid B)\right)d\tau \geq \int_0^{\tau_{m^*}} (1+r)d\tau \tag{14}$$

Given $\tau_{m^*} < 1$, we have

$$\int_0^1 \frac{\partial}{\partial B}\left(Q_{\tilde{X}}(\tau \mid B)\right)d\tau > 1+r \tag{15}$$

---

the revenue-space is a mathematical device and distinct from exogenous state-space Ω. For example, one can construct the many partitions L even when $M = 2$.

[27] $\frac{\partial}{\partial B}\left(\int_0^{\tau_{m^*}}\left(Q_{\tilde{X}}(\tau) - B(1+r) + c\right)d\tau\right) = \int_0^{\tau_{m^*}}\left(\frac{\partial Q_{\tilde{X}}(\tau)}{\partial B} - (1+r)\right)d\tau + \left(Q_{\tilde{X}}(\tau_{m^*})c - B(1+r)\right)\frac{d\tau_{m^*}}{dB}$; where $Q_{\tilde{X}}(\tau_{m*}) = B(1+r)$ or $\tau_{m^*} = F_{Z|\tilde{X}}^{-1}(B(1+r))$.

Equation (15) asserts that, given *r* and $\tilde{X}$, a decision to lend *B* rupees is guided by the rule that credit-led premium up to yield quantile $\tau_{m^*} \in (0,1)$ will cumulatively exceed $1+r$.

We are aware that matrix $\tilde{X}$ is lender's annotation of production technology afforded by representative borrower or borrower-group in view of an inability to fully observe farmer attributes $\vec{X}_i$ in equation (3): input use, land quality, farmer demography and local weather. It is quite likely that $\tilde{X} \neq \vec{X}_i$, which clarifies that information asymmetry operates through a coarse representation of borrowers' repayment ability. The idea is that $\tilde{X}$ is likely a noisy proxy for household-level yield covariates, $\vec{X}_i$, rationalizes credit rationing such that aggregate population metrics (equation (15)) may lead to heterogeneous credit access among borrowers with similar productivity who belong to different groups. We now contrast credit-led revenue premium for *farmer i* (instead of a population group annotated by $\tilde{X}$) against the marginal cost 1 + *r*:

$$\int_0^1 \hat{\alpha}_{1\tau}(X_i \mid B)d\tau > 1+r \tag{16}$$

Inequality (16) suggests that there exists revenue-level $z_i^* = F_{\vec{X}_i}^{-1}(\tau_i^*)$ up to which quantiles-based credit-led revenue premium will cumulatively exceed repayment liability. In other words, farmer *i* will be credit rationed upon realizing end-of-season revenues greater than $F_{Z|X_i}^{-1}(\tau_i^*)$, i.e., with probability $1-\tau_i^*$, with equation (16) implying $\int_0^1 \hat{\alpha}_{1\tau}(X_i \mid B)d\tau =$ $\int_0^{\tau_i^*} \hat{\alpha}_{1\tau}(X_i \mid B)d\tau + \int_{\tau_i^*}^1 \hat{\alpha}_{1\tau}(X_i \mid B)d\tau$ with $\int_0^{\tau_i^*} \hat{\alpha}_{1\tau}(X_i \mid B)d\tau = 1+r$. Hence, our next result:

Result 4 (*Credit rationing odds*): A conditional quantile-level τ_i^* fully characterizes the ex-ante likelihood of credit rationing for farm household *i* through the conditional state-contingent revenue levels exceeding $F_{\vec{X}_i}^{-1}(\tau_i^*)$. That is,

i. If $\tau_i^* \in [0,1)$, then borrower i faces ex-ante odds of credit rationing equal to $1-\tau_i^*$.

ii. If $\tau_i^* \to 1$, then we can infer that borrower i is not likely to be credit rationed.

We posit $1-\tau_i^*$ as a metric for the degree of credit rationing faced by household *i* as τ_i^* reflects the extent of mismatch between the lender's group-based assessment of a borrower's repayment ability (based on $\tilde{X}$) and the borrower's actual repayment capacity (based on $\vec{X}_i$). In effect, farmers having similar *productive* attributes may receive distinct loan contracts by virtue of their group affiliation conceptualized by the lender through imperfect or noisy proxy for productivity $\tilde{X}$. Carter (1988) characterized rationing as 'statistical discrimination' which plays out through unequal credit allocation across farmer-groups. Below we will characterize ex-ante odds of rationing (hence statistical discrimination) based on sample estimate of τ_i^*. SI Section 8 (S8) details the steps for computing τ_i^*.

<u>*2.3.3.1 Identifying the channels of statistical discrimination by totally differentiating the LPC*</u>

We re-write lender's expected profits as $E\rho = B(1+r)(1-\tau_{m^*}) + c\tau_{m^*} + \int_0^{\tau_{m^*}} Q_{Z|\tilde{X}}(\tau)d\tau$ where $\tau_{m^*} = F_{Z|\tilde{X}}^{-1}(B(1+r))$. Totally differentiating $E\rho$ with respect to τ, *B*, *r* and *c*:

$$dE\rho = \int_0^{\tau_{m^*}} \left(\frac{\partial Q(\tau)}{\partial B} dB + \frac{\partial Q(\tau)}{\partial \tau} d\tau \right) d\tau + \left[Q(\tau_{m^*}) - \left(B(1+r) - c \right) \right] d\tau_{m^*} +$$
$$(1+r)(1-\tau_{m^*})dB + B(1-\tau_{m^*})dr + \tau_{m^*}dc.$$

Given that $Q(\tau_{m^*}) = B(1+r)$; $\partial Q(\tau)/\partial B = \hat{\alpha}_{1\tau}$ (equation 3); and $d\tau_{m^*} = (\partial\tau_{m^*}/\partial B)dB + (\partial\tau_{m^*}/\partial r)dr$, we will now investigate statistical differentiation in credit allocation through the trade-off between $dB \neq 0$ and distinct farmer-groups annotated by $d\tau$ while keeping $E\rho$ unchanged. We set $dr = 0$ to recognize market-driven and regulatory interest rate cap in developing countries (see Section 2) and $dc = 0$ to denote the reality of undercollateralized loans with a limited scope of demanding higher collateral in lieu of output risk (Carter 1988). That is,

$$dE\rho = \underbrace{\int_0^{\tau_{m^*}} \left(\alpha_{1\tau} dB + \frac{\partial Q_{Z|\tilde{X}}(\tau)}{\partial \tau} d\tau \right) d\tau}_{\text{TI}} + \underbrace{(1+r)(1-\tau_{m^*})dB}_{\text{T2}} + cd\tau_{m^*}. \quad (17)$$

Equation (17) quantifies the contribution of two additive components in determining the expected lender profits: (1) term T1 represents lender payoff derived from lower quantiles of state-contingent revenues such that $\tau \in [0, \tau_{m^*})$ where $z_m + c < B(1+r)$; and (2) term T2 represents lender payoff from the higher revenue quantiles where $z_m < B(1+r)$. A closer look at these two additive components reveals concavity of lender profits whereby lender profits increase proportional to the borrower revenue function in the non-repayment region (T1) but remain constant once borrower revenue crosses over to the repayment region (T2). Therefore, as *B* increases, the credit-led quantile-based yield/revenue impacts positively affect expected lender returns in T1 but not in T2 (where marginal lender profits do not change with higher borrower revenues). Finally, we assume $d\tau_{m^*} = 0$, meaning that the marginal shift in revenue distribution does not shift the cut-off state m^* which we set as exogenously set by nature (see Section 2).

To characterize statistical discrimination we assess TI in equation (17) which reveals a trade-off between marginal credit allocation and a marginal shift in revenue schedule. Figure 2.5 shows that when $d\tau < 0$ we have $\partial Q_{Z|\tilde{X}}(\tau) \big/ \partial\tau \, d\tau < 0 \;\forall \tau$ (or inferior revenue schedule) which may be compensated through $dB > 0$, however the extent of this trade-off will depend on the size of $\alpha_{1\tau}$. From the lender's perspective, we have learnt that farmers may be classified into 'superior' and 'inferior' groups through covariates $\ell^{\sup}$ and $\ell^{\inf}$ which represent a subset of farmer characteristics ($\tilde{X}$) that represent borrower's repayment potential but are noisy proxies for farm-level productivity (e.g., wealth or land size), such that $\alpha_{1\tau,\ell^{\inf}} \le \alpha_{1\tau,\ell^{\sup}}$ (which leads to a testable empirical hypothesis below). Overall, we find, given *r* and *c*, that the lender will offer lower credit allocation to inferior groups relative to their 'superior' counterparts in achieving the same level of expected profits (i.e., $dE\rho$).

In our sample, we expect the 'offset effect' of *dB* to be higher among some farmer-groups relative to others; such as larger vs. smaller farmers (Carter 1988); landowners vs. renters (Regassa et al. 2023); asset-rich vs. asset-poor farmers (Luan and Bauer 2016). Below we will specify empirical hypotheses to test non-uniformity in credit allocation and odds of rationing, along with relevant production impacts, from our sample.

2.3.4 Empirical Hypotheses

Hypothesis 1 (*Skewed credit allocation across farmer-groups*): Given *r* and $\tilde{X}^{\text{productivity}}$, credit allocation per acre (*B*) is on-average higher for 'superior' farm groups relative to 'inferior' farm groups.

$H_{0|\tilde{r}_i,\tilde{X}^{\text{productivity}}} : B^{\sup} = B^{\inf}$

$H_{A|\tilde{r}_i,\tilde{X}^{\text{productivity}}} : B^{\sup} > B^{\inf}$

Here $\tilde{r}_i \in \{\tilde{r}_i^{BPLR}, \tilde{r}_i^{ISS}\}$ where BPLR denotes benchmark prime lending rate regime applicable prior to 2006-07 and ISS denotes interest subvention regime applicable post 2007, $\tilde{r}_i^{BPLR} = \{r_i : r_i \in [12,15] \,\&\, t \le 7\}$ and $\tilde{r}_i^{ISS} = \{r_i : r_i \in [6,9] \,\&\, t > 7\}$. $\tilde{X}^{\text{productivity}}$ consists of borrower characteristics observable to the lender that proxy agricultural productivity (for e.g., irrigation access or regional historical yields). $B^{\sup}$ ($B^{\inf}$) denotes average loan amount for $\ell^{\sup}$ (asset-rich) and $\ell^{\inf}$ (asset-poor) farmer groups respectively.

Hypothesis 1A (*Skewed credit rationing odds across farmer-groups*): Given *r* and $\tilde{X}^{\text{productivity}}$, odds of being rationed are on-average lower for 'superior' farm groups relative to 'inferior' farm groups.

$$H_{0|\tilde{r}_i,\tilde{X}^{\text{productivity}}} : \tau^{*\sup} = \tau^{*\inf}$$

$$H_{A|\tilde{r}_i,\tilde{X}^{\text{productivity}}} : (1-\tau^{*\sup}) < (1-\tau^{*\inf})$$

Here $(1-\tau^{*\sup})$ ($(1-\tau^{*\inf})$) denote average odds of rationing for $\ell^{\sup}$ (asset-rich) and $\ell^{\inf}$ (asset-poor) farmer groups respectively.

We document statistical discrimination by summarizing loan amounts per acre and odds of rationing for richer versus poorer farmers by indicators of agricultural productivity observable to the lender: irrigation access, regional rainfall pattern and regional agricultural productivity, separately at the regulated interest rate boundary for each interest rate regime in our sample. Patterns revealed by a statistical comparison of average loan amounts and extent of rationing *within* observable productivity categories *across* farmers' wealth groups can inform on statistical discrimination or lack thereof. However, our analysis is challenged by low statistical power due to the small sample size which also limits our attempts at more granular analysis.

Hypothesis 2 *(Asymmetric input response to credit rationing)*: Rationed farmers facing a binding credit constraint reduce the use of risk-substituting inputs, while the effect on risk-complementing inputs is indeterminate.

To test this, we utilize a t-test to compare the average input application levels across rationed $(x_{j,1} \equiv E(x_{ji} \mid 1-\tau_i^* > 0))$ and non-rationed $(x_{j,0} \equiv E(x_{ji} \mid 1-\tau_i^* \approx 0))$ households. Mathematically, $H_0^3 : x_{j,1} = x_{j,0}$ and $H_A^{3,\text{risk-substitute}} : x_{j,1} < x_{j,0}$ while $H_A^{3,\text{risk-complement}} : x_{j,1} \neq x_{j,0}$.

2.4 Results and Discussion:

Aside: Revisions Summary (Based on Examination Committee Comments)

Based on the comments received, we have revised the quasi-experimental design from a single-crop analysis to a cropping-systems based framework that retains crop-specificity by crop categories and better reflects the diversified multiple-cropping systems prevalent in semi-arid and tropical regions. Unlike conventional credit impact studies that focus on aggregate farm output and thereby mask heterogeneity across cropping systems, the revised framework incorporates cropping-system heterogeneity into the matching procedure, as well as the analysis of the incidence and extent of credit rationing among semi-arid Indian farmers.

This revision involved classifying 47406 distinct cropping instances involving 127 distinct crops into 31 distinct multiple cropping systems arising from five crop categories: Millets, Cereals, Pulses, Fruits and Vegetables and Commercial crops (e.g., cotton and oil seeds[28]) practiced by 1438 unique households over 14 years. Unsurprisingly, millets-based cropping systems (which are dominated by sorghum-based cropping systems) emerged among the top cropping systems (by area) accounting for over 44 percent of area cultivated across the years in the VDSA dataset. Among millets, Sorghum is the largest crop category accounting for ~92% percent of cultivated area (see Table 2, SI Section 10 (I)).

Based on the frequency of these cropping systems and counterfactual availability within these cropping system categories, we seek out a quasi-experimental design for studying credit impacts for all 'major' crops. Major crops are defined as crops that account for at least 5 percent of cropping area across the years. Based on this definition, we obtain 9 major crops (including Sorghum) which together account for more than 80% of cultivated area across the years. Of these, based on counterfactual availability, we rule out the possibility of studying credit impacts for all crops except Sorghum and Pigeonpea. While Pigeonpea is a major crop (accounting for 8% of cultivated area over the years), Pigeonpea is not a major crop within a household's yearly cropping portfolio. Specifically, the average Pigeonpea cultivating household is not a principal cultivator of Pigeonpea and in fact allocates <20 percent of cultivated area in a season-year to Pigeonpea. In contrast, cultivators of sorghum are also (typically) principal cultivators of Sorghum (i.e., allocate at least 50 percent or more area to sorghum in the season-year they cultivate it). This assessment (outlined in SI Section 10 (I)) rules out the possibility of studying credit impacts for all crops except Sorghum. For Sorghum, we classify the 12025 distinct cropping instances (i.e., 8512 unique HH-Crop-Year IDs) associated with 520 sorghum growing farmers across 14 years into three distinct cropping system categories:

(1) **Sorghum (Rabi/Kharif) + Pulses/Commercial (Kharif):** This includes two types of cropping systems. The first is a typical intra-seasonal cropping system where Kharif sorghum is cropped (in row proportions) with pulses (e.g., pigeonpea and blackgram) and

[28] For details of crops included in each crop category, please see Table 1, Appendix 1(I).

commercial crops (e.g., groundnut) to improve productivity and profitability under rainfed conditions (Maruthi et al. 2017). The second is a typical sequential inter-seasonal cropping system where households cultivating sorghum in Rabi, crop pulses and other commercial crops in Kharif reflecting the common use of short-duration pulse crops in multiple cropping sequences semi-arid India (Singh et al. 2009). Notably, cases where the household grown only one crop type in Kharif i.e., either pulses or commercial crops are rare.

(2) **Sorghum Only (Rabi) / Sorghum (Rabi) + Cereals (Rabi):** This is a subsistence-oriented cropping system critical for food consumption and household nutritional security (Kholova et al. 2013). Sorghum is typically cultivated under rainfed conditions utilizing stored soil moisture (Seetharama et al. 1990; Mudalgiriyappa et al. 2012) either by itself (i.e., sorghum-only cultivators) or it may be intra-seasonally cropped in overlap with cereals (i.e., wheat) contingent on relatively higher moisture availability (or irrigation availability) than that required for sorghum-only cultivation.

(3) **Sorghum (Overall):** This includes all cases described in (1) and (2).

SI Section 10 (I) provides further details of this data construction exercise and a snapshot of the main results. Importantly, all results are robust to the exclusion of Sorghum (Kharif) cultivation in cropping system (1). Moreover, the results also hold if Sorghum-only households are considered as a separate system.

Major Revision 2

Across both chapters 2 and 3, an important area of progress has been the construction of village-level seasonal price variables using geocoded and disaggregated *mandi*[29]-level data for the empirical analysis. Earlier versions of these papers relied on district-level, season-year-wise farm harvest prices which were also identified as a potential explanation for the perverse (negative) relationship observed between past output prices and current input use in the instrumental variable regression results in Paper 2. To address this issue, we scraped mandi names and associated market information for VDSA districts and their contiguous districts from the AgMarknet portal. This was followed by manual geocoding in Google

[29] In Hindi, mandi refers to a trading hub or market, particularly for agricultural produce. Mandis are notified physical wholesale markets where crops, fruits, and vegetables are sold directly by farmers. Regulated mandis are notified, physical markets "typically designated as the principal, and in some cases the only state-sanctioned sites for overseeing the critical "first transaction" between primary producers and buyers of agricultural produce" (Krishnamurthy, 2020; p. 179).

Earth for approximately 200 mandis for which location information was unavailable, resulting in a daily, commodity-wise database of prices and arrivals for 600 mandis spanning 14 years. I then harmonized commodity classifications across the VDSA crop data and mandi commodity listings to construct summary measures for modal prices using the nearest five mandis for each village. The resulting dataset provides crop-wise and cropping-pattern-wise prices at the village–season–year level, which will be incorporated into the analysis going forward and are expected to improve the precision with which local market conditions are captured. SI Section 10 (II) details the process and presents an overall summary of the data constructed.

2.4.1 Credit Attainment

Table 2.3 summarizes the results for credit attainment. As expected, farmers' economic and social endowments are positively and significantly associated with the likelihood of obtaining credit (Feder et al. 1990; Foltz 2004; Kumar et al. 2020). An increase in household wealth raised the odds of credit attainment, albeit modestly. Specifically, a ₹1,000 increase in annual non-farm income (i.e., a ~1 percent increase in average monthly non-farm income) improved the odds of credit attainment by 1.02 to 1.04 times, with similar patterns observed for the value of livestock and other durable assets. Likewise, an increase in land ownership (measured as the proportion of owned land in total operational holdings) improved the odds of credit attainment by 1.02 to 6 times (on-average, ~2.5 times). Prior borrowing experience associated with a reduction in transaction costs and informational asymmetry (Petrick 2004; Choudhary and Jain 2022), increased odds of credit attainment between 1.2 to 1.7 times. Operational landholding was also positively associated with odds of credit attainment but the impact was insignificant.

Consistent with Swain (2002), we find that farmers operating plots with erosive, saline, or alkaline soils were nearly 10 percent less likely to attain formal credit compared to farmers operating plots with good-quality soils. Social hierarchy, proxied by caste, remains an enduring predictor of credit attainment (e.g., Rao (2018), Kumar et al. (2020)) such that farmers belonging to scheduled castes and scheduled tribes on-average had 34% lower odds of accessing formal credit relative to upper caste farmers. On the supply side, our findings were inconclusive across two measures of institutional credit supply at the district level: the density of rural and semi-urban bank branches per 1,000 population and the amount of agricultural credit disbursement. An increase in bank branches was associated with a small but significant decline in formal credit attainment possibly due to banking data at higher spatial granularity not reflecting credit access in all villages (RBI 2019), while an increase in agricultural credit disbursement had no significant effect.

2.4.2 Heterogenous Impacts, Rationing Occurrence and Ex-Ante Odds of Rationing:

First considering the average effect (ATT), we find that credit attainment increased the average sorghum yield by ~70 kilograms per acre, or ~25 percent compared to the baseline, no-credit scenario. However as expected, this locational effect masks significant heterogeneity in impacts across the yield distribution (see QTT estimates plotted in Figure 2.6 (Panel A)). Based on the interpretive framework of Koenker and Hallock (2001), our results show that credit impacts increase across quantiles, are bounded at roughly 25% of counterfactual yields up to the 60th percentile, but climb to about 45% of baseline levels in the higher quantiles. This pattern of heterogeneity indicates that credit not only shifts the average yield upward but also expands the upper tail of the yield distribution, implying increased yield variability among borrowers[30]. As shown in Figure 2.6 (Panel B), the yield distribution for non-borrowers is steeper, indicating a narrower range of yield values and consequently significantly lower variability compared to borrowers[31]. Thus credit induces both location and scale changes in the yield distribution.

Converted into revenue, formal credit attainment leads to an on-average increase in sorghum revenue per acre of approximately ₹585 (ATT), which represents a 26.4% gain over the baseline no-credit scenario. This average effect masks a significant heterogeneity in impacts across quantiles between ₹209 - ₹835, equivalent to 9.4 - 40% of average baseline revenue per acre. While the relative impact of credit is positive in alignment with previous literature (see Table 1, SI S1), the rationality of borrowing in terms of absolute gains from credit utilized for sorghum production is unclear. Considering the average effect, with an average loan size of ₹5524 per acre, the incremental revenue per rupee borrowed is ₹0.11, only slightly exceeding the on-average per-rupee interest cost of ₹0.10. The positive net gain implies that formal credit constraints are indeed binding for the average sorghum-grower but the extent of the constraint is low. This weakly binding constraint reflects low profitability from sorghum (on its own) which aligns with its role as a primary consumption crop at the

[30] Given our rejection of rank similarity, the larger credit effects in the higher quantiles are not effects for individual farmers in these quantiles but the effect on the distribution itself.
[31] Jimi et al. (2019), Kumar et al. (2020), and Hutchins (2020) report similarly positive effect of formal credit on rice yields, rice and wheat yields and corn yields in Bangladesh, India, and the United States respectively. Further details in Table 1 (SI, Section 6).

household level (Proietti et al. 2015; Hossain et al. 2022). Moreover, it suggests that observed borrowing behaviour may be influenced by expected returns from other crops in the household's mixed-cropping portfolio (Townsend 1994).

It is important to recognize that this inference regarding binding credit constraints is based on an average of yield realizations across households. However, the average credit impact is only directly applicable to any individual borrower under the assumption of a "common" effect (Heckman et al., 1997). In our case, since credit has improved yields in the higher quantiles substantially more than the yields in the lower quantiles, assuming rank invariance, the median credit impact is likely to be lower than the average treatment impact, and the inference that credit constraints are binding on average applies only for a *smaller* subset of borrowers (Bedoya et al. 2018), not the entire sample. Quantile treatment effects thus offer a richer lens for assessing sample-wide credit rationing based solely on ex-post average outcomes. Under the assumption of rank invariance, i.e., a scenario where each borrower retains their respective rank in the counterfactual distribution, quantile effects can help us move beyond the assessment of credit constraint incidence based on ex-post realizations to a more nuanced, ex-ante evaluation.

Specifically, using quantile treatment effects we can construct the entire distribution of potential yields. The *expected* impact of credit for any household is given as a weighted sum of the individual impacts across quantiles (see equation (16)). The specific quantile that an individual household occupies, and thus the realized impact, depends on its rank in the conditional yield distribution, which itself is a function of both $\vec{X}_i$ and u_i. Using the distributional estimates, we can compute the odds that under credit a household attains a yield (or equivalently revenue level) exceeding the break-even threshold dictated by the cost of borrowing which can be interpreted as an ex-ante likelihood that credit constraints are binding (i.e., τ_i^*). This approach provides a more precise understanding of rationing incidence, capturing both the heterogeneity of credit impacts and the probabilistic nature of binding credit constraints within the sample.

Figure 2.6 (Panel C) summarizes the findings of τ_i^* in our sample for borrowers who (a) grow only sorghum (i.e., do not intercrop), (b) intercrop sorghum with pulses/grains and (c) intercrop sorghum with vegetables or commercial crops (e.g., cotton). As expected, we find that non-intercropping borrowers are almost-never rationed. However, after accounting for the fungibility of agriculturally productive credit and considering the composite marginal credit returns from crops intercropped with sorghum, we find that ~48% of Sorghum borrowing farmers faced strictly positive odds of binding credit constraints or rationing. The distribution of rationing extent reflects the significant heterogeneity in rationing incidence among borrowers in our sample. In general, we find that credit constraints are driven by the returns on vegetables and other commercial crops intercropped with sorghum.

2.4.2.1 Statistical Discrimination:

Table 2.4 presents the distribution of loan amounts and the odds of credit rationing by farmer wealth category, conditional on lender-observable indicators of agricultural productivity and stratified by interest rate regime (ISS and BPLR). In line with statistical discrimination, there is a positive association between borrower wealth and loan amount per acre, even after controlling for observable productivity proxies. Asset-rich borrowers consistently obtain higher average loan amounts than asset-poor borrowers across both regimes, with the disparity being more pronounced under the BPLR regime, where lenders exercise greater pricing flexibility. Notably, higher measured productivity such as irrigation access and regional historical yields partially attenuates wealth-associated loan differences, but does not fully offset them. For example, the gap in loan amounts by wealth remains statistically significant even with irrigation access, while higher regional productivity marginally reduces the gap but not significantly. The unconditional comparison of rationing odds indicates that wealthier borrowers are significantly less likely to be credit rationed. This pattern persists at the regulated interest rate boundary within the ISS regime and across productivity indicators, though regime-specific differences are not statistically significant. In contrast, under the BPLR regime, the directional relationship reverses across productivity categories, suggesting regime-dependent heterogeneity in rationing mechanisms, potentially attributable to lender discretion in interest rate setting. Overall, these results suggest persistent, regime-dependent, wealth-based disparities in both credit allocation and rationing, partially mitigated by productivity observables.

2.4.2.2 Rationing and Input Use:

Table 2.5 summarizes input use by rationing status. We find that credit rationing impairs farmer’s adaptive capacity by limiting access to better technologies like high-yielding variety seeds and irrigation. Specifically, rationed borrowers are 20% more likely to utilize local seeds and 26% less likely to use irrigation. Rationed households also utilize a higher amount of female labour which is typically cheaper (Kundu and Das 2019) while male labour hours are no different. Fertilizer emerges as a risk-complementing input (Akinseye et al. 2023) whereby fertilization frequency and extent of use are higher among rationed borrowers relative to non-rationed borrowers.

2.4.2.3 Robustness

We assess the quality of our matching strategy by examining the differences in covariate means across the treated and counterfactual groups prior to and after matching. Our results (see Table 1(S9) and Table 2(S9)) show that matching through the propensity score has resulted in a reduction in bias across all covariates. However, some covariates remain unbalanced on-average between the treated and control groups. Following Caliendo and Kopeinig (2008) and Stuart (2010), we include these covariates in our yield regressions to assess the robustness of our main findings. All our main results hold with this balance adjustment (see Table 3(S9), Figure 1(S9) and Figure 2(S9)). We have presented the main results with a year-by-year matching and 0.05 calliper. All our main results and the matching quality assessment are robust to a stricter calliper of 0.01 and a village-year by village-year matching criterion. All the standard errors are clustered at the village-level.

Until May 2020, agricultural crop insurance and formal credit were compulsorily linked in India whereby farmers borrowing from institutional sources were mandated to buy crop insurance under the National Agricultural Insurance Scheme (MoAFW 2014). To account for this, we estimate the formal credit impact excluding districts which received any insurance payout in each year and find that the credit impact is relatively higher in this sub-sample while the pattern on impacts across quantiles remains the same (see Figure 3(S9)). Finally, our calculations for τ_i^* rely on yield regression results through the process detailed in SI Section 8. We expect similar results to hold if we regress total farm revenue (instead of sorghum yields) (see Figure 4 (S9) Panels A and B) and if we regress on credit amount instead of a binary indicator (Figure 5(S9)).

2.5 Summary and Conclusions:

This study employs quantile regression models in conjunction with propensity score matching to estimate the heterogeneous impact of agricultural credit on crop yield quantiles for a representative sample of sorghum-growing farmers in semi-arid and tropical India. We exploit the within-sample variation in the credit-yield linkage to augment sample-wide detection of binding credit constraints beyond average *ex-post* realized outcomes. To this end, we propose *ex-ante* odds of credit rationing as a device for measuring the incidence of binding credit constraints. We interpret these odds in a state-contingent technology framework revealing structural implications of rationing for farmers' adaptive capacity under uncertainty. We further use these odds to motivate hypotheses on statistical discrimination as a rationing mechanism.

Our findings reveal that credit access is disproportionately skewed toward better-endowed households, and while borrowing raises yields, the gains are concentrated in the upper quantiles. Nearly half of borrowers exhibit positive ex-ante odds of rationing, and rationed households adopt less irrigation and high-yielding seeds, indicating constrained adaptive choices. Credit allocation further reflects statistical discrimination, with wealthier farmers receiving larger loans and facing lower rationing odds even at similar productivity levels. Ex-ante rationing odds provide a practical device for indirect rationing detection in a more nuanced manner than incidence based on ex-post outcomes detection. This can be particularly valuable in developing-country settings where direct survey-based elicitation of credit constraints is rare, allowing policymakers to infer *potential* rationing with limited data. Viewing rationing through a state-contingent lens reveals how reduced form linkages between inputs and credit cannot capture the full understanding of how rationing constrains farmers' capacity to adjust to uncertainty.

An important caveat of our analysis that it may underestimate the impact of relaxing credit constraints since it relies on the observed borrowing behaviour of farmers which may not be an accurate representation of the credit constraints faced by a farmer. For example, a non-borrowing farmer may be risk-rationed, i.e., not be borrowing for fear of losing their collateral (Boucher et al. 2009). Overall, our findings raise important questions for agricultural credit targeting policies in developing regions.

2.6 Tables

Variable Acronym	Description	Mean (SD) $D_{i_k t_k}^{Formal}=1$ N = 282	$D_{i_k t_k}^{Formal}=0$ N = 647
Economic Status ($\vec{Z}^{eco}$)			
$size_{it}$	Size of operational holding (acres).	8.9 (8.3)	6.77 (6.45)
own_{it}	Ratio of land area owned to the total operational holding.	0.9 (0.24)	0.87 (0.3)
nf_{it}	Real value of non-farm income (in rupees, base year = 1985).	9809 (11878)	8487 (8950)
$cons_{it}$	Real value of consumer durables (in rupees, base year = 1985).	15710 (21804)	7863 (11587)
liv_{it}	Real value of owned livestock (in rupees, base year = 1985)	6426 (6514)	4160 (5157)
Household-Specific Time Invariant Controls ($\vec{Z}^{fe}$)			
age_i	Age of household head (in years)	48.63 (12.06)	49.61 (13.67)
edu_i	Number of years of formal education of household head.	5.51 (4.95)	3.9 (4.26)
$prior_i$	Count of instances of credit access prior to treatment.	2.01 (2.42)	0.75 (1.43)
$soil_i$	Percentage share of plots with erosive, saline, or infertile soils.	41.2 (47.17)	46.2 (47.2)
$caste_i$	=1 if i belongs to the nomadic tribes or scheduled tribes, else 0.	0.15 (0.35)	0.18 (0.39)
Institutional Production Environment ($\vec{Z}^{ins}$)			
$kharif_{it}$	=1 in the Kharif season, 0 otherwise.	0.40 (0.49)	0.19 (0.39)
$loans_t$	Amount of agricultural credit disbursed (in rupees).	60889 (31714)	59116 (30778)
rs_{dt}	District-level number of banks in rural and semi-urban areas.	138 (45)	152 (32)

Table 2.1: Summary Statistics (Credit Attainment)

Notes:

- Caste classification sourced from Ministry of Social Justice and Empowerment, Government of India.
- Banking variables sourced from RBI.
- indexes districts.

Variable Acronym	Description	Mean (SD)	
		$D_{i_k t_k}^{Formal}=1$ N = 230	$D_{i_k t_k}^{Formal}=0$ N = 520
Y_{it}	Sorghum yield (kg/acre)	408.6 (428.7)	280.91 (286.69)
Inputs ($\vec{X}_{it}^{input}$)			
$nitro_{it}$	Nitrogen application (kg/acre).	0.26 (0.38)	0.23 (0.42)
$potash_{it}$	Potassium application (kg/acre).	0.02 (0.1)	0.02 (0.09)
$firr_{it}$	Count of irrigation instances per acre.	0.65 (1.68)	0.41 (0.38)
$local_{it}$	=1 if type of seed used is local, else 0.	0.28 (0.45)	0.26 (0.44)
intercrop_{it}	=1 if sorghum is intercropped (i.e., cultivated with another crop on the same plot), else 0.	0.47 (0.5)	0.41 (0.49)
Land Quality ($\vec{X}_{i}^{land}$)			
$soil\_ind_i$	=1 if soil on one or more of sorghum plots is erosive, saline, or alkaline.	0.47 (0.5)	0.53 (0.49)
$deep_i$	=1 if depth of soil on one or more sorghum plots is more than 1.1 metres and 0 otherwise.	0.31 (0.46)	0.42 (0.49)
Weather ($\vec{X}_{vs(t)}^{weather}$) + Controls			
$temp_{vs(t)}$	Season-wise average temperature (in degree Celsius).	25.78 (1.42)	25.29 (1.16)
$rain_{vs(t)}$	Season-wise total rainfall (in millimetres).	325.33 (342.03)	207.91 (269.49)
$kharif_{s(t)}$	=1 if season is Kharif, else 0.	0.35 (0.48)	0.2 (0.4)
$krf_{vs(t)}$	$kharif_{s(t)} \times rf_{vs(t)}$	267.5 (380.12)	139.44 (294.81)
$kt_{vs(t)}$	$kharif_{s(t)} \times temp_{vs(t)}$	9.71 (13.2)	5.5 (10.96)
ki_{it}	$kharif_{s(t)} \times irr_{it}$ where irr_{it} =1 if one or more sorghum plots is irrigated, else 0.	0.01 (0.09)	0.003 (0.06)
Farmer Experience ($\vec{X}_{it}^{farmer}$)			
age_{it}	Age of household head (in years)	48.87 (12.18)	49.24 (13.31)
edu_i	Number of years of formal education of household head.	4.78 (4.61)	3.99 (4.35)

Table 2.2: Summary Statistics (Yield Regression) pertaining to 0.05-calliper matched dataset. $s(t)$ denotes season.

Dependent Variable: D_{it}	Estimate (Standard Error) **Formal** **N = 920, Log Likelihood =** -460.43
Economic Status ($\vec{Z}^{eco}$)	
$size_{it}$	0.02* (0.01)
own_{it}	0.94* (0.55)
nf_{it}	0.2E-3*** (7.75E-6)
$cons_{it}$	1.4E-4** (7E-6)
liv_{it}	6.8E-4*** (2.21E-4)
Household-Specific Time Invariant Controls ($\vec{Z}^{fe}$)	
age_i	0.007 (0.007)
edu_i	0.05* (0.03)
$prior_i$	0.34*** (0.09)
$soil_i$	-0.003* (0.001)
$caste_i$	-0.41* (0.23)
Institutional Production Environment ($\vec{Z}^{ins}$)	
$kharif_{it}$	0.73 (0.46)
$loans_t$	5.16E-6 (4.08E-6)
rs_{dt}	-0.01** (0.004)
constant	-1.76 (1.19)

Table 2.3: Credit Attainment Regression Results

Notes: Standard errors are clustered at the village-level. Stars indicate *p*-values. ***: $p<0.01$; **: $p<0.05$; * $p<0.10$

(Interest Rate)	Lender's Observables Indicators of Productivity	Irrigation Application		Irrigation & Rainfall Levels			Agricultural Productivity	
		Yes	No	Low Rainfall, No Irrigation	Low Rainfall, With Irrigation †	High Rainfall, No Irrigation	High	Low
	Loan Amount by Wealth Categories							
(ISS): 6-9%	**Low** 1227 (1253) [22]	₹1232 (899) [8]	₹1224 (1449) [14]	₹1334 (1535) [12]	₹1232 (899) [8]	₹560 (598) [2]	₹1359 (1447) [3]	₹1206 (1263) [19]
(ISS): 6-9%	**High** 1514 (1152) [34]	₹1180 (902) [10]	₹1653 (1232) [24]	₹1784 (1259) [21]	₹1180 (902) [10]	₹741 (405) [3]	₹389 (0) [1]	₹1548 (1152) [33]
(BPLR): 12-15%	**Low** 1217** (1394) [44]	₹1598** (1298) [17]	₹977 (1422) [27]	₹1396 (1965) [9]	₹1598** (1298) [17]	₹768 (1065) [18]	₹820 (1121) [16]	₹1444** (1500) [28]
(BPLR): 12-15%	**High** 2482 (3189) [40]	₹ 3146 (2712) [22]	₹1670 (3602) [18]	₹ 2731 (4973) [9]	₹3250 (2734) [21]	₹609 (559) [9]	₹716 (572) [9]	₹2995 (3453) [31]
	Odds of Rationing by Wealth Categories							
(ISS): 6-9%	**Low** 0.25 (0.38) [22]	0.36 (0.38) [8]	0.19 (0.38) [14]	0.07 (0.27) [12]	0.36 (0.38) [8]	0.88 (0.14) [2]	0.81 (0.16) [3]	0.17 (0.33) [19]
(ISS): 6-9%	**High** 0.15 (0.32) [34]	0.33 (0.41) [10]	0.08 (0.24) [24]	0.02 (0.06) [21]	0.32 (0.41) [10]	0.56 (0.49) [3]	0 (0) [1]	0.16 (0.32) [33]
(BPLR): 12-15%	**Low** 0.36 (0.44) [44]	0.59* (0.41) [27]	0 (0) [17]	0.15 (0.31) [9]	0 (0) [17]	0.82 (0.25) [18]	0.80 (0.27) [16]	0.12 (0.29) [28]
(BPLR): 12-15%	**High** 0.35 (0.40) [40]	0.63 (0.39) [18]	0.12 (0.24) [22]	0.38 (0.44) [9]	0.09 (0.19) [21]	0.87 (0.09) [9]	0.86 (0.11) [9]	0.20 (0.33) [31]

Table 2.4: Statistical Discrimination by Wealth

Notes:

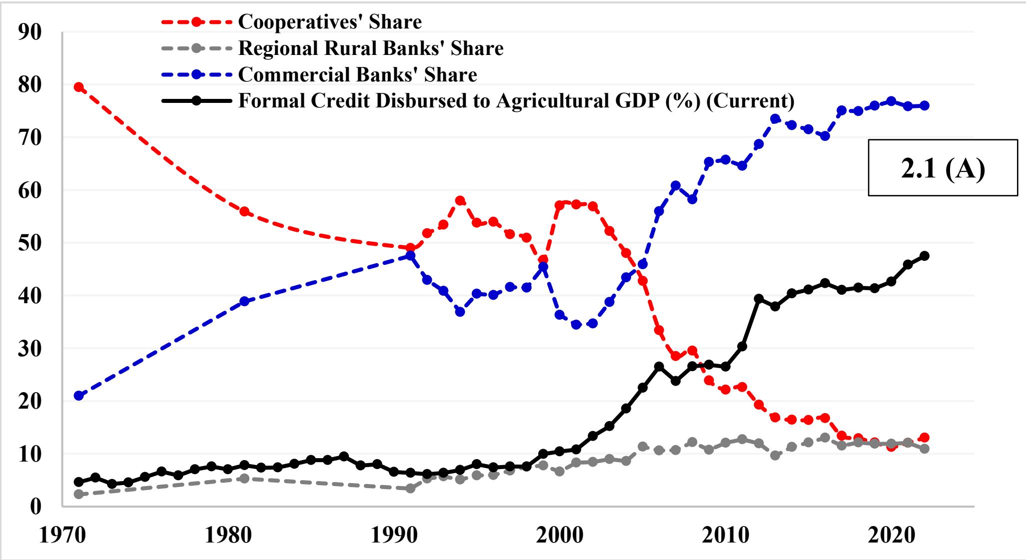
- Wealth index is total real value of livestock, agricultural equipment, consumer durables, consumption expenditure, non-farm income and savings. Low Wealth: ₹1487- ₹5807; High Wealth: ₹5807- ₹33828.
- †**:** High Rainfall, With Irrigation category is excluded due to negligible sample size.
- (.) contain standard errors. [.] indicate sample size. Stars indicate *p*-values of the difference in average loan amounts between wealth categories. *: $p<0.10$ | **: $p<0.05$ | ***: $p<0.01$.
- Agricultural productivity is measured by the district-wise average yields for the last five years ("threshold yield"). A district is classified as low productivity in a given year if its threshold yield is below the median threshold yield for that year.

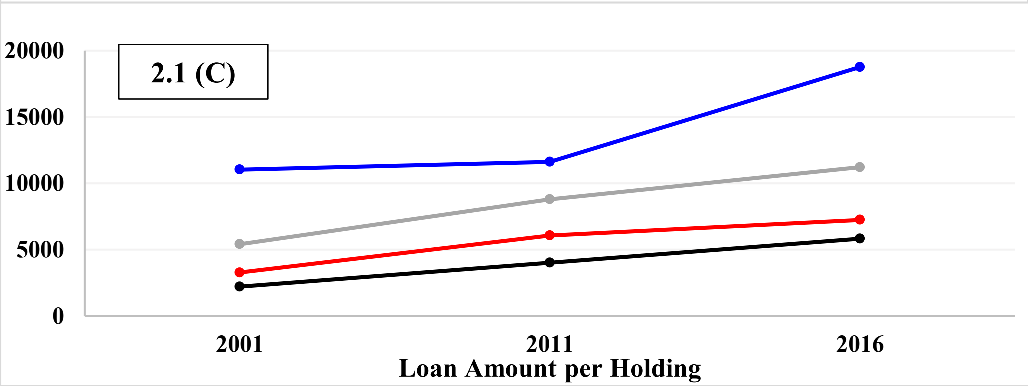
Input Type	Average Difference (Non-Rationed – Rationed)
Local Seed Use	-0.20*** (0.06)
Labour Hours per acre (Male)	-0.17 (0.35)
Labour Hours per acre (Female)	-1.51*** (0.52)
Irrigation Access (Yes/No)	0.26*** (0.06)
Irrigation Frequency	0.43*** (0.18)
Nitrogen Application (kg/acre)	-0.16*** (0.05)
Fertilization Frequency	-2.3*** (0.4)

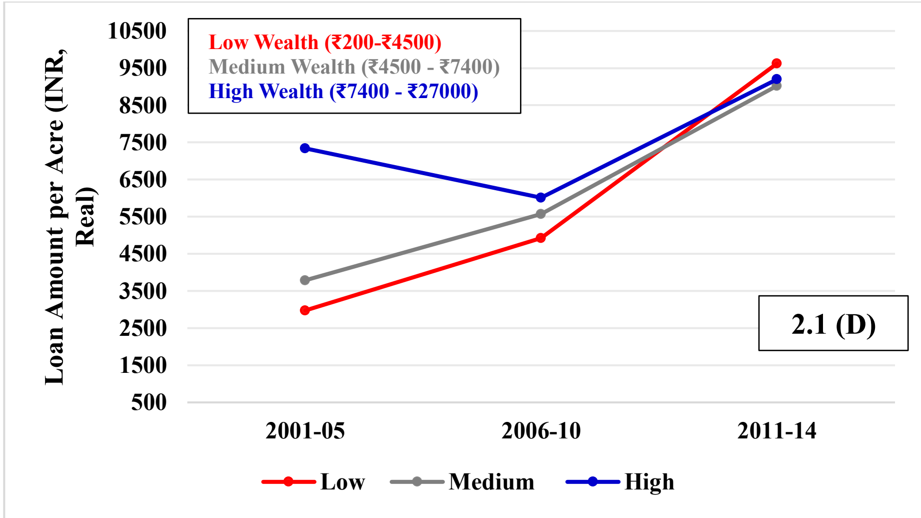
Table 2.5: Input Use by Credit Rationing Status

Notes: Stars indicate *p*-values. *: *p*<0.10 | **: *p*<0.05 | ***: *p*<0.01

2.7 Figures:







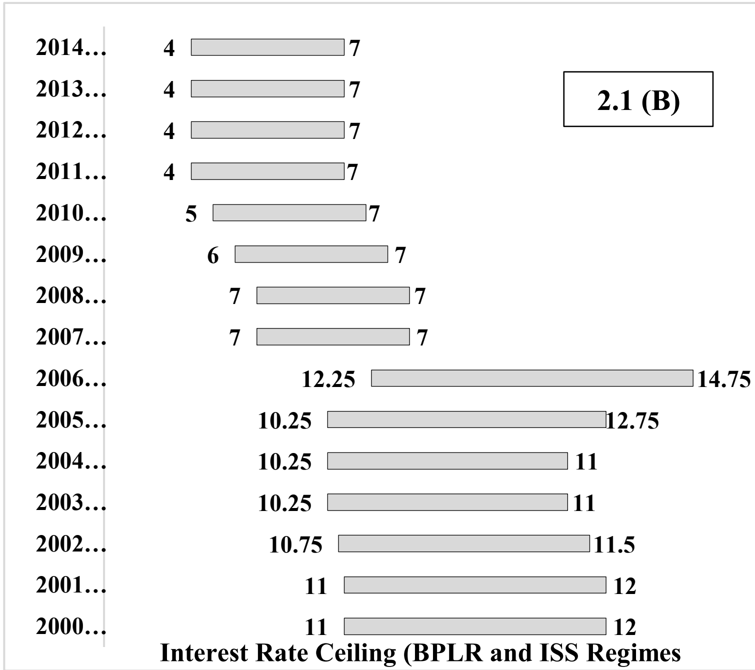


Figure 2.1: Formal Agricultural Credit in India

(A): Credit Disbursal Volume and Shares | **Data Source:** RBI Database for the Indian Economy (RBI-DBIE); **(B):** Interest Rate Ceilings (BPLR and ISS Regimes) | **Data Source:** RBI-DBIE; **(C):** Average formal loan amount per holding by land size in semi-arid districts (INR; real, base year = 1986) | **Data Source:** Input Surveys, MoAFW; **(D):** Average formal loan amount per hectare by wealth quartiles in VDSA Data (INR; real, base year = 1986)

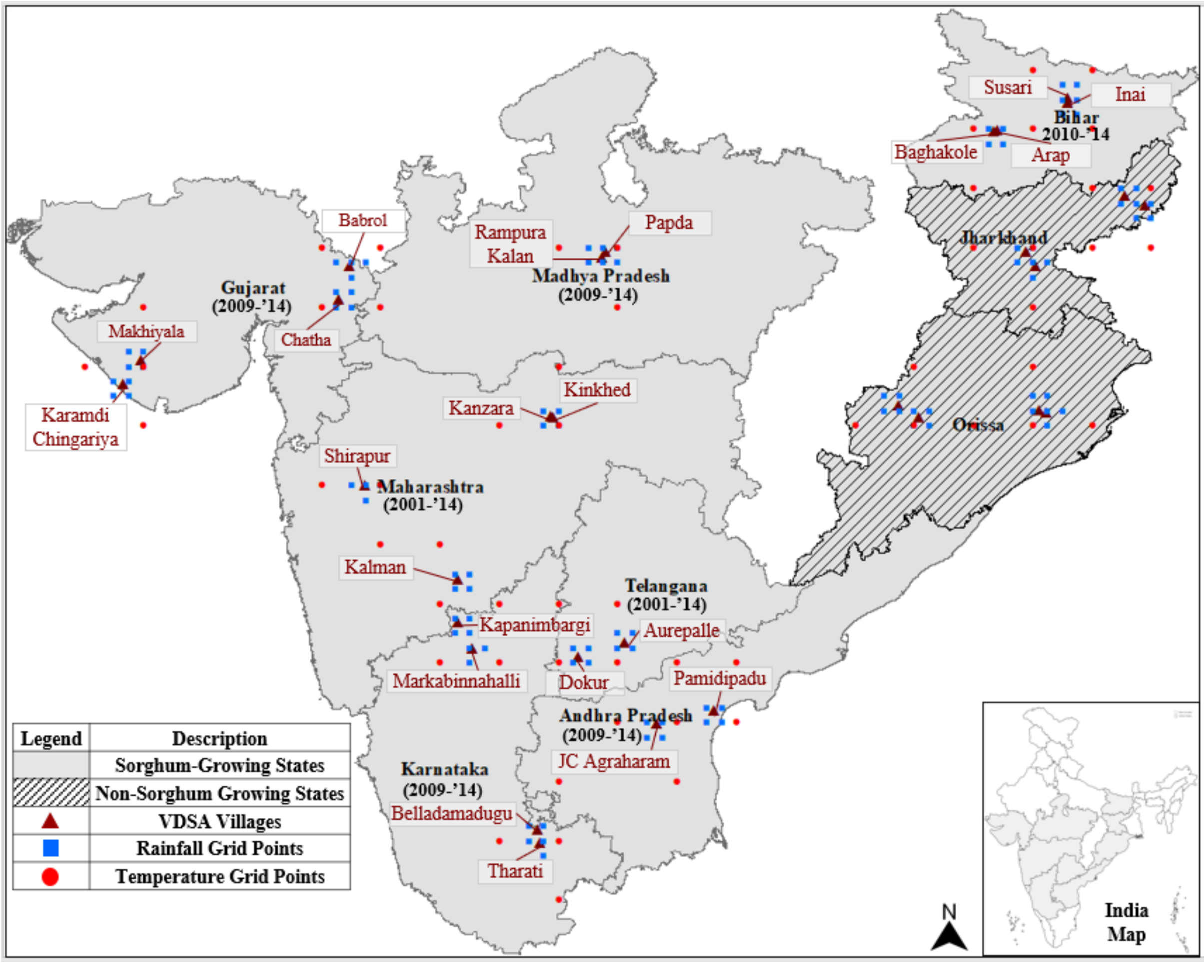
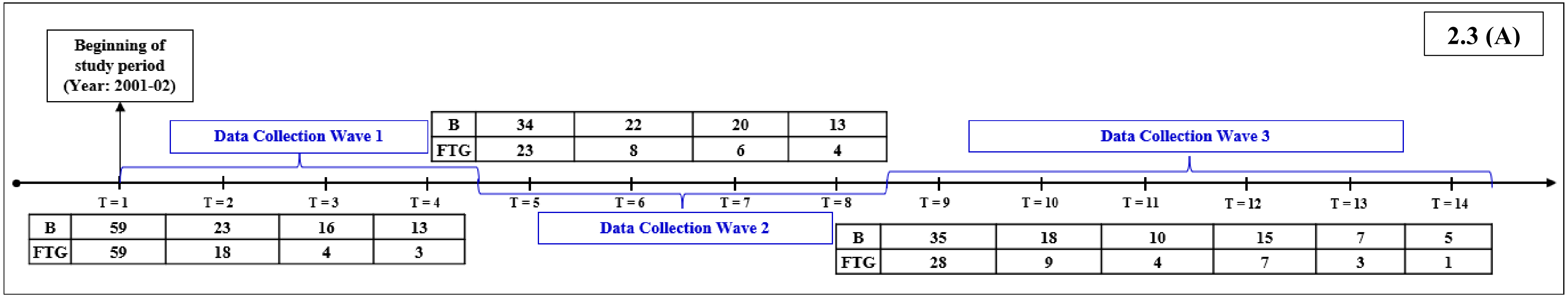


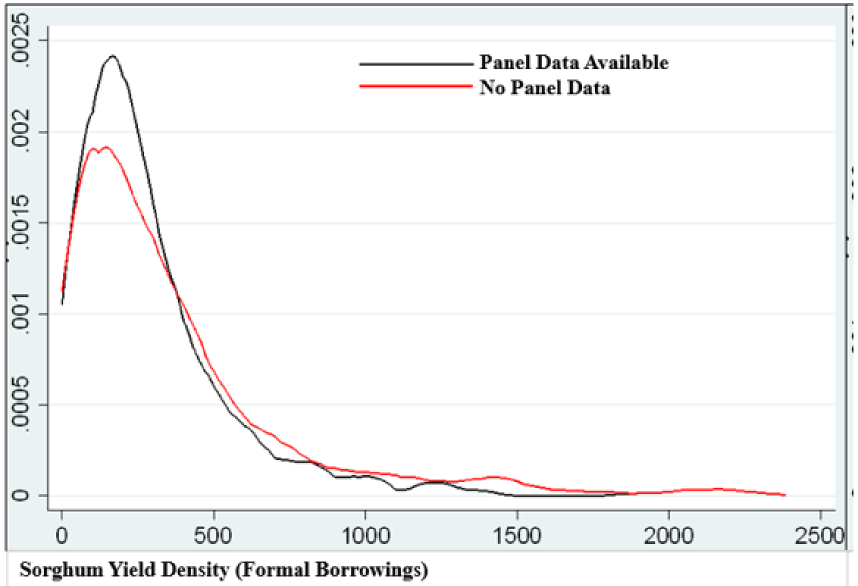
Figure 2.2: Study Area

Notes:

- State names are mentioned in **black** and villages are labelled in **brown**.
- State-wise years of data availability are mentioned in parentheses.
- The study area constitutes four villages spread across four talukas in two districts in each state except Madhya Pradesh where we have only two villages (in two talukas of one district).
- The state of Telangana was carved out of Andhra Pradesh in 2014.

Figure 2.3: Data Details **(A):** Borrowing Behaviour of Sorghum-Growing Households; **(B):** Sorghum Yield Density (With and Without Panel Possibility). **Note:** B represents the count of first-time borrowers in T while FTG represents the count of first-time borrowers which grow Sorghum first time in T.





2.3 (B)

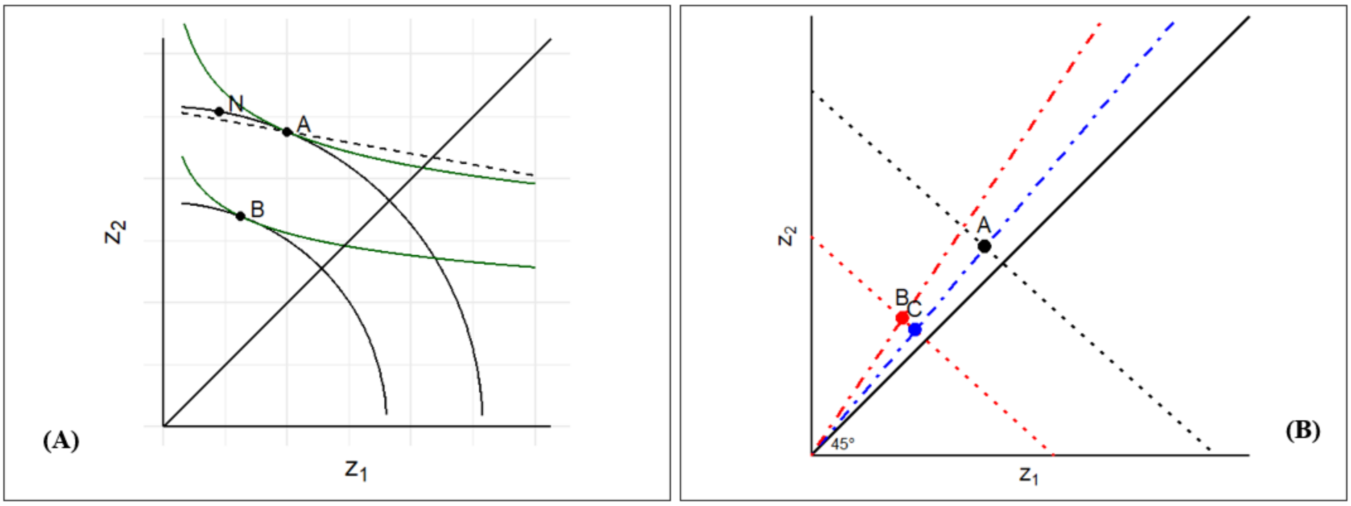


Figure 2.4: Illustration of Binding Credit Constraint in a State-Contingent Technology framework.
Note:

- **Dashed Line:** Fair-Odds Line
- **Dot-Dashed Line:** Revenue Mix Line(s)
- **Solid Curve:** Indifference Curve
- **Solid Line:** Bisector

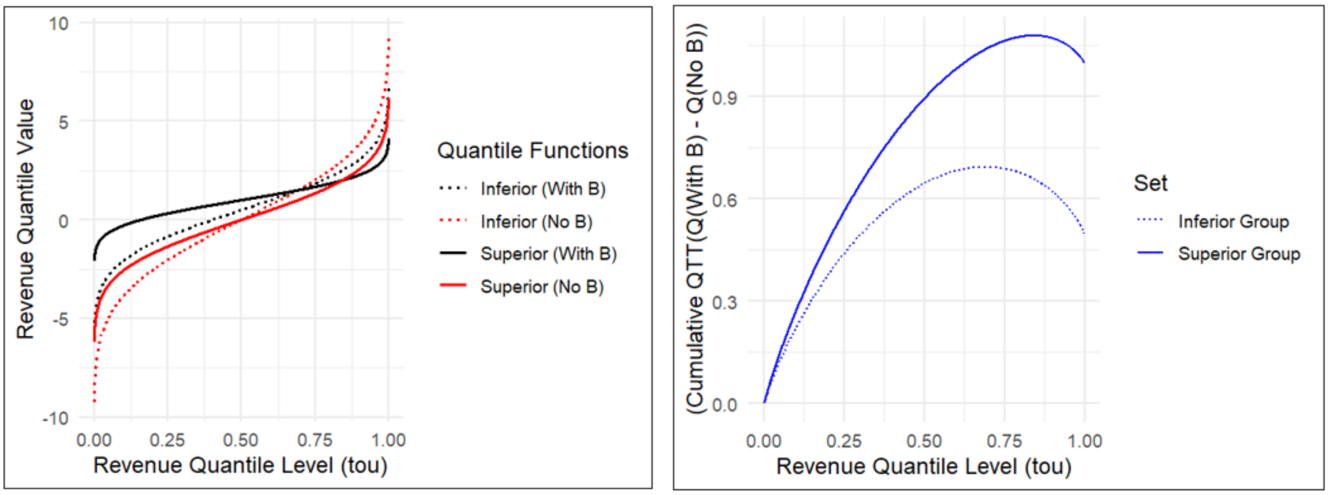
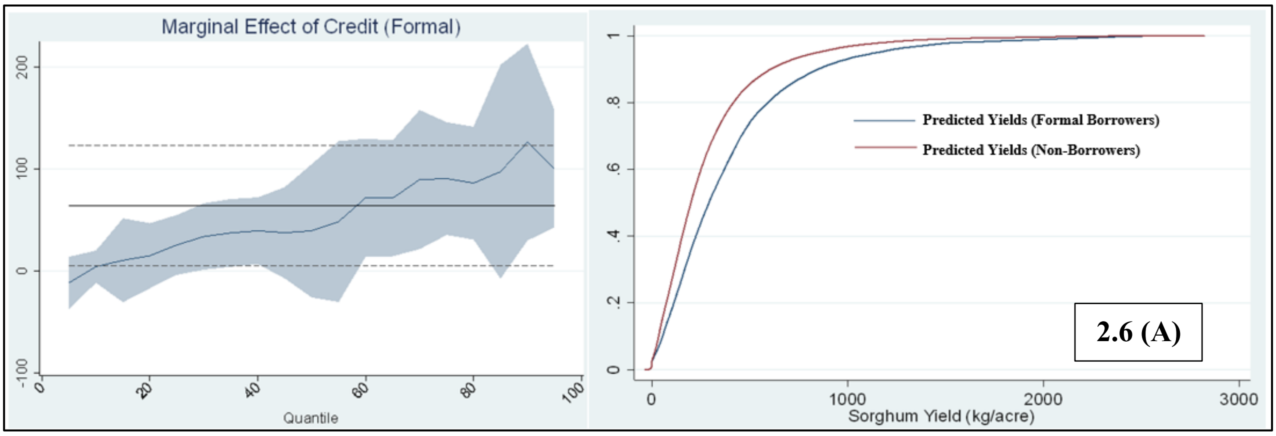
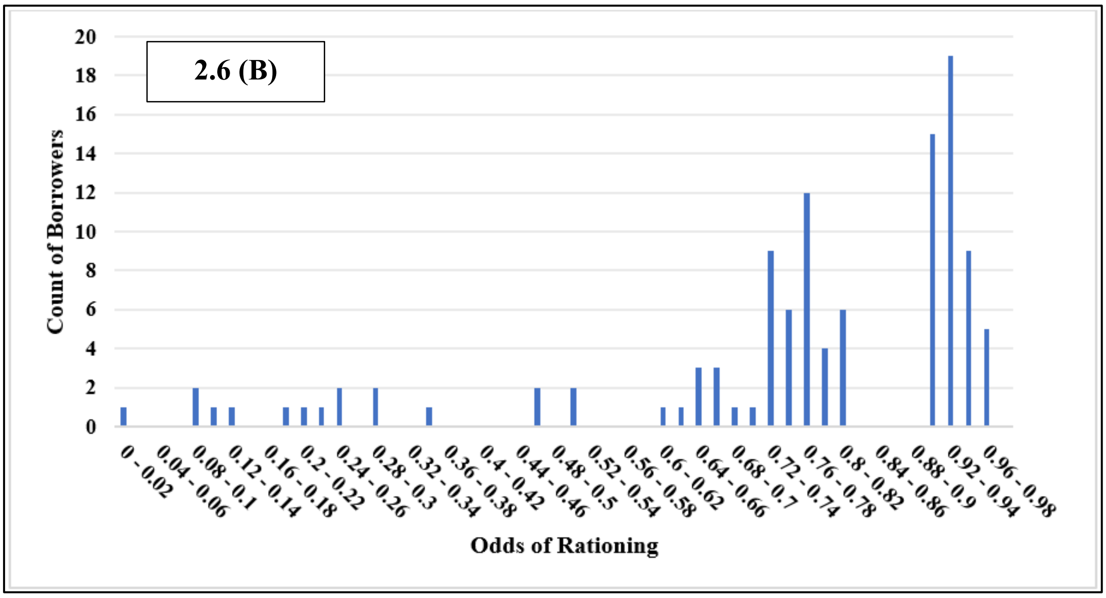


Figure 2.5: Illustration of Statistical Discrimination Term 1 (Equation (17))

(A): Quantile Functions for Revenue | **(B):** Term 1 (Equation (17))

Notes: Dashed Line: Inferior Group | **Solid Line:** Superior Group





Intercropping Status	Count of Borrowers	Count of Rationed Borrowers (i.e. $(1-\tau_i^* > 1)$)	Average odds of rationing incidence among *rationed* borrowers
No Intercropping	85	0	0
Sorghum intercropped with Pulses/Grains	37	3	0.08
Sorghum intercropped with Vegetables/Commercial Crops	109	109	0.77

2.6 (C)

Figure 2.6: Results | **(A):** Marginal effect of formal credit attainment across Sorghum *yield* quantiles; **(B):** Predicted Yields CDF for Formal Credit versus Non-Borrowers**; (C):** Distribution of Rationing Odds among Rationed Borrowers ($1-\tau_i^* \to 1$).

Notes: Grey band represents the 95% confidence interval. Solid and dashed lines represent the linear regression (OLS) estimate and associated 95% confidence interval respectively.

2.8 References

Abate, Gashaw Tadesse, Shahidur Rashid, Carlo Borzaga, and Kindie Getnet. 2016. "Rural Finance and Agricultural Technology Adoption in Ethiopia: Does the Institutional Design of Lending Organizations Matter?" *World Development* 84 (April): 235–53. https://doi.org/10.1016/j.worlddev.2016.03.003.

Ali Daniel & Deininger K. 2012. "Causes and Implications of Credit Rationing in Rural Ethiopia : The Importance of Spatial Variation." *Policy Research Working Paper Series 6096, The World Bank.* https://ideas.repec.org/p/wbk/wbrwps/6096.html.

Akinseye, Folorunso M., Birhanu Zemadim Birhanu, Hakeem A. Ajeigbe, Madina Diancoumba, Karamoko Sanogo, and Ramadjita Tabo. 2023. "Impacts of Fertilization Management Strategies on Improved Sorghums Varieties in Smallholder Farming Systems in Mali: Productivity and Profitability Differences." *Heliyon* 9 (3): e14497. https://doi.org/10.1016/j.heliyon.2023.e14497.

Ali, Daniel., Deininger, K., & Duponchel, M. (2014). Credit Constraints and Agricultural Productivity: Evidence from rural Rwanda. *The Journal of Development Studies*, 50(5), 649–665. https://doi.org/10.1080/00220388.2014.887687.

Angrist, Joshua, and Jörn-Steffen Pischke. 2008. "Chapter 7: Quantile Regression." In *Mostly Harmless Econometrics: An Empiricist's Companion*, by Joshua Angrist and Jörn-Steffen Pischke, 268-291. Princeton University Press.

Appau, Adriana, Jeffrey Drope, Firman Witoelar, Jenina Joy Chavez, and Raphael Lencucha. 2019. "Why Do Farmers Grow Tobacco? A Qualitative Exploration of Farmers Perspectives in Indonesia and Philippines." *International Journal of Environmental Research and Public Health* 16 (13): 2330. https://doi.org/10.3390/ijerph16132330.

Arpino, Bruno, and Massimo Cannas. 2016. "Propensity Score Matching With Clustered Data. An Application to the Estimation of the Impact of Caesarean Section on the Apgar Score." *Statistics in Medicine* 35 (12): 2074–91. https://doi.org/10.1002/sim.6880.

Arora, G., Wolter, P. T., Feng, H. & Hennessy, D. A., 2016. Role of Ethanol Plants in Dakotas Land Use Change: Incorporating Flexible Trends in the Difference-in-Difference Framework with Remotely-Sensed Data, Iowa: Iowa State University. Accessed here.

Basak, Nirmalendu, Biswapati Mandal, Arvind Kumar Rai, and Piu Basak. 2021. "Soil Quality and Productivity Improvement: Indian Story." *Proceedings of the Indian National Science Academy* 87 (1): 2–10. https://doi.org/10.1007/s43538-021-00007-8.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2023. "Selection Into Credit Markets: Evidence From Agriculture in Mali." *Econometrica* 91 (5): 1595–1627. https://doi.org/10.3982/ecta18916.

Besley, Timothy. 1994. "HOW DO MARKET FAILURES JUSTIFY INTERVENTIONS IN RURAL CREDIT MARKETS?" *The World Bank Research Observer* 9 (1): 27–47. https://doi.org/10.1093/wbro/9.1.27.

Bhattarai, Madhusudan, Gajendra Singh, Hiroyuki Takeshima, and Ravindra S. and Shekhawat. 2020. "Chapter 3: Farm machinery use and the agricultural machinery industries in India: Status, evolution, implications, and lessons learned." In *An evolving paradigm of agricultural mechanization development: How much can Africa learn from Asia?*, by Xinshen Diao, Hiroyuki Takeshima and Xiaobo and Zhang, 97-138. CGIAR.

Bhende, M. J. 1986 "Credit Markets in Rural South India." *Economic and Political Weekly* 21(38): A119-A124. https://www.jstor.org/stable/4376153.

Binswanger, Hans P., and Mark R. Rosenzweig. 1986. "Behavioural and Material Determinants of Production Relations in Agriculture." *The Journal of Development Studies* 22 (3): 503–39. https://doi.org/10.1080/00220388608421994.

Boucher, Stephen R., Michael R. Carter, and Catherine Guirkinger. 2008. "Risk Rationing and Wealth Effects in Credit Markets: Theory and Implications for Agricultural Development." *American Journal of Agricultural Economics* 90 (2): 409–23. https://doi.org/10.1111/j.1467-8276.2007.01116.x.

Boucher, Stephen R., Catherine Guirkinger, and Carolina Trivelli. 2009. "Direct Elicitation of Credit Constraints: Conceptual and Practical Issues With an Application to Peruvian Agriculture." *Economic Development and Cultural Change* 57 (4): 609–40. https://doi.org/10.1086/598763.

Braverman, Avishay, and J.Luis Guasch. 1986. "Rural Credit Markets and Institutions in Developing Countries: Lessons for Policy Analysis From Practice and Modern Theory." *World Development* 14 (10–11): 1253–67. https://doi.org/10.1016/0305-750x(86)90104-x.

Burgess, Robin, and Rohini Pande. 2005. "Do Rural Banks Matter? Evidence From the Indian Social Banking Experiment." *American Economic Review* 95 (3): 780–95. https://doi.org/10.1257/0002828054201242.

Caliendo, Marco, and Sabine Kopeinig. 2008. "SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING." *Journal of Economic Surveys* 22 (1): 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2020. "Difference-in-Differences With Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

Carter, Michael R. 1988. "Equilibrium Credit Rationing of Small Farm Agriculture." *Journal of Development Economics* 28 (1): 83–103. https://doi.org/10.1016/0304-3878(88)90015-6.

Carter, Michael R. 1989. "The Impact of Credit on Peasant Productivity and Differentiation in Nicaragua." *Journal of Development Economics* 31 (1): 13–36. https://doi.org/10.1016/0304-3878(89)90029-1.

Carter, Michael R., and Keith D. Wiebe. 1990. "Access to Capital and Its Impact on Agrarian Structure and Productivity in Kenya." *American Journal of Agricultural Economics* 72 (5): 1146–50. https://doi.org/10.2307/1242523.

Carter, Michael R., and Pedro Olinto. 2003. "Getting Institutions 'Right' for Whom? Credit Constraints and the Impact of Property Rights on the Quantity and Composition of Investment." *American Journal of Agricultural Economics* 85 (1): 173–86. https://doi.org/10.1111/1467-8276.00111.

Chambers, RG & Quiggin, J 2000, *Uncertainty, Production, Choice and Agency: The State-Contingent Approach*. vol. 1, First edn, Cambridge University Press, New York.

Chambers, Robert G., and John Quiggin. 2001. "Decomposing Input Adjustments Under Price and Production Uncertainty." *American Journal of Agricultural Economics* 83 (1): 20–34. https://doi.org/10.1111/0002-9092.00134.

Chang, Ting-Hsuan, and Elizabeth A. Stuart. 2022. "Propensity Score Methods for Observational Studies With Clustered Data: A Review." *Statistics in Medicine* 41 (18): 3612–26. https://doi.org/10.1002/sim.9437.

Choudhary, M. Ali, and Anil Jain. 2022. "Finance and Inequality: The Distributional Impacts of Bank Credit Rationing." *Journal of Financial Intermediation* 52 (September): 100997. https://doi.org/10.1016/j.jfi.2022.100997.

Ciaian, Pavel, Jan Fałkowski, and D'Artis Kancs. 2012. "Access to Credit, Factor Allocation and Farm Productivity." *Agricultural Finance Review* 72 (1): 22–47. https://doi.org/10.1108/00021461211222114.

David, Cristina C., and Richard L. Meyer. 1979. "Measuring the Farm Level Impact of Agricultural Loans in Low Income Countries: A Review Article." June 1, 1979. https://kb.osu.edu/items/60fedb66-bc65-5d86-b67f-4e427ece7a2a.

Deininger, Klaus & Olinto Pedro. 2000. "Why Liberalization Alone Has Not Improved Agricultural Productivity in Zambia : The Role of Asset Ownership and Working Capital Constraints." *Ideas.Repec.Org*, March. https://ideas.repec.org/p/wbk/wbrwps/2302.html.

Diagne, Aliou & Zeller Manfred. 2001. "Access to Credit and Its Impact on Welfare in Malawi." *Ideas.Repec.Org*. https://ideas.repec.org/p/fpr/resrep/116.html.

Dong, Yingying & Shen, Shu. 2018. "Testing for Rank Invariance or Similarity in Program Evaluation." *The Review of Economics and Statistics* 100(1): 78-85. https://doi.org/10.1162/REST_a_00686

Eswaran, Mukesh, and Ashok Kotwal. 1989. "Credit as Insurance in Agrarian Economies." *Journal of Development Economics* 31 (1): 37–53. https://doi.org/10.1016/0304-3878(89)90030-8.

Feder, Gershon, Lawrence J. Lau, Justin Y. Lin, and Xiaopeng Luo. 1990. "The Relationship Between Credit and Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium." *American Journal of Agricultural Economics* 72 (5): 1151–57. https://doi.org/10.2307/1242524.

Frandsen, Brigham R., and Lars J. Lefgren. 2018. "Testing Rank Similarity." *The Review of Economics and Statistics* 100 (1): 86–91. https://doi.org/10.1162/rest_a_00675.

Frölich, Markus. 2007. "Propensity Score Matching Without Conditional Independence Assumption—with an Application to the Gender Wage Gap in the United Kingdom." *Econometrics Journal* 10 (2): 359–407. https://doi.org/10.1111/j.1368-423x.2007.00212.x.

Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica* 75 (1): 259–76. https://www.jstor.org/stable/4123114.

Foltz, J. 2004. "Credit Market Access and Profitability in Tunisian Agriculture." *Agricultural Economics* 30 (3): 229–40. https://doi.org/10.1016/j.agecon.2002.12.003.

Fuwa, Nobuhiko, Christopher Edmonds, and Pabitra Banik. 2007. "Are Small-scale Rice Farmers in Eastern India Really Inefficient? Examining the Effects of Microtopography on Technical Efficiency Estimates." *Agricultural Economics* 36 (3): 335–46. https://doi.org/10.1111/j.1574-0862.2007.00211.x.

Guirkinger, Catherine, and Stephen R. Boucher. 2008. "Credit Constraints and Productivity in Peruvian Agriculture." *Agricultural Economics* 39 (3): 295–308. https://doi.org/10.1111/j.1574-0862.2008.00334.x.

Gulati, Ashok; Juneja, Ritika (2019) : Agricultural credit system in India: Evolution, effectiveness and innovations, ZEF Working Paper Series, No. 184, University of Bonn, Center for Development Research (ZEF), Bonn. http://hdl.handle.net/10419/206974

Hossain, Md. Saddam, Md. Nahidul Islam, Md. Mamunur Rahman, Mohammad Golam Mostofa, and Md. Arifur Rahman Khan. 2022. "Sorghum: A Prospective Crop for Climatic Vulnerability, Food and Nutritional Security." *Journal of Agriculture and Food Research* 8 (April): 100300. https://doi.org/10.1016/j.jafr.2022.100300.

Heckman, J. J., J. Smith, and N. Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *The Review of Economic Studies* 64 (4): 487–535. https://doi.org/10.2307/2971729.

Hossain, Md. Saddam, Md. Nahidul Islam, Md. Mamunur Rahman, Mohammad Golam Mostofa, and Md. Arifur Rahman Khan. 2022. "Sorghum: A Prospective Crop for Climatic Vulnerability, Food and Nutritional Security." *Journal of Agriculture and Food Research* 8 (April): 100300. https://doi.org/10.1016/j.jafr.2022.100300.

Huntington-Klein, N. (2021). The Effect: An Introduction to Research Design and Causality. CRC Press: Taylor & Francis Group.

Isom, W.H. and Worker, G.F. (1979) Crop Management in Semi-Arid Environments. In Hall, A.E., Cannell, G.H. and Lawton, H.W., Eds., Agriculture in Semi-Arid Environments, Ecological Studies 34, Springer-Verlag, Berlin, Heidelberg, New York 200-223. Scientific Research Publishing. https://www.scirp.org/reference/referencespapers?referenceid=2780318.

Jaffee, Dwight M., and Thomas Russell. 1976. "Imperfect Information, Uncertainty, and Credit Rationing." *The Quarterly Journal of Economics* 90 (4): 651. https://doi.org/10.2307/1885327.

Jimi, Nusrat Abedin, Plamen V. Nikolov, Mohammad Abdul Malek, Subal Kumbhakar. 2019. "The Effects of Access to Credit on Productivity: Separating Technological Changes From Changes in Technical Efficiency." *Journal of Productivity Analysis* 52 (1–3): 37–55. https://doi.org/10.1007/s11123-019-00555-8.

Karthick, and Madheswaran S. 2018. "Access to Formal Credit in the Indian Agriculture: Does Caste Matter?" *Journal of Social Inclusion Studies* 4 (2): 169–95. https://doi.org/10.1177/2394481118814064.

Kjenstad, Einar, and Xunhua Su. 2014. "Credit Rationing by Loan Size: A Synthesized Model." *The Quarterly Review of Economics and Finance* 55 (September): 20–27. https://doi.org/10.1016/j.qref.2014.08.001.

Kochar, Anjini. 1997. "An Empirical Investigation of Rationing Constraints in Rural Credit Markets in India." *Journal of Development Economics* 53 (2): 339–71. https://doi.org/10.1016/s0304-3878(97)00020-5.

Koenker, Roger, and Kevin F Hallock. 2001. "Quantile Regression." *The Journal of Economic Perspectives* 15 (4): 143–56. https://doi.org/10.1257/jep.15.4.143.

Kumar, Anjani, Ashok K. Mishra, Vinay K. Sonkar, and Sunil Saroj. 2020. "Access to Credit and Economic Well-Being of Rural Households: Evidence From Eastern India." *Journal of Agricultural and Resource Economics* 45 (1): 145–60. https://www.jstor.org/stable/26869031.

Kumar, V., Joshi, S., Pant, N.C., Sangwan, P., Yadav, A.N., Saxena, A., Singh, D. (2019). Molecular Approaches for Combating Multiple Abiotic Stresses in Crops of Arid and Semi-Arid Region. In: Singh, S., Upadhyay, S., Pandey, A., Kumar, S. (eds) Molecular Approaches in Plant Biology and Environmental Challenges. Energy, Environment, and Sustainability. Springer, Singapore.

Kumar, Sunil Mitra. 2013. "Does Access to Formal Agricultural Credit Depend on Caste?" *World Development* 43 (December): 315–28. https://doi.org/10.1016/j.worlddev.2012.11.001.

Kumbhakar, S. C., and R. Bokusheva. 2009. "Modelling Farm Production Decisions Under an Expenditure Constraint." *European Review of Agricultural Economics* 36 (3): 343–67. https://doi.org/10.1093/erae/jbp031.

Kundu, Amit, and Sangita Das. 2019. "Gender Wage Gap in the Agricultural Labor Market of India: An Empirical Analysis." *Journals.Econsciences.Com*, July. https://doi.org/10.1453/jepe.v6i2.1872.

Liverpool, Lenis Saweda O., and Alex Winter-Nelson. 2010. "Poverty Status and the Impact of Formal Credit on Technology Use and Wellbeing Among Ethiopian Smallholders." *World Development* 38 (4): 541–54. https://doi.org/10.1016/j.worlddev.2009.11.006.

Liu, Xiangping, and Lori Lynch. 2011. "Do Agricultural Land Preservation Programs Reduce Farmland Loss? Evidence From a Propensity Score Matching Estimator." *Land Economics* 87 (2): 183–201. https://www.jstor.org/stable/41307208.

Liu, Zinan, and Juzhong Zhuang. 2000. “Determinants of Technical Efficiency in Post-Collective Chinese Agriculture: Evidence From Farm-Level Data.” *Journal of Comparative Economics* 28 (3): 545–64. https://doi.org/10.1006/jcec.2000.1666.

Luan, Do Xuan, and Siegfried Bauer. 2016. “Does Credit Access Affect Household Income Homogeneously Across Different Groups of Credit Recipients? Evidence From Rural Vietnam.” *Journal of Rural Studies* 47 (August): 186–203. https://doi.org/10.1016/j.jrurstud.2016.08.001.

Menapace, Luisa, Gregory Colson, and Roberta Raffaelli. 2012. “Risk Aversion, Subjective Beliefs, and Farmer Risk Management Strategies.” *American Journal of Agricultural Economics* 95 (2): 384–89. https://doi.org/10.1093/ajae/aas107.

Mohan, Rakesh. 2006. “Agricultural Credit in India: Status, Issues and Future Agenda.” *Economic and Political Weekly* 41 (11). https://www.jstor.org/stable/4417965.

MoAFW. 2014. Report of the Committee to Review the Implementation of Crop Insurance Schemes in India. Accessed here.

Nguyen, Thanh-Tung, Trung Thanh Nguyen, and Ulrike Grote. 2022. “Credit, Shocks and Production Efficiency of Rice Farmers in Vietnam.” *Economic Analysis and Policy* 77 (December): 780–91. https://doi.org/10.1016/j.eap.2022.12.018.

O’Donnell, Christopher J., Robert G. Chambers, and John Quiggin. 2009. “Efficiency Analysis in the Presence of Uncertainty.” *Journal of Productivity Analysis* 33 (1): 1–17. https://doi.org/10.1007/s11123-009-0143-9.

Petrick, M. 2004. “A Microeconometric Analysis of Credit Rationing in the Polish Farm Sector.” *European Review of Agricultural Economics* 31 (1): 77–101. https://doi.org/10.1093/erae/31.1.77.

Pray, Carl & Nagarajan, Latha. (2009). Pearl Millet and Sorghum Improvement in India. Accessed here.

Press Information Bureau. 2014. Notification: Credit to Farmers Under Interest Subvention Scheme. Accessed here.

Proietti, Ilaria, Chiara Frazzoli, and Alberto Mantovani. 2015. “Exploiting Nutritional Value of Staple Foods in the World’s Semi-Arid Areas: Risks, Benefits, Challenges and Opportunities of Sorghum.” *Healthcare* 3 (2): 172–93. https://doi.org/10.3390/healthcare3020172.

Rajakaruna, N., & Boyd, R. S. (2019). Edaphic Factor. Encyclopedia of Ecology, 361-367.

Rao, Y Mohan, K Ravi Chand, VR Kiresur, and MCS Bantilan. 2011. *Documentation of Second-Generation Village Level Studies (VLS) in India (2001/02-2004/05).* VDSA Data Manual, Hyderabad: ICRISAT. Accessed here.

Rao, Shankar. 2017. “Caste Discrimination and Agricultural Performance in India.” *Economic and Political Weekly* 52 (25/26): 32–38. https://www.jstor.org/stable/26696038.

Rasmussen, Svend. 2003. “Criteria for Optimal Production Under Uncertainty. The State-contingent Approach.” *Wiley Online Library*, December. https://doi.org/10.1111/j.1467-8489.2003.t01-1-00223.x.

Rasmussen, Svend, 2004. "Optimizing Production under Uncertainty: Generalisation of the State-Contingent Approach and Comparison of Methods for Empirical Application," *Unit of Economics Working Papers 24184, Royal Veterinary and Agricultural University, Food and Resource Economic Institute.*

Regassa, Mekdim D., Mohammed B. Degnet, and Mequanint B. Melesse. 2023. “Access to Credit and Heterogeneous Effects on Agricultural Technology Adoption: Evidence From Large Rural Surveys in Ethiopia.” *Canadian Journal of Agricultural Economics/Revue Canadienne D Agroeconomie* 71 (2): 231–53. https://doi.org/10.1111/cjag.12329.

Reserve Bank of India. 1954. All India Rural Credit Survey Report. Accessed here.

Reserve Bank of India. 2009a. Master Circular: Interest Rates on Advances. Accessed here.

Reserve Bank of India. 2009b. Report of the Working Group on Benchmark Prime Lending Rate. Accessed here.

Reserve Bank of India. 2017. Guidelines on Kisan Credit Cards. Accessed here.

Reserve Bank of India. 2019. Report of the Internal Working Group to Review Agricultural Credit. Accessed here.

Resnick, Sidney. 1999. "*A Probability Path*" Beijing : World Publishing Corporation. https://archive.org/details/probabilitypathm0000resn.

Reyes-Cabrera, Joel, Curtis B. Adams, Jackson Nielsen, and John E. Erickson. 2023. "Yield, Nitrogen, and Water-use Efficiency of Grain Sorghum With Diverse Crown Root Angle." *Field Crops Research* 294 (March): 108878. https://doi.org/10.1016/j.fcr.2023.108878.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41. https://doi.org/10.2307/2335942.

Sial, Maqbool H., and Michael R. Carter. 1996. "Financial Market Efficiency in an Agrarian Economy: Microeconometric Analysis of the Pakistani Punjab." *The Journal of Development Studies* 32 (5): 771–98. https://doi.org/10.1080/00220389608422439.

Shukla, Sumedha & Arora, Gaurav. 2022. "A Rank Similarity Test for Quantile Treatment Effects in Conjunction With Propensity Score Matching: An Application to Heterogeneous Crop Yield Impacts of Agricultural Credit." *AgEcon Search (University of Minnesota, USA)*. https://doi.org/10.22004/ag.econ.322569.

Siamwalla, Ammar, Chirmsak Pinthong, Nipon Poapongsakorn, Ploenpit Satsanguan, Prayong Nettayarak, Wanrak Mingmaneenakin, and Yuavares Tubpun. 1990. "The Thai Rural Credit System: Public Subsidies, Private Information, and Segmented Markets." *The World Bank Economic Review* 4 (3): 271–95. https://doi.org/10.1093/wber/4.3.271.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets With Imperfect Information." *The American Economic Review* 71 (3): 393–410. https://www.jstor.org/stable/1802787.

Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25 (1): 1–21. https://doi.org/10.1214/09-sts313.

Swain, Ranjula Bali. 2002. "Credit Rationing in Rural India." Journal of Economic Development, Chung-Ang University. 27(2): 1-20 https://ideas.repec.org/a/jed/journl/v27y2002i2p1-20.html.

Townsend, Robert M. 1994. "Risk and Insurance in Village India." *Econometrica* 62 (3): 539. https://doi.org/10.2307/2951659.

Wolfstetter, Elmar. 1970. "Topics in Microeconomics: Industrial Organization, Auctions and Incentives." *Cambridge University Press*. https://doi.org/10.1017/cbo9780511625787.

Wollburg, Philip, Marco Tiberti, and Alberto Zezza. 2020. "Recall Length and Measurement Error in Agricultural Surveys." *Food Policy* 100 (December): 102003. https://doi.org/10.1016/j.foodpol.2020.102003.

Zeeshan, None, Mohd Arshad Khan, Tuba Shamim, and Noorsaba Khatoon. 2025. “Caste Disparities in Institutional Credit Access and Their Implications for Sustainable Development in Rural India.” *Discover Sustainability* 6 (1). https://doi.org/10.1007/s43621-025-01805-6.

Zinman, Jonathan. 2009. “WHERE IS THE MISSING CREDIT CARD DEBT? CLUES AND IMPLICATIONS.” *Review of Income and Wealth* 55 (2): 249–65. https://doi.org/10.1111/j.1475-4991.2009.00321.x.

2.9 Supplementary Information

Section 1: Literature Summary

Reference	Study Years	Geography	Endogeneity Resolution (#)	Research Question	Rationing Elicitation	Credit Impact and Conclusion about Rationing
Carter (1989)	1981	Nicaragua	ESR	Assess impact of credit on small farm food production.	Indirect	Weak evidence of positive credit impacts, confirmed presence of credit constraints.
Feder et al. (1990)	1987	China	ESR	Assess effect of credit on output supply.	Direct	~37% households reported being credit constrained. Positive credit effect on output of constrained households.
Jappelli (1990)	1983	United States	N.A.	Assess determinants of credit constraints in US economy.	Direct	19% households report being rationed. Wealth and age significantly matter for rationing.
Sial and Carter (1996)	1987-88	Pakistani Punjab	ESR	Examine financial market inefficiency for small farms by estimating shadow price of capital.	Indirect	For a random individual, ~200% rate of return on first rupee borrowed, implied significant credit constraints.
Kochar (1997)	1981-82	India	ESR, IV	Examine land rental choices to assess if formal credit constraints bind.	Indirect	No significant role of credit on amount of land leased or tenurial status after accounting for resource endowments, suggests non-binding credit constraints.
Deininger and Olinto (2000)	1993/94-1994/95	Zambia	Panel Data, IV	Examine drivers of modest effect of liberalization in agriculture.	Indirect	Suboptimal input use despite positive marginal returns indicates significant credit constraints. Credit access increased cultivated area and TFP.
Liu and Zhang (2000)	1990	China	IV	Examine drivers of inter-farm efficiency differences.	Indirect	Credit increases output but impact is lower in high-income regions. Suggests weaker constraints in wealthier regions.
Diagne and Zeller (2001)	1995 (3 rounds)	Malawi	Differentiate borrowing ability and access using credit limits data.	Examine credit impact on income and food security.	Direct	>40% households reported being credit constrained. Credit impact on net farm income is insignificant due to lack of supporting infrastructure (e.g., rainfall).

Reference	Study Years	Geography	Endogeneity Resolution	Research Question	Rationing Elicitation	Credit Impact and Conclusion about Rationing
Khandker and Faruqee (2003)	1995	Pakistan	IV	Assess the cost-effectiveness of subsidized rural credit delivery.	N.A.	Positive credit impacts on consumption, vary by landholding size.
Foltz (2004)	1995	Tunisia	ESR	Investigate impact of credit constraints on profitability and investments.	Direct	45% households reported being credit constrained. Constraints reduce profits but not investments.
Petrick (2004)	2000	Poland	Inverse-Mills Ratio	Investigate credit rationing on Polish farms.	Direct	42% households reported being credit constrained. Marginal willingness to pay for credit was 209% net of principal for constrained households.
Petrick (2004)	2000	Poland	Inverse Mills Ratio	Analyse effect of credit access on farm investment.	Direct	45% farmers reported being credit rationed. Marginal effect of credit on investment is less than 1, suggest diversion of credit for non-productive purposes, particularly for smaller loans.
Banerjee and Duflo (2004)	1996-2002	India	Experimental, Difference-in-Differences	Estimate credit constraints incidence using variation in access to targeted lending.	Indirect	Argue that both constrained and unconstrained firms can absorb targeted lending but only the former would use it to expand production. Credit increases sales and profits, i.e., constraints bind.
Blancard et al. (2006)	1994 - 2001	France	Profit Function, IV	Investigates rationing incidence among French farmers.	Indirect	Observed expenditures are lower than optimal, suggests constraints bind. Average shadow price of capital exceeds interest rates.
De Mel, McKenzie and Woodruff (2008)	2005 (4 rounds)	Sri Lanka	Experimental	Measure returns to capital for micro-enterprises.	Indirect	Average real return to credit is 5.7% per month, higher than the market interest rate, suggests binding credit constraints.
Guirkinger and Boucher (2008)	1997, 2003	Peru	ESR, Panel Data	Evaluate the performance of rural credit market in Peru.	Direct	56% and 43% households reported being credit constrained across the study years. Relaxing credit constraints would increase output by 26%.

Reference	Study Years	Geography	Endogeneity Resolution	Research Question	Rationing Elicitation	Credit Impact and Conclusion about Rationing
McKenzie and Woodruff (2008)	2005, 2006 (Panel)	Mexico	Experimental	Assess the impact of capital constraints on micro-enterprise earnings.	Indirect	Returns to credit were 20-33% per month i.e., about 3-5 times higher than market interest rate, suggests rationing.
Briggeman et al. (2009)	2004-05	United States	PSM	Assess production impact of credit constraints for (non-) farm sole proprietorships.	Direct	5% farm households reported being credit constrained. Credit constraints lower production value by ~3%.
Karlan and Zinman (2009)	2004	South Africa	Experimental	Assess the impact of consumer credit expansion.	Indirect	Expanding credit supply improved household welfare, suggests binding credit constraints.
Kumbhakar and Bokusheva (2009)	1999-2003	Russia	Indirect Production Function, IV	Examine output and technical efficiency losses due to budget constraints.	Indirect	Observed expenditures lower than optimal i.e., constraints bind for majority of farms; potential output loss of 20%. Constraint severity varies by farmer attributes (e.g., ownership)
Liverpool and Winter-Nelson (2010)	1994-2004	Ethiopia	Panel Data, IV	Examines the impact of formal credit on technology use, consumption and asset growth.	Indirect	No impact of credit on technology use and consumption for some households but strictly positive impacts on others. Implies heterogeneity in constraints faced by households. Suggest asset-based classifications for program targeting.
Islam et al. (2011)	2009	Bangladesh	ESR	Investigate credit constraints impact on farm efficiency.	Direct	No significant credit impact on rice farmers' efficiency.
Dong et al. (2012)	2008	China	ESR	Examine impact of credit constraints on household productivity and income.	Direct	~17% households reported being credit constrained. Constraints lower productivity (income) by 32% (23%).
Ali and Deininger (2012)	2007	Ethiopia	ESR	Study impact of rationing in semi-formal credit markets on resource allocation and crop productivity.	Direct	~60% households reported being credit constrained. Eliminating constraints improved productivity by 11% in productive regions but no effect in drought-prone regions.
Ciaian et al. (2012)	2005	CEE Transition Countries	PSM	Studies impact of credit access on farm inputs and efficiency.	Indirect	Credit access raises total factor productivity and variable input use, but not cultivated area, and reduces labour use. Suboptimal use of profitable inputs reveals asymmetric credit constraints.

Reference	Study Years	Geography	Endogeneity Resolution	Research Question	Rationing Elicitation	Credit Impact and Conclusion about Rationing
Kumar et al. (2013)	2008-09	India, China	N.A.; Self-Reported Impacts	Investigate impact of credit constraints on capital formation, input application, consumption smoothing etc.	Direct	31% (41%) Chinese (Indian) households reported loan rejections. Binding credit constraints adversely affect production and livelihood.
Ali, Deininger and Duponchel (2014)	2011	Rwanda	ESR	Identify credit constraints among rural households and their impact on efficiency.	Direct	Almost *all* (71%) households report being constrained in formal (semi-formal) sector. Elimination of credit constraints could increase output by 17%.
Karlan et al. (2014)	2008-12	Ghana	Experimental	Examine agricultural investment decisions after relaxing credit and risk constraints.	Indirect	Cash grants have a smaller effect on farm investment relative to rainfall insurance grant, suggests liquidity constraints are not as binding as typically assumed.
Abate et al. (2016)	2012	Ethiopia	PSM	Examine the effect of formal credit on agricultural technology adoption.	Indirect	Credit access increases adoption and extent of technology use. Heterogenous impacts by farm size and institution type. Sub-optimal adoption of profitable technology suggests binding constraints.
Kumar et al. (2017)	2013	India	IV	Examine impact of credit on farm income and household expenditures.	N.A.	Credit has positive effects on farm income and consumption expenditure.
Sekyi, Demanban and Honya (2019)	2012	Ghana	ESR	Assess informal credit impact on farm productivity.	N.A.	Credit access had a positive impact on yield.
Jimi et al. (2019)	2012, 2014	Bangladesh	Experimental	Assess credit impact on rice productivity.	Indirect	Credit constraint relaxation increased yields by 14%. 11% due to technological change and 3% due to technical efficiency.
Kumar et al. (2020)	2018	India	ESR	Assess credit impact on household annual income.	N.A.	Credit access is strongly associated with household wealth and demography. Positive effect on income.

Reference	Study Years	Geography	Endogeneity Resolution	Research Question	Rationing Elicitation	Credit Impact and Conclusion about Rationing
Kumar, Sonkar and Saroj (2020)	2018	India	ESR	Assess formal and informal credit impacts on farm income and crop yields.	N.A.	Credit access increased household income and crop yields. Heterogeneity by education and caste but neutral to scale.
Missiame et al. (2021)	2016	Ghana	ESR	Assess impact of credit on technical efficiency of Cassava farmers.	N.A.	Credit access significantly improves technical efficiency.
Hutchins (2023)	1920-40	United States	Panel Data, Fixed Effects	Analyse impact of production credit associations on agricultural yield and input use.	Indirect	Credit access significantly increased crop revenue per acre (7-14%), corn yields (9%) and tractor use (1-2%). Positive effects suggest binding credit constraints.
Beaman et al. (2023)	2010-12	Mali	Experimental	Examine if selection into lending is predictive of heterogenous returns to capital.	Indirect	Returns to capital are higher among borrowers relative to non-borrowers. Evidence suggests credit rationing as high-return poor households remain excluded from borrowing.
Haryanto et al. (2023)	2014	Indonesia	PSM	Assess credit impact on maize farm productivity.	N.A.	Credit access improves farm productivity with effects more pronounced for formal credit.
Regassa et al. (2023)	2013	Africa	IV	Assess credit impact on adoption and use intensity of chemical fertilizers.	Direct	~73% households report being credit constrained. Credit access increases adoption and intensity of chemical fertilizers use. Heterogeneity by source of credit and credit rationing status.

Table 1 (SI Section 6): Overview of literature on impact of credit and constraints elicitation. **Source:** Author's Compilation.

Notes:

- **(#) Abbreviations:** ESR: Endogenous Switching Regression Framework; IV: Instrumental Variable Framework**;** PSM: Propensity Score Matching Framework**;** N.A. : Not Applicable
- This review comprises of empirical studies examining credit (constraint) impacts on economic outcomes at the farm/micro level. Papers examining group lending are outside the scope.
- **Direct elicitation** exploits a series of questions on credit market participation, including borrowing needs, experience, and overall perceptions to assign constrained/ unconstrained status to a household (see Figure 1 (SI Section 7)). Specific questions may differ across studies depending on context.

Figure 1 (SI Section 1): Direct Elicitation Approach (Source: Ali et al. (2014))

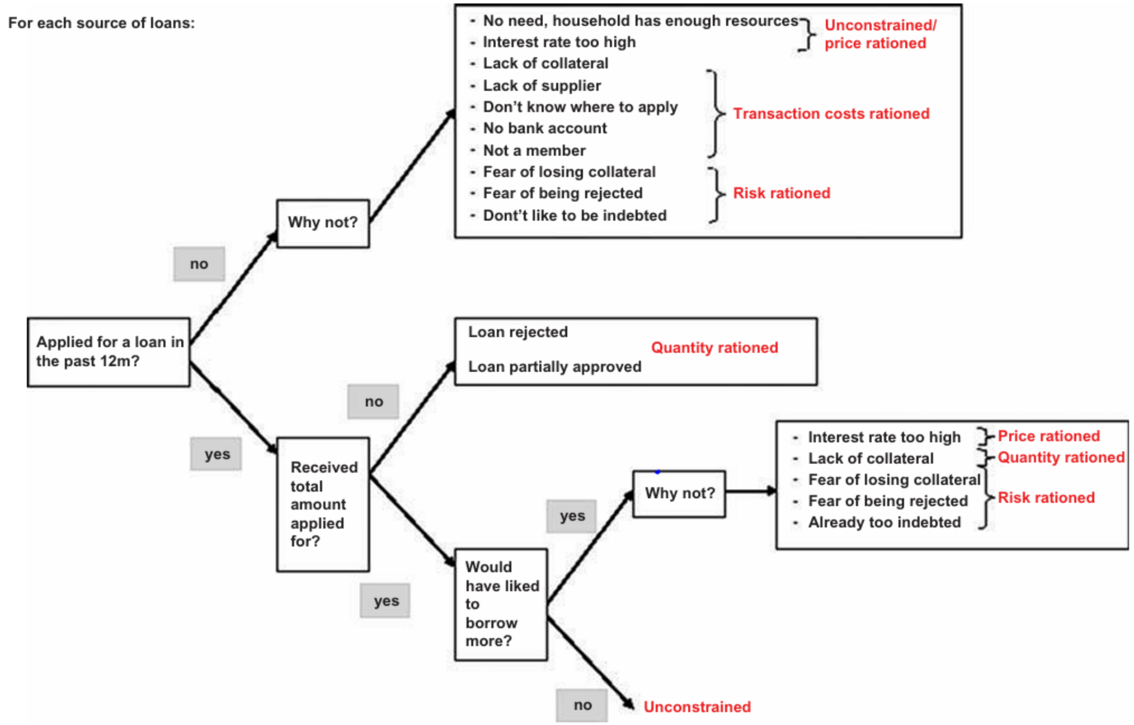


Figure 2. The direct elicitation approach.
Source: World Bank survey 2011.

References (SI, Section 1):

Abate, Gashaw Tadesse, Shahidur Rashid, Carlo Borzaga, and Kindie Getnet. 2016. "Rural Finance and Agricultural Technology Adoption in Ethiopia: Does the Institutional Design of Lending Organizations Matter?" *World Development* 84 (April): 235–53. https://doi.org/10.1016/j.worlddev.2016.03.003.

Ali Daniel & Deininger K. 2012. "Causes and Implications of Credit Rationing in Rural Ethiopia : The Importance of Spatial Variation." *Ideas.Repec.Org*, June. https://ideas.repec.org/p/wbk/wbrwps/6096.html.

Ali, Daniel., Deininger, K., & Duponchel, M. (2014). Credit Constraints and Agricultural Productivity: Evidence from rural Rwanda. *The Journal of Development Studies*, 50(5), 649–665. https://doi.org/10.1080/00220388.2014.887687.

Banerjee, Abhijit, and Esther Duflo. 2014. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program." *The Review of Economic Studies* 81 (2): 572–607. https://doi.org/10.1093/restud/rdt046.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2023. "Selection Into Credit Markets: Evidence From Agriculture in Mali." *Econometrica* 91 (5): 1595–1627. https://doi.org/10.3982/ecta18916.

Blancard, Stéphane, Jean-Philippe Boussemart, Walter Briec, and Kristiaan Kerstens. 2006. "Short- and Long-Run Credit Constraints in French Agriculture: A Directional Distance Function Framework Using Expenditure-Constrained Profit Functions." *American Journal of Agricultural Economics* 88 (2): 351–64. https://doi.org/10.1111/j.1467-8276.2006.00863.x.

Briggeman, Brian C., Charles A. Towe, and Mitchell J. Morehart. 2008. "Credit Constraints: Their Existence, Determinants, and Implications for U.S. Farm and Nonfarm Sole Proprietorships." *American Journal of Agricultural Economics* 91 (1): 275–89. https://doi.org/10.1111/j.1467-8276.2008.01173.x.

Carter, Michael R. 1989. "The Impact of Credit on Peasant Productivity and Differentiation in Nicaragua." *Journal of Development Economics* 31 (1): 13–36. https://doi.org/10.1016/0304-3878(89)90029-1.

Ciaian, Pavel, Jan Fałkowski, and D'Artis Kancs. 2012. "Access to Credit, Factor Allocation and Farm Productivity." *Agricultural Finance Review* 72 (1): 22–47. https://doi.org/10.1108/00021461211222114.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence From a Field Experiment." *The Quarterly Journal of Economics* 123 (4): 1329–72. https://doi.org/10.1162/qjec.2008.123.4.1329.

Deininger, Klaus & Olinto Pedro. 2000. "Why Liberalization Alone Has Not Improved Agricultural Productivity in Zambia : The Role of Asset Ownership and Working Capital Constraints." *Ideas.Repec.Org*, March. https://ideas.repec.org/p/wbk/wbrwps/2302.html.

Diagne, Aliou & Zeller Manfred. 2001. "Access to Credit and Its Impact on Welfare in Malawi." *Ideas.Repec.Org*. https://ideas.repec.org/p/fpr/resrep/116.html.

Dong, Fengxia, Jing Lu, and Allen M. Featherstone. 2012. "Effects of Credit Constraints on Household Productivity in Rural China." *Agricultural Finance Review* 72 (3): 402–15. https://doi.org/10.1108/00021461211277259.

Feder, Gershon, Lawrence J. Lau, Justin Y. Lin, and Xiaopeng Luo. 1990. "The Relationship Between Credit and Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium." *American Journal of Agricultural Economics* 72 (5): 1151–57. https://doi.org/10.2307/1242524.

Foltz, J. 2004. "Credit Market Access and Profitability in Tunisian Agriculture." *Agricultural Economics* 30 (3): 229–40. https://doi.org/10.1016/j.agecon.2002.12.003.

Guirkinger, Catherine, and Stephen R. Boucher. 2008. "Credit Constraints and Productivity in Peruvian Agriculture." *Agricultural Economics* 39 (3): 295–308. https://doi.org/10.1111/j.1574-0862.2008.00334.x.

Haryanto, Tri, Wahyu Wisnu Wardana, Iqram Ramadhan Jamil, Annisaa Rizky Dwi Brintanti, and Kabiru Hannafi Ibrahim. 2023. "Impact of Credit Access on Farm Performance: Does Source of Credit Matter?" *Heliyon* 9 (9): e19720. https://doi.org/10.1016/j.heliyon.2023.e19720.

Hutchins, Jared. 2022. "The US Farm Credit System and Agricultural Development: Evidence From an Early Expansion, 1920–1940." *American Journal of Agricultural Economics* 105 (1): 3–26. https://doi.org/10.1111/ajae.12290.

Jappelli, Tullio. 1990. "Who Is Credit Constrained in the U. S. Economy?" *The Quarterly Journal of Economics* 105 (1): 219. https://doi.org/10.2307/2937826.

Jimi, Nusrat Abedin, Plamen V. Nikolov, Mohammad Abdul Malek, Subal Kumbhakar. 2019. "The Effects of Access to Credit on Productivity: Separating Technological Changes From Changes in Technical Efficiency." *Journal of Productivity Analysis* 52 (1–3): 37–55. https://doi.org/10.1007/s11123-019-00555-8.

Karlan, Dean, and Jonathan Zinman. 2009. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies* 23 (1): 433–64. https://doi.org/10.1093/rfs/hhp092.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. "Agricultural Decisions After Relaxing Credit and Risk Constraints." *The Quarterly Journal of Economics* 129 (2): 597–652. https://doi.org/10.1093/qje/qju002.

Khandker, S. 2003. "The Impact of Farm Credit in Pakistan." *Agricultural Economics* 28 (3): 197–213. https://doi.org/10.1016/s0169-5150(03)00017-3.

Kochar, Anjini. 1997. "An Empirical Investigation of Rationing Constraints in Rural Credit Markets in India." *Journal of Development Economics* 53 (2): 339–71. https://doi.org/10.1016/s0304-3878(97)00020-5.

Kumbhakar, S. C., and R. Bokusheva. 2009. "Modelling Farm Production Decisions Under an Expenditure Constraint." *European Review of Agricultural Economics* 36 (3): 343–67. https://doi.org/10.1093/erae/jbp031.

Kumar, Anjani, Ashok K. Mishra, Sunil Saroj, and P.K. Joshi. 2017. "Institutional Versus Non-institutional Credit to Agricultural Households in India: Evidence on Impact From a National Farmers' Survey." *Economic Systems* 41 (3): 420–32. https://doi.org/10.1016/j.ecosys.2016.10.005.

Kumar, Anjani, Ashok K. Mishra, Vinay K. Sonkar, and Sunil Saroj. 2020. "Access to Credit and Economic Well-Being of Rural Households: Evidence From Eastern India." *Journal of Agricultural and Resource Economics* 45 (1): 145–60. https://www.jstor.org/stable/26869031.

Kumar, Anjani, Vinay K. Sonkar, and Sunil Saroj. 2020 “Access to Credit in Eastern India. Implications for the Economic Well-being of Agricultural Households.” *Economic and Political Weekly* 55 (21): 1-21. https://www.epw.in/journal/2020/21/special-articles/access-credit-eastern-india.html.

Liu, Zinan, and Juzhong Zhuang. 2000. “Determinants of Technical Efficiency in Post-Collective Chinese Agriculture: Evidence From Farm-Level Data.” *Journal of Comparative Economics* 28 (3): 545–64. https://doi.org/10.1006/jcec.2000.1666.

Liverpool, Lenis Saweda O., and Alex Winter-Nelson. 2009. “Poverty Status and the Impact of Formal Credit on Technology Use and Wellbeing Among Ethiopian Smallholders.” *World Development* 38 (4): 541–54. https://doi.org/10.1016/j.worlddev.2009.11.006.

McKenzie, David, and Christopher Woodruff. 2008. “Experimental Evidence on Returns to Capital and Access to Finance in Mexico.” *The World Bank Economic Review* 22 (3): 457–82. https://www.jstor.org/stable/40282284.

Missiame, Arnold, Rose A. Nyikal, and Patrick Irungu. 2021. “What Is the Impact of Rural Bank Credit Access on the Technical Efficiency of Smallholder Cassava Farmers in Ghana? An Endogenous Switching Regression Analysis.” *Heliyon* 7 (5): e07102. https://doi.org/10.1016/j.heliyon.2021.e07102.

Petrick, Martin. 2004. “Farm Investment, Credit Rationing, and Governmentally Promoted Credit Access in Poland: A Cross-sectional Analysis.” *Food Policy* 29 (3): 275–94. https://doi.org/10.1016/j.foodpol.2004.05.002.

Petrick, M. 2004. “A Microeconometric Analysis of Credit Rationing in the Polish Farm Sector.” *European Review of Agricultural Economics* 31 (1): 77–101. https://doi.org/10.1093/erae/31.1.77.

Regassa, Mekdim D., Mohammed B. Degnet, and Mequanint B. Melesse. 2023. “Access to Credit and Heterogeneous Effects on Agricultural Technology Adoption: Evidence From Large Rural Surveys in Ethiopia.” *Canadian Journal of Agricultural Economics/Revue Canadienne D Agroeconomie* 71 (2): 231–53. https://doi.org/10.1111/cjag.12329.

Sekyi, Samuel, Paul Bata Domanban, and George Kwame Honya. 2019. “The Impact of Informal Credit on Rural Agricultural Productivity in the Savannah Ecological Zone of Ghana.” *African Journal of Economic and Management Studies* 11 (2): 301–15. https://doi.org/10.1108/ajems-03-2019-0121.

Sial, Maqbool H., and Michael R. Carter. 1996. “Financial Market Efficiency in an Agrarian Economy: Microeconometric Analysis of the Pakistani Punjab.” *The Journal of Development Studies* 32 (5): 771–98. https://doi.org/10.1080/00220389608422439.

Section 2: Data and Context

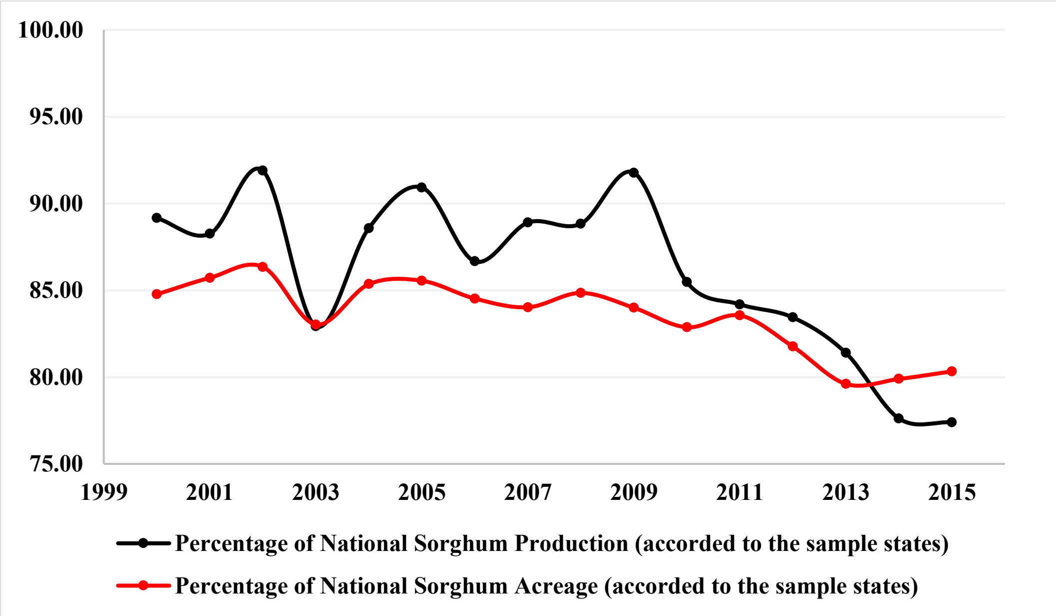


Figure 1: Percentage of National Sorghum Production and Acreage accorded to the sample states. | **Data Source(s):** Agricultural Statistics at a Glance, Directorate of Economics and Statistics, Ministry of Agriculture and Farmers' Welfare (MoAFW)

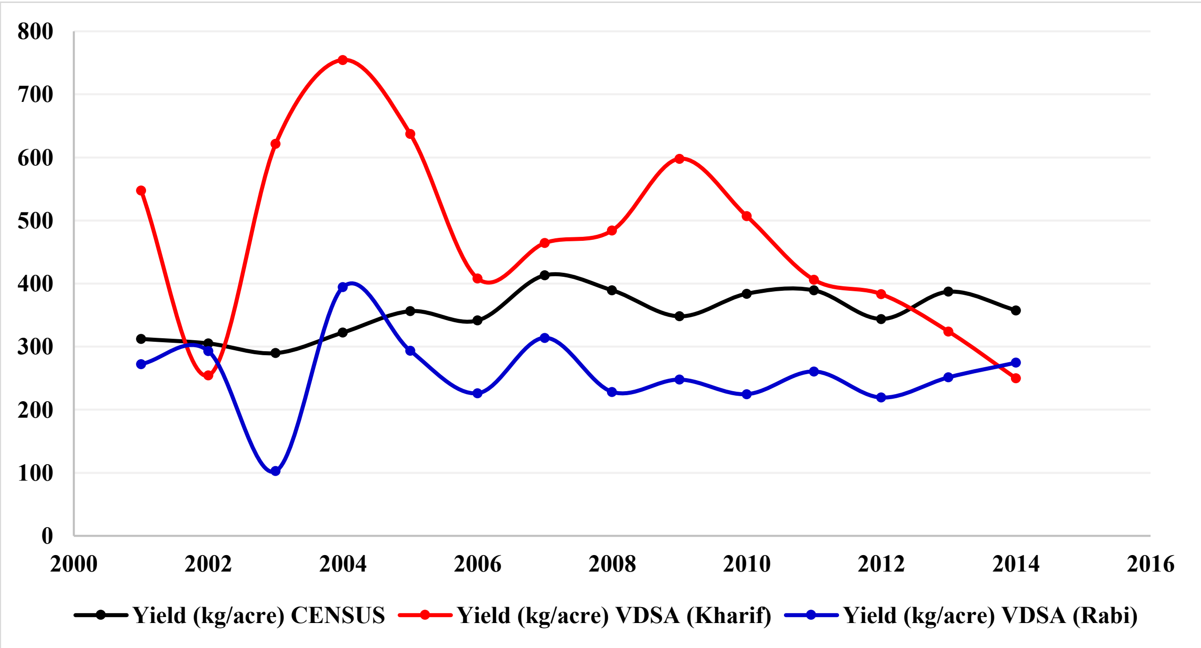


Figure 2: Average Annual Sorghum Yield (kg/acre) – National and VDSA Sample.

Data Source(s): Agricultural Statistics at a Glance, Directorate of Economics and Statistics, Ministry of Agriculture and Farmers' Welfare (MoAFW)

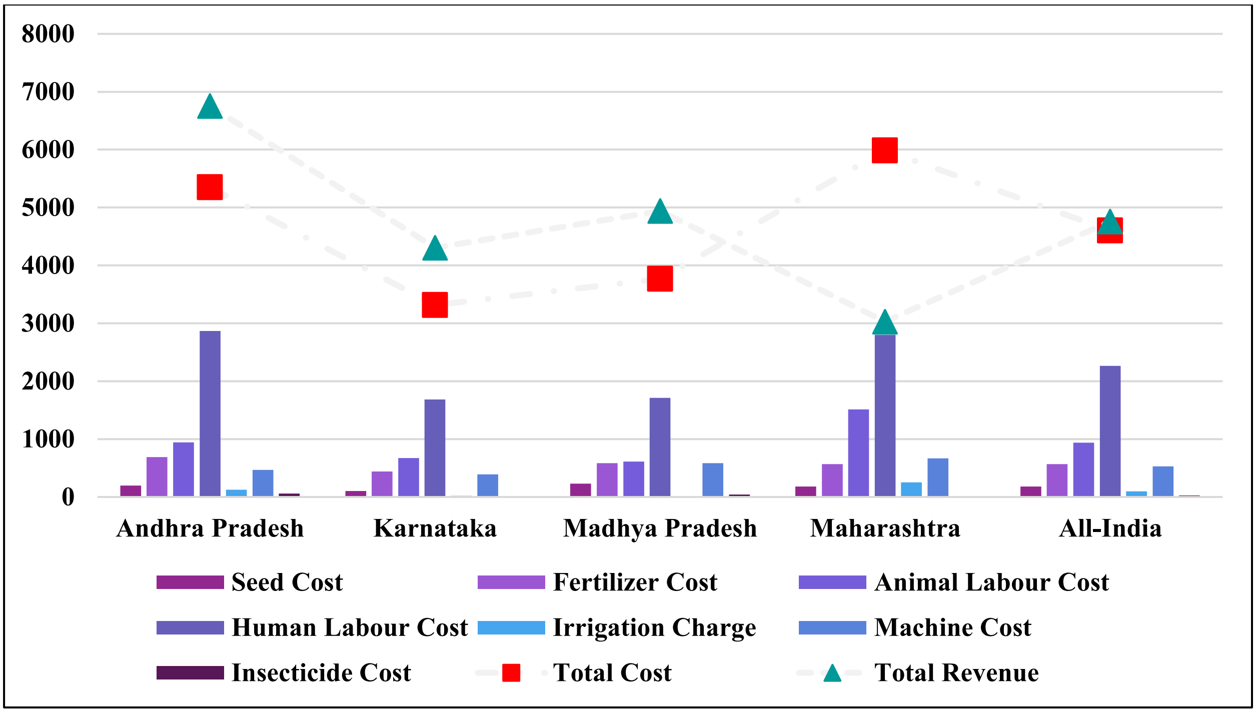


Figure 3: Breakup of Cost of Production of Sorghum (Averaged over 2001-'14). **Data Source(s):** Comprehensive Analysis of Cost of Production Surveys, Ministry of Agriculture and Farmers' Welfare (MoAFW)

Section 3:

Section 3A: Data Structure Similarities between Sorghum and Other Major Crops in the VDSA Dataset

Sorghum and Other Major Crops:

In this section, we outline the similarities in the data structure between Sorghum and other major crops in the VDSA data.

First, define major crops as the set of crops accounting for ~60% of cropping instances and over 65% of cropped area over the study period. Table 1(S2) lists the major crops and the percentage of cropping instances and area they account for.

Crop Name	Instances of Cultivation (Percentage of Total Cropping Instances)	Area under Cultivation (Percentage of Total Area under Cropping)
Sorghum	7.5	12.4
Cotton	6.8	12.2
Paddy	20.4	11.3
Wheat	9.9	10.8
Soybean	4.4	10.0
Pigeonpea	11.0	8.1
All other crops individually account for less than five percent of cropping instances and are kept out of this analysis.		

Table 1(S3): Major crops in the VDSA Dataset

Table 2(S2) provides the household-level distribution of cropping frequency across all major crops. Evidently, for all crops except wheat and paddy, the median household cultivates the crop in only three years during the study period. Even in case of wheat and paddy, the median household cultivates the crop in four and five years during the study period respectively.

Cropping Frequency	Sorghum	Cotton	Paddy	Wheat	Soybean	Pigeonpea
1	125	98	76	126	106	205
2	70	64	56	105	16	112
3	59	51	76	56	14	97
4	47	41	102	72	16	63
5	40	30	287	128	18	85
6	41	27	43	53	27	85
7	35	25	12	19	17	31
8	34	15	8	14	28	20
9	11	6	5	11	9	17
10	17	12	6	5	.	12
11	7	7	6	1	2	14
12	6	11	8	3	5	8
13	14	14	5	1	1	9
14	14	16	5	.	.	18

Table 2(S3): Cropping frequency across major crops in the VDSA dataset.

Table 3(S2) provides the year-wise count of first-time formal borrowers for all major crops. The highlighted rows (in red) for each crop represent a year where the count of borrowing households is higher than the potential counterfactuals available in that year, i.e., the counterfactual data is insufficient. For all the crops except Pigeonpea, we observe insufficiency of counterfactual data in at least four out of the fourteen study periods. A similar pattern holds for informal borrowing.

Year	Count of First-Time Borrowers (Formal)					
	Sorghum	Cotton	Paddy	Wheat	Soybean	Pigeonpea
2001	59	39	7	37	2	46
2002	23	16	8	10	2	33
2003	16	15	11	1	2	34
2004	13	13	10	1	3	19
2005	34	25	16	26	3	41
2006	22	11	5	21	11	15
2007	20	11	3	17	19	18
2008	13	19	7	4	30	28
2009	35	16	4	39	15	32
2010	18	16	80	49	6	30
2011	10	17	54	44	7	20
2012	15	4	32	16	8	8
2013	7	17	11	13	7	16
2014	5	7	11	15	5	9

Table 3(S3): Year-wise count of First-Time Borrowers across major crops in the VDSA dataset.

Finally, Table 4(S2) provides the percentage of households which have at least one year of pre-borrowing and post-borrowing data across major crops to check for the possibility of a panel data analysis. However, similar to Sorghum, such a strategy would lead to major data losses.

Crop Name	Households with at least 1 year of pre-borrowing & post-borrowing data (Percentage of total borrowers: Formal)	Households with at least 1 year of pre-borrowing & post-borrowing data (Percentage of total borrowers: Informal)
Paddy	23	32
Wheat	18	22
Soybean	15	16
Sorghum	25	33
Pigeonpea	22	27
Cotton	21	25

Table 4(S3): Percentage of households with at least 1 year of pre-borrowing & post-borrowing data across major crops in the VDSA dataset.

Section 3B: Data Construction

We will now provide details on data construction.

Constructing Weather Data from IMD grids

To construct this data, we use year-wise 0.25°-by-0.25° grids of daily total rainfall (in mm.) and 1°-by-1° grids of daily average temperature (in degree celsius) between 2001-'14 provided by IMD.

First, for each village we find the four nearest grid points (see Figure 2). On-average, the temperature grid points are located ~55 kilometres away from the village centroid, while the rainfall grid points are ~15 kilometres away from the village centroid. Next, at each grid point in each year, we obtain the total rainfall and average temperature in the Kharif and Rabi months. Finally, we construct the village-level, season-wise, yearly rainfall and temperature as an average of the total rainfall and average temperature recorded on the four nearest grid points respectively.

Constructing Plot-Level Data from Landholding File

We combine plot-level information on crop inputs and output for each season-year with information on household-level borrowings and demographic characteristics for the year. Further, we designed and implemented an algorithmic merging procedure in order to incorporate information on plot-level soil and land characteristics - plot slope, soil depth, and soil quality (as indicated by the alkalinity and erosivity of soil) - from the landholding file. This procedure was necessary because a merge based on the household's plot identifier in a year (as given in the file) only provided landholding information for *less than 40 percent* of the entries in the crop output file. We were able to match ~98 percent of sorghum entries in the crop output file with the corresponding landholding information. For the remaining 2 percent data, we substituted missing information at the plot-level with household-level aggregates.

As part of the algorithm, we broke down the merging process into six steps. In the first step, we constructed a plot identifier which corresponded to the plot code when available, and was substituted with plot name in the cases where plot code was missing. On the basis of this revised plot identifier, we performed a many-to-one[32] merge between the crop output and landholding files respectively which furnished matches for ~36 percent of the data. In the second step, we revised the plot identifier generated in the first step to include only the first letter of plot code (in the crop output file) to account for the plot naming nomenclature followed in the VDSA data collection whereby a plot was divided into sub-plots or sub-sub-plots based on the cultivated crop(s) or irrigation status (Binswanger & Jodha 1978). Since the sub-plots or sub-sub-plots are derived from the same main plot, we extracted the soil quality information for these subplots based on the available information for the main plot. A merge on the basis of this revised identifier yielded a match for ~40.6 percent of data.

In the third step, we revised the plot identifier generated in the first step[33] to remove any information on plot serial number which was rendered redundant because plot serial number information was only available in the landholding file and <u>not</u> the crop output file (VDSA 2013). Merging based on this revised identifier provided matches for ~3.5 percent of the data. In the fourth step, we accounted for cases where arbitrary symbols were used while concatenating plot codes with sub-plot codes or serial numbers by retaining only the initial portion of the plot identifier string. Based on this revised identifier, we merged ~2.9 percent of the original data. In the fifth step, we constructed an alternate plot identifier which corresponded to the plot name when available, and was substituted with plot code in cases where plot name was missing. This alternate plot identifier yielded merges for ~0.5 percent of the data.

In the sixth step, for the remaining 16 percent of data, we undertook a manual merging exercise which furnished matches for ~84 percent of this leftover data (i.e., ~14 percent of the original data). As part of the manual merging, we were able to incorporate landholding information in cases where: (a) the plot names were spelled inconsistently across the two files but were phonetically similar (for example *chalka, chhalka,* and *chalaka* all refer to the same plot); (b) plot names were recorded with vernacular identifiers as part of the plot name in one

[32] Note that at each step of the merging algorithm, before merging entries from the crop output file with the landholding file, we removed any rows in the landholding file which were <u>duplicated based on plot identifier in a season-year</u> to avoid erroneous merges and <u>always</u> relied on a <u>many-to-one</u> merge between the two files. The information from these deleted, duplicated rows - if useful - was incorporated at the sixth step of the merging algorithm – i.e., the manual merging stage.

[33] This entailed removing any special characters belonging to the following set: $\{(,),-,0,1,2,3,4,5,6,7,8,9\}$

of the files but not the other (for example, a piece of farm-land can be referred to as *shet/sheth/khet* in languages spoken in Maharashtra); (c) plot codes were same but recorded in the serial number column in the landholding file; and (d) serial number (in the landholding file) was same as the plot code recorded in the crop output file. Often, a combination of these conditions was used to obtain the match hence a manual merging process was deemed most suitable.

Section 4: Testing for Rank Similarity

Rank Invariance

$\alpha_{1\tau} = Q_\tau(Y_{1i_k} \mid \vec{X}'_{i_k t_k}, D_{i_k t_k} = 1) - Q_\tau(Y_{0i_k} \mid \vec{X}'_{j_k t_k}, D_{j_k t_k} = 0)$ represents the impact of credit attainment on the τ^{th} conditional yield quantile which will correspond with the treatment effect for individual households lying in the τ^{th} yield quantile only if the impact of credit attainment for τ^{th} proportion of the population is less than or equal to this difference, i.e., $F^{-1}_{Y_{1,it} - Y_{0,it}}(\tau) = \alpha_{1\tau}$. Due to the non-linearity of the quantile operator, this equivalence requires each household to maintain their rank in the outcome (i.e., yields) distribution regardless of treatment status. This condition is termed *rank invariance* or *rank preservation* and it is a statistically "strong"[34] and untestable assumption because the joint distribution of potential outcomes is unknown (Firpo 2007; Angrist and Pischke 2009). Let $U_0 = F_0(Y)$ and $U_1 = F_1(Y)$ be the potential ranks by treatment status, i.e., $U_0 \sim U(0,1)$, and $U_1 \sim U(0,1)$ by construction. Rank invariance holds for the matched sample if and only if:

$(U_0 \mid \vec{X} = x, u = u) = (U_1 \mid \vec{X} = x, u = u)\ \forall (x,u)$ where $\vec{X}$ and u are the observed and unobserved determinants of yields respectively.

Intuitively, rank invariance implies that the estimated conditional QTT for any quantile can only be interpreted for households located in that quantile if the same households lie in that quantile before and after receiving treatment (Bedoya et al. 2018; Chalak 2019).

Illustrating the consequences of Rank Invariance

To understand rank invariance, consider a hypothetical dataset of five subjects who have received fertilizer subsidy from the government. We want to estimate the effect of fertilizer

---

[34] The intuition for the need for rank invariance is clarified through a numerical and empirical example in SI Section 4. SI Section 5 presents a thought experiment illustrating the statistically restrictive nature of the rank invariance assumption.

subsidy on yields on average as well as yields on the median. Table 1(S4) below presents the data on the subjects' yields when they receive subsidy (column 1) and what their yields would have been had they not received subsidy (column 2)[35]. The average treatment effect is given by the difference in means of the potential outcomes (460 - 280) which is equal to the average of the difference in potential outcomes, i.e., 180[36]. Now, consider the treatment effect at the median which would be given by the difference in the medians of the potential outcomes (500 – 300 = 200) which is higher than the median of the subject-level differences in potential outcomes (i.e., 100).

This is because of the *mobility effect* whereby subject I who would have had the lowest yield (among the subjects) without subsidy, has the highest yield among them when they do receive subsidy. The notion of *mobility effect* captures *shifts in subjects' ranking with and without treatment* and is precisely the source of dissimilarity between the difference-in-quantiles of potential outcomes and the quantile-of-difference in potential outcomes. To reconcile this difference and obtain the unconditional quantile treatment effects, rank invariance assumes that the mobility effect is zero for each subject.

Subject	Yield with Subsidy	Yield without Subsidy	Difference in Potential Outcomes (i.e., yields with and without subsidy)
I	600	100	500
II	500	300	200
III	500	400	100
IV	500	500	0
V	200	100	100
Average	460	280	180
Median	500	300	100

Hypothetical data on potential outcomes (yield) for five subjects.

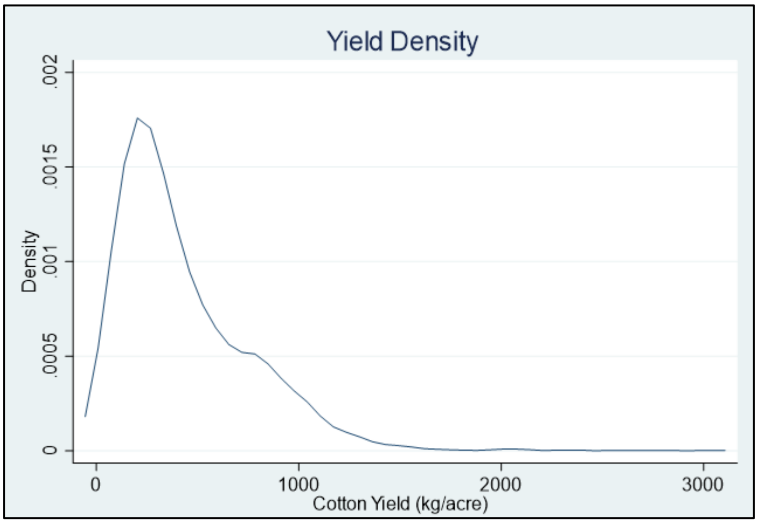
Now, let's illustrate the need for rank invariance through an empirical example. Suppose we have a simple linear regression of the form: $Y_i = \alpha + \beta F_i + \varepsilon_i$ where Y_i is the yield, F_i is

[35] Counterfactual outcomes are hypothetical as no subject can be observed in both treatment and control state.
[36] Follows from the linearity of the expectation operator (Firpo 2007)

the fertilizer application (Diammonium Phosphate (DAP)) per acre for the i^{th} household. First, let's plot the unconditional density of yields.

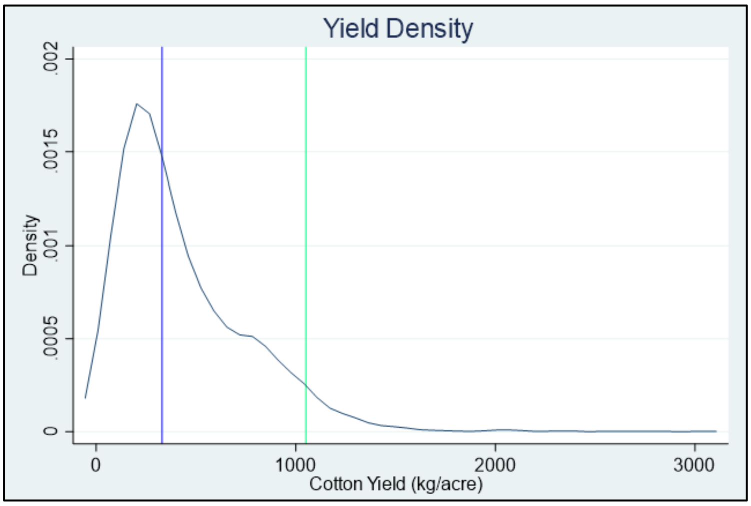
P5	P10	P25	P50	P75	P90	P95	P99
58.67	100.41	195.01	337.12	600	888.29	1050	1425

Cotton Yield(kg/acre) Summary



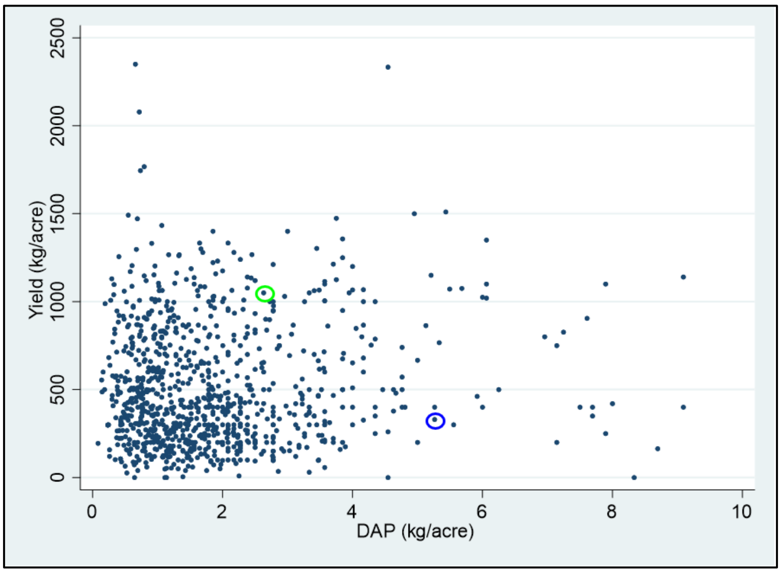
Unconditional Cotton Yield(kg/acre) Density Plot

Next, pick two households B and G where B (yield = 337 kg/acre) belongs to a relatively lower (i.e., 50th) quantile of the unconditional yield distribution and G (yield = 1050 kg/acre) belongs to one of the relatively higher (i.e., 95th) quantiles.



Picking two households B and G

Now we condition our yields on the level of fertilizer application, i.e., for each level of fertilizer application, we obtain a conditional yield distribution. This would be represented as a density plot as above (Figure 2(S4)) but for each level of fertilizer application separately. As we can observe in Figure 3(S4) below, the level of fertilizer application is 5.26 kg/acre for B and 2.38 kg/acre for G (see the blue and green circles respectively):



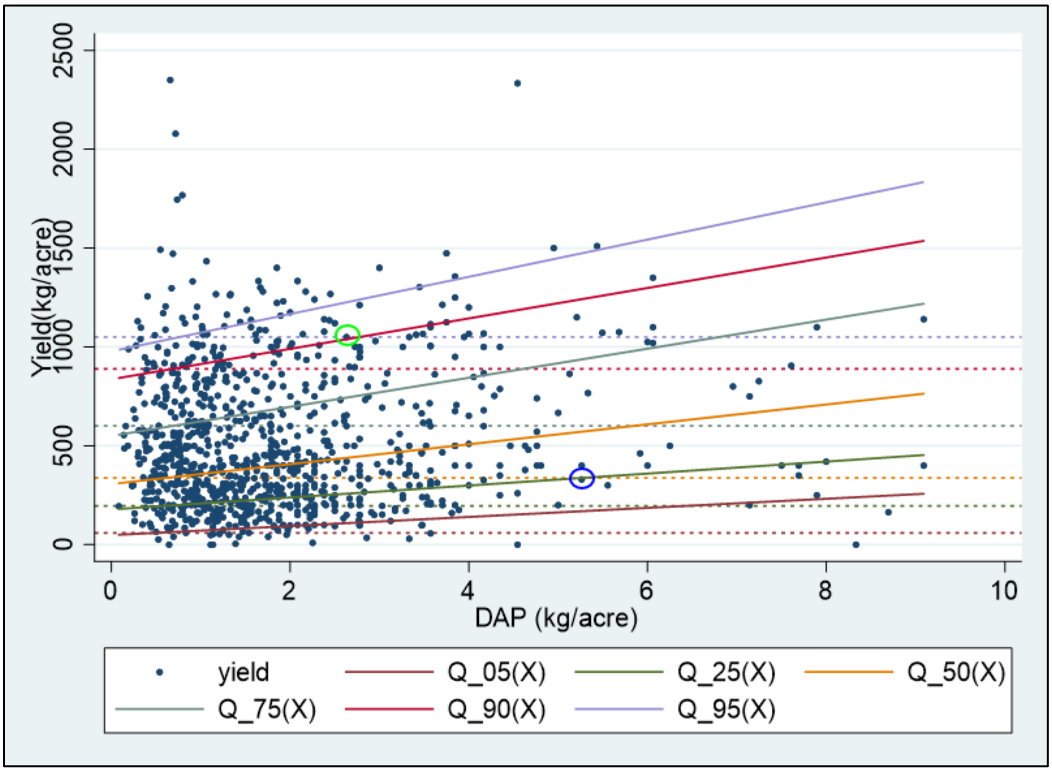
Conditioning Yields on Fertilizer (DAP) Application

Let us now look at the conditional quantile estimates of DAP on cotton yields.

Dependent Variable = Yield (kg/acre)	Coefficient (SE)					
	Q5	Q25	Q50	Q75	Q90	Q95
DAP (kg/acre)	22.99*** (4.04)	30.25*** (3.77)	50.26*** (5.32)	73.5*** (9.89)	76.97*** (10.31)	94.14*** (15.09)
Constant	48.14*** (5.47)	177.9*** (5.11)	306.6*** (7.21)	550*** (13.4)	836.36*** (13.96)	978*** (20.44)

Quantile Regression Estimates for Yield as a function of Fertilizer Application

Plotting these conditional quantile estimates (see Figure 4(S4)) we can see that apparently, neither household B or G is doing quite as well among their peers in the 5.26kg/acre and 2.38 kg/acre fertilizer usage brackets. In fact, they lie in the 25th and 90th percentile respectively as compared to their ranks in the unconditional yield distribution, i.e., 50th and 95th percentile respectively. Thus, once we condition on another variable, the ranking of households has shifted in this new distribution. Specifically, both the households considered in this example have a lower place in the conditional distribution as compared to the unconditional distribution. The question is, how does this change our interpretation of the QR coefficient estimates?



Plotting Quantile Regression (QR) Estimates of Yields Conditional on Fertilizer (DAP) Application

As evident with the two households taken as an example, since we cannot comment on where a household will be in the outcome distribution before and after a treatment (in our case additional DAP usage), we can only make statements about the distribution as a whole. Hence in the above example, while $\hat{\beta}_{0.5} = 50.27$ can be interpreted as the difference in the median yields due to an additional kilogram of DAP being used per acre, it **cannot** be interpreted as the effect of DAP for households having median yields at the population level because we don't know which households are still in the 50th quantile after using an additional kilogram of DAP per acre. The only way we can match the two interpretations is if the same households were retained at the median under both the conditional and unconditional distributions. This is exactly what rank invariance assumes. Under rank invariance, the QR coefficient can be interpreted as the coefficient for individual households located below quantile τ .

Statistically Restrictive Nature of Rank Invariance:

Let's consider the statistically restrictive nature of rank invariance through a thought experiment, adapted from Dong and Shen (2018). Consider a random sample of farmers and their observationally equivalent duplicates, i.e., clones. Suppose 'treatment' is a binary indicator for being a clone. Naturally, this 'treatment' should not have any effect on yields. However, random chance induced by idiosyncratic shocks (e.g., drought), can lead to a situation whereby the farmer and their clone may not have exactly equal yields in one instance of the experiment. Nevertheless, in infinite repetitions, the farmer and their clone would have the same distribution of yields. Rank similarity captures this notion and relaxes rank invariance by allowing for random deviations of each subject's rank away from her counterfactual self so long as the ranks distribution (conditional on factors influencing treatment status) is identical across treatment states.

Rank Similarity:

Rank similarity relaxes rank invariance by allowing for random deviations in a subject's rank away from their counterfactual self so long as the distribution of potential ranks among observationally equivalent households is the *same* across treatment states, i.e., $(U_0 \mid \vec{X} = x, u = u) \sim (U_1 \mid \vec{X} = x, u = u)$. Assuming that the distribution of unobservables is the same at the same rank of the potential outcome distribution given $\vec{X}$, the condition above provides the testable implication of rank similarity: $F_{U_1|X}(\tau) = F_{U_0|X}(\tau)$, which states for the matched set of households, the relative yields distribution conditional on yield determinants is identical across treatment states.

Statistical Test for Rank Similarity

Dong and Shen (2018) proposed a test for rank similarity where endogeneity in treatment was resolved using instrumental variable regressions. We adapt their framework to infer upon rank similarity when treatment endogeneity is resolved through PSM. First, we obtain the conditional yields distribution for the treated (i.e., $\hat{Y}_1$) and counterfactual households (i.e., $\hat{Y}_0$) through a linear regression of yields ($Y_{i_k t_k}$) on the covariate vector $\vec{X}'_{i_k t_k} = \{\vec{X}'^{inp}_{i_k t_k}, \vec{X}'^{we}_{i_k t_k}, \vec{X}'^{land}_{i_k}, \vec{X}'^{farmer}_{i_k}\}$ defined in model(3). The empirical cumulative distribution functions (CDFs) for $\hat{Y}_0$ and $\hat{Y}_1$ provide an estimate of the potential rank distributions, i.e.,

$\hat{U}_0$ and $\hat{U}_1$. We then compare $\hat{U}_0$ and $\hat{U}_1$ for the set of matched pairs of treated and counterfactual households using the two-sample Kolmogorov-Smirnov (K-S) distributional similarity test (Wilcox 1997; Dodge 2008).

The null hypothesis for the K-S test states that the distribution of ranks across treated and counterfactual groups are equal against the alternate that the distributions are not equal. The test statistic is $D_{mn} = \left[\frac{mn}{(m+n)}\right] \sup_y \left|\hat{U}_1(y) - \hat{U}_0(y)\right|$ where m and n are sample sizes for the treated and counterfactual groups respectively. If we reject rank similarity, then $\alpha_{1\tau}$ represents the change in the τ^{th} yield quantile by virtue of borrowing but may not be representative of the credit impact for households occupying the τ^{th} yield quantile (Borgen et al. 2022).

<u>*Results*</u>:

Table 1(S3) summarizes the results from the Kolmogorov Smirnov Distribution test. We reject the null hypothesis that the ranks for treated and control groups are similar. Figure 1 (S3) displays the empirical CDF of ranks for the treated and control groups respectively. Overall, the ranks for the treated group are significantly higher than the ranks in the control group. To see this, fix a level of rank (say 0.6) and note that the cumulative density values for the (informal) treated group is close to 0.45 whereas for the (informal) control group it is about 0.65. This implies that roughly 45% of households in the treated group have yield ranks lower than 0.6 whereas the same proportion of households for the control group is higher (i.e., about 65%). In this manner, the control group displays a higher concentration of low ranks as compared to the treated group and therefore rank similarity does not hold. Hence, while the quantile regression estimates represent the impact of credit on the τ^{th} conditional yield quantile $\forall \tau$, these may not represent the impact for the households located in the τ^{th} yield quantile.

Group	m	n	D-Statistic	p-value
Treatment: Informal Credit Access	897	931	0.18	0
Treatment: Formal Credit Access	1919	1968	0.21	0

Table 1 (S4): Results for the Rank Similarity Test using the Two Sample Kilmogorov-Smirnov Distribution Test

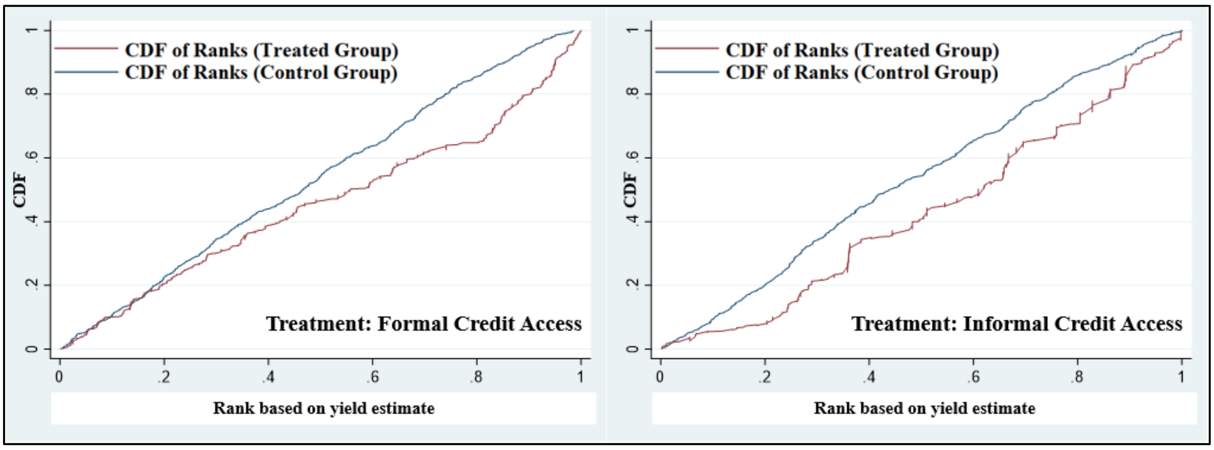


Figure 1 (S4): Empirical Rank CDFs for Formal and Informal Credit Access

Section 5: How are cluster-level fixed effects equivalent to within-cluster exact matching?

Suppose $Y_{it} = \alpha_1 D_{it} + \alpha_2 \vec{X}_{it} + u_{it}$ where $u_{it} = \lambda_t + \varepsilon_{it}$ such that λ_t are time-specific, individual-invariant components of the error in yields and ε_{it} are the strictly exogenous, normally distributed ($\varepsilon_{it} \sim N(0, \sigma^2 I_{NT})$) random errors such that $E(\varepsilon_{it} \mid X_{it}, D_{it}, m_i, s_t) = 0 \;\forall i \in \{1,2,...N\}$ & $\mathrm{t} \in \{1,2,...T\}$. λ_t and ε_{it} are orthogonal, i.e., independent of each other. Here we have assumed that the leftover random component of u_{it}, i.e., ε_{it} is homoscedastic. This assumption can be easily relaxed to incorporate village-level clustering in ε_{it} , which we do as part of our empirical strategy. Further, other than reducing the number of counterfactual households within a village-year or year, matching on the propensity of credit access (such that $|\hat{\Pr}(D_{it} = 1 \mid \vec{Z}_{it}) - \hat{\Pr}(D_{it} = 0 \mid \vec{Z}_{it})| \le k$ where calliper $k \in \{0.01, 0.05\}$) does not change the argument described here. Hence for ease of exposition, we do not consider the matching on propensity of credit access explicitly. Intuitively, u_{it} can be understood as the unobservable error left over *after matching* loanee and non-loanee households based on their propensity to access credit.

Under **within-cluster matching**, household i in year t can only be matched to another household j in t, depending on the closeness between the propensity scores of i and j. The impact of credit on yields for the matched sample is:
$E(Y_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - E(Y_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T})$ where $\{A_t\}_{t=1}^{T}$ is a set of T dummy variables such that $A_t = 1$ in year t and 0 otherwise.

$$
\begin{aligned}
& E(Y_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - E(Y_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) \\
&= (\alpha_1 + \alpha_2 \vec{X}_{it}) + E(u_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - \alpha_2 \vec{X}_{it} - E(u_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) \\
&= \alpha_1 + E(\lambda_t + \varepsilon_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - E(\lambda_t + \varepsilon_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) \\
&= \alpha_1 + E(\lambda_1 + \lambda_2 + ...\lambda_T + \varepsilon_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - E(\lambda_1 + \lambda_2 + ...\lambda_T + \varepsilon_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) \\
&= \alpha_1 + (\lambda_1 + \lambda_2 + ...\lambda_T) + E(\varepsilon_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - (\lambda_1 + \lambda_2 + ...\lambda_T) - E(\varepsilon_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) \\
&= \alpha_1 + E(\varepsilon_{it} \mid D_{it} = 1, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) - E(\varepsilon_{it} \mid D_{it} = 0, \vec{X}_{it}, \{A_t = 1\}_{t=1}^{T}) \\
&= \alpha_1
\end{aligned}
$$

Here the fourth equality follows from the fact that $\forall t$, λ_t is a constant given $A_t = 1$ and the fifth equality follows from the strict exogeneity assumption on ε_{it} . Under strict exogeneity, $E(\varepsilon_{it} \mid X_{it}, D_{it}) = 0$ in each t and therefore, $E(\varepsilon_{it} \mid X_{it}, D_{it}, A_t = 1) = 0 \ \forall t$.

Under a **cluster-level fixed effects regression**, we will attempt to estimate λ_t by including T time-specific dummy variables $\{A_t\}_{t=1}^{T}$ where $A_t = 1$ in year t and 0 otherwise.

So $Y_{it} = \alpha_1 D_{it} + \alpha_2 X_{it} + \sum_t \gamma_t A_t + \varepsilon_{it}$. Here, the impact of credit is given by:

$$
\begin{aligned}
& E(Y_{it} \mid D_{it} = 1, X_{it}, A_t) - E(Y_{it} \mid D_{it} = 0, X_{it}, A_t) \\
& = \alpha_1 + \alpha_2 X_{it} + \sum_t \gamma_t A_t + E[\varepsilon_{it} \mid D_{it} = 1, X_{it}, A_t] - \alpha_2 X_{it} - \sum_t \gamma_t A_t + E[\varepsilon_{it} \mid D_{it} = 0, X_{it}, A_t] \\
& = \alpha_1 + E[\varepsilon_{it} \mid D_{it} = 1, X_{it}, A_t] - E[\varepsilon_{it} \mid D_{it} = 0, X_{it}, A_t] \\
& = \alpha_1
\end{aligned}
$$

Here, the final equality follows from the strict exogeneity assumption

Clearly, the two approaches are equivalent and yield the impact of credit on crop yields as measured by α_1 .

Section 6: Numerical Example

Consider a sorghum-growing farmer operating in a risky environment with two states: $\Omega = \{\text{dry}(1), \text{humid}(2)\}$. The farmer applies irrigation (x_1) and fertilizer (x_2) to produce output. The farmer's production in each state is a Cobb-Douglas function of the input, i.e., $y_1 = 2.5x_1^{0.5}x_2^{0.3}$ and $y_2 = 4x_1^{0.3}x_2^{0.5}$ where the output elasticity of irrigation (fertilizer) is higher in the dry (humid) state respectively. Input and output prices are $p_1 = 10$, $p_2 = 4$ and $p_y = 11$ respectively. The risk-averse farmer's beliefs about the states are $(\psi_1, \psi_2) = (0.4, 0.6)$ and they maximize the expected utility from the state-contingent vector of net-returns (s_1, s_2) given by $W(s_1, s_2) = \psi_1 u(s_1) + \psi_2 u(s_2)$ where $u(.) = s_1^{1-\gamma} / (1-\gamma)$ and $\gamma = 2$. In the absence of a budget constraint the risk-averse farmer chooses $\vec{z}^* = (207.6, 366.5)$ and $\vec{x}^* = (10.4, 17)$ corresponding to a total expenditure of ~172 units. Holding $x_1 = x_1^* = 10.4$, state-wise production, total revenue and net revenues are plotted against x_2 in panels (a), (b) and (c) respectively of Figure 1(S5). Panel (d) (Figure 1(S5)) plots the net revenue curve depicting the tradeoff between state-contingent revenues and the net-revenue bisector. Since state 2 is 'good', the net-revenue curve lies *above* the bisector. Point *A* (black) depicts the unconstrained optimal while point *B* (red) denotes the optimal net revenues if the farmer were instead constrained to 140 units of expenditure ($\vec{z}' = (175.9, 312)$).

Figure 2 (S6) (Panel (a)) depicts the risk-neutral farmer's unconstrained equilibrium and both the unconstrained and constrained equilibrium for the risk-averse farmer. As expected the latter occurs at the point of tangency between the isocost curve and the farmer's indifference curve(s) and the the fair-odds line cuts the isocost frontier from below. Figure 2 (S6) (Panel (b)) presents a decomposition of the total shift in revenues due to a binding credit constraint into a contraction effect (along the revenue-mix ray to point *C* (blue)) and a pure-risk effect (along the fair-odds line and *away* from the bisector) from *C* to *B*. Figure 3(S6) presents the input shift underlying the revenue decomposition where the contraction effect dominates for each input.

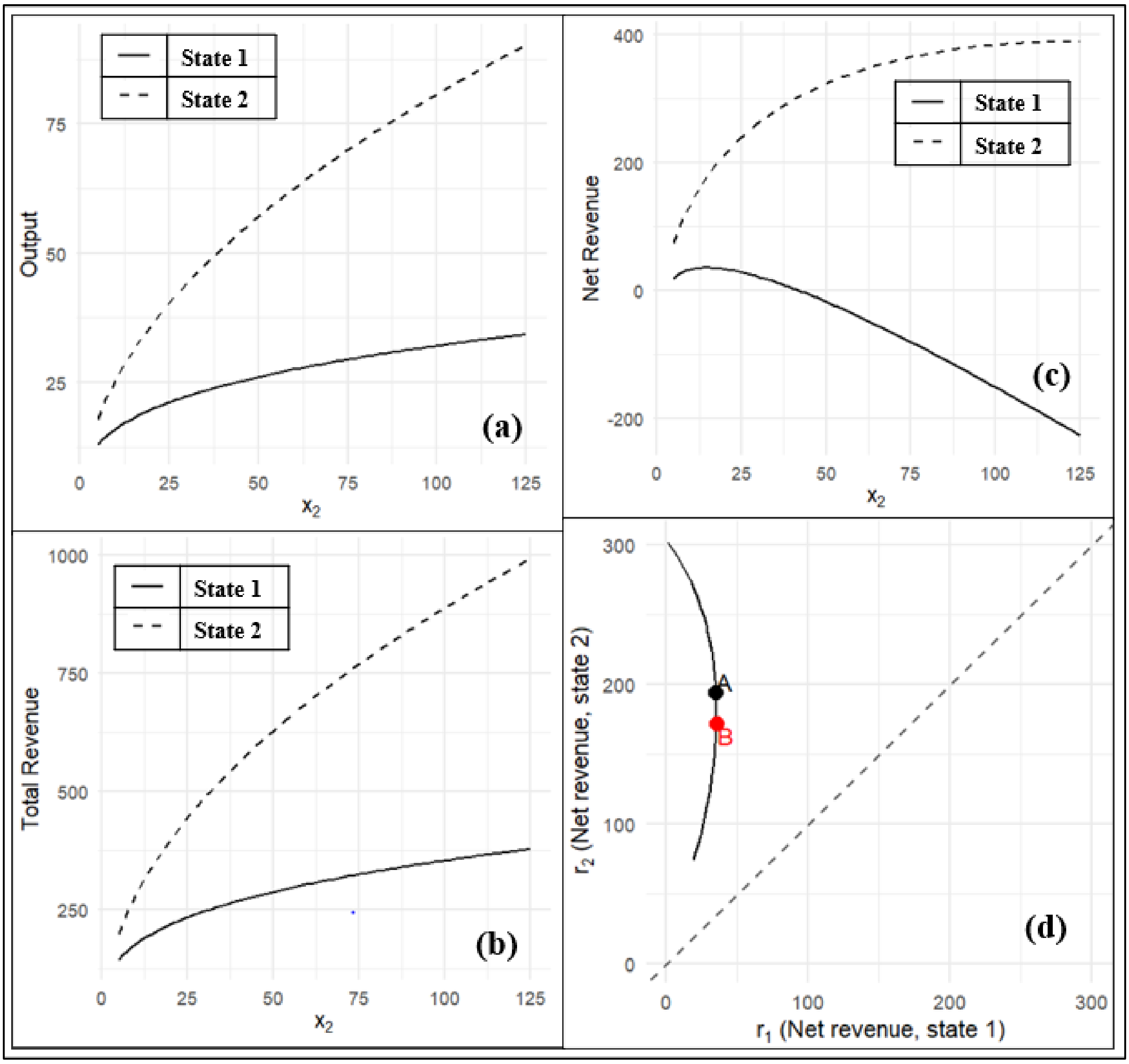


Figure 1(S6): Production, Revenue and Net Returns (Numerical Example, Credit Rationing Results)

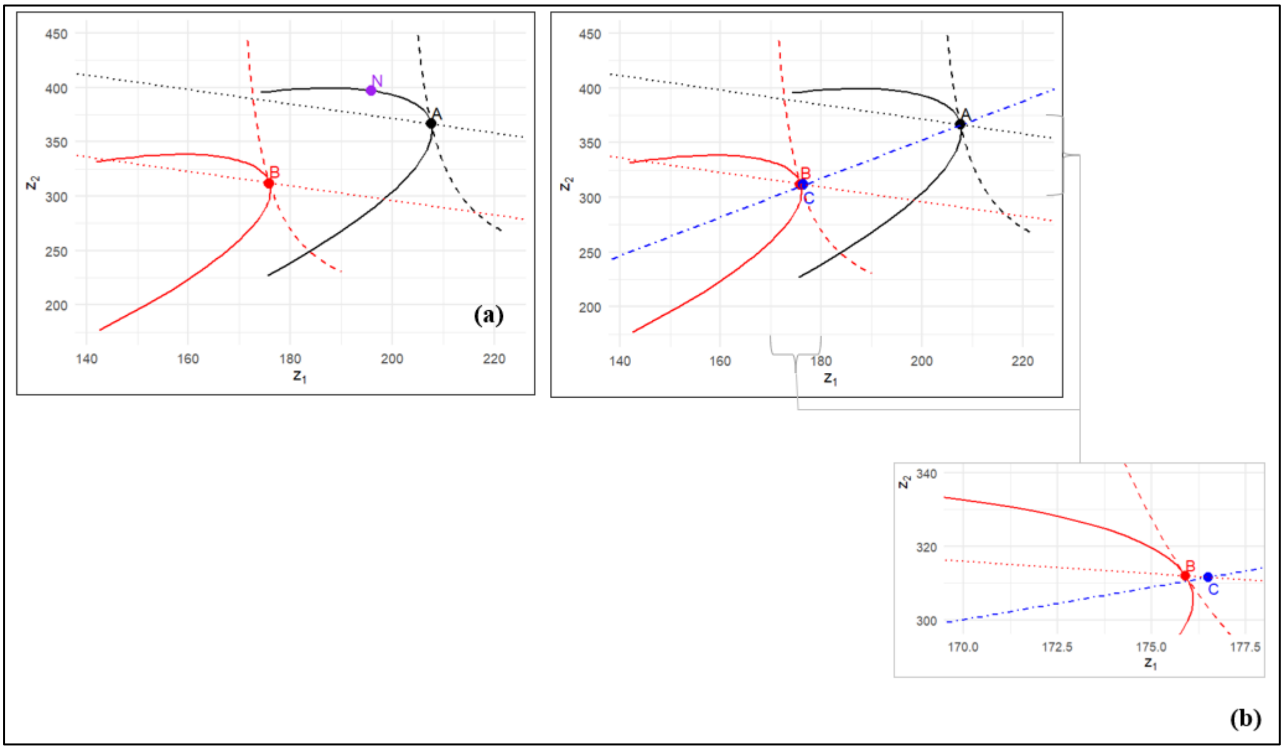
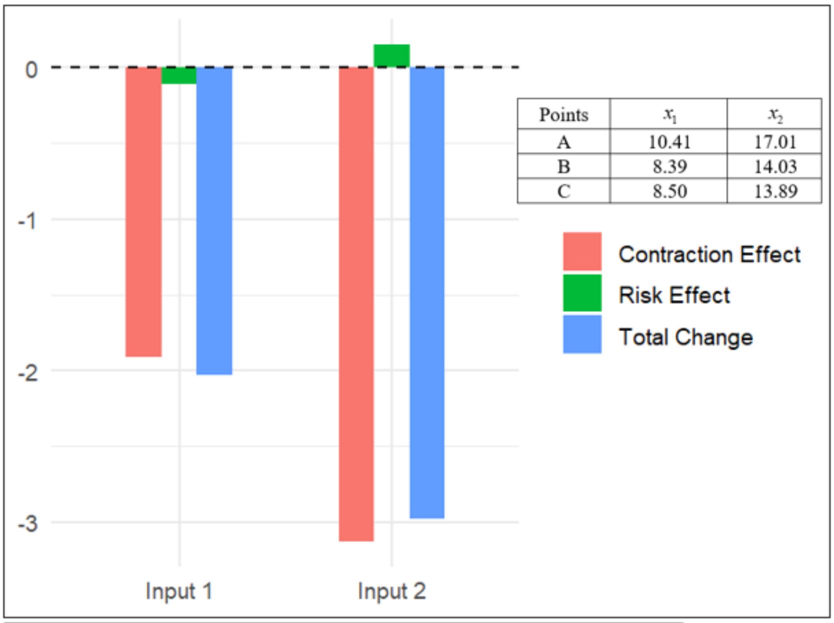


Figure 2 (S6): Constrained and Unconstrained Equilibrium (Numerical Example, Credit Rationing)

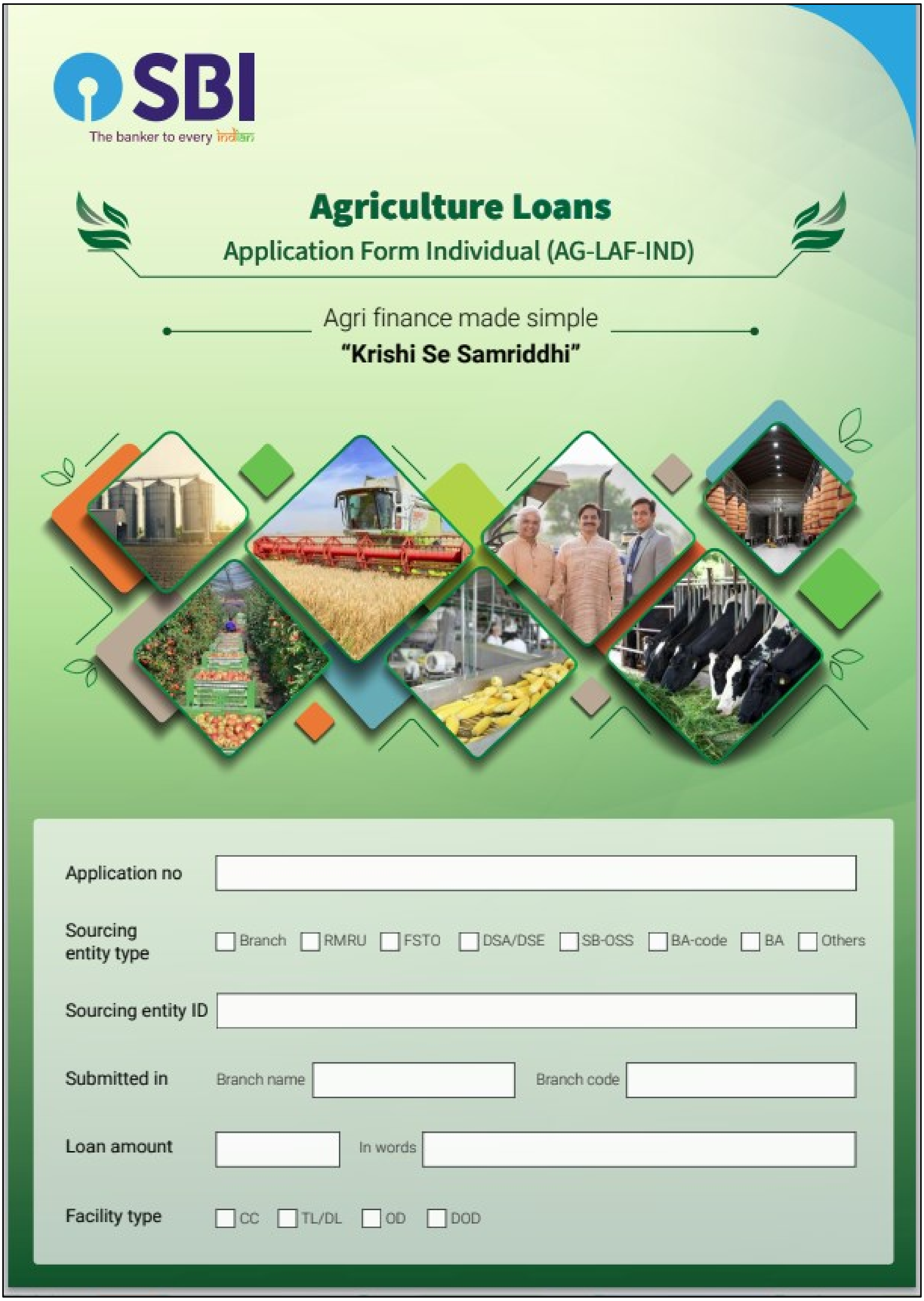


Points	x_1	x_2
A	10.41	17.01
B	8.39	14.03
C	8.50	13.89

Input Decomposition	x_1	x_2
$\Delta x_j^{\text{contraction}} = (x_j^C - x_j^A)$	-1.91	-3.13
$\Delta x_j^{\text{risk}} = (x_j^B - x_j^C)$	-0.11	0.15
Total Change ($\Delta x_j = (x_j^B - x_j^A)$)	-2.03	-2.98

Figure 3 (S6): Input Decomposition (Numerical Example, Credit Rationing)

Section 7A: Loan Application Form Details (Source: State Bank of India, Accessed here)



SBI

The banker to every Indian

Agriculture Loans

Application Form Individual (AG-LAF-IND)

Agri finance made simple

"Krishi Se Samriddhi"

Application no	
Sourcing entity type	☐ Branch ☐ RMRU ☐ FSTO ☐ DSA/DSE ☐ SB-OSS ☐ BA-code ☐ BA ☐ Others
Sourcing entity ID	
Submitted in	Branch name ____ Branch code ____
Loan amount	____ In words ____
Facility type	☐ CC ☐ TL/DL ☐ OD ☐ DOD

Individual- AG-LAF-IND

Agricultural Loans

Application form

Application form ☐☐☐☐☐☐☐☐☐☐☐☐

Sourcing entity type: ☐ Branch ☐ RMRU ☐ FSTO ☐ DSA/DSE

☐ SBOSS ☐ BA ☐ Others

Sourcing entity ID: ______________________

Submitted in: Branch name: ______________________

Branch code: ☐☐☐☐☐☐☐☐

Loan amount (in Rs.): ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

(in words) Rupees ______________________

Facility type: ☐ CC ☐ TL/DL ☐ OD ☐ DOD

SBI

Individual- AG-LAF-IND

BORROWER'S PROFILE

Applicant(s) details:
(Co-applicant details to be captured in separate form)

PHOTO

Existing customer: ☐ Yes ☐ No

If yes, CIF no.

Name: First Name / Middle Name / Last Name

Son/Daughter/ Wife of:

Date of Birth: **Age :** **Years**

Educational Qualifications: ______

Signature

Full Name(s) of the Co-applicants	Relation with Applicant	Age (Years)	Educational Qualifications
Co-applicant - 1			
Co-applicant - 2			

Family details of applicant:

Names of the family members	Relationship	Whether dependent (Yes/No)	Annual income (Rs Lakhs)

Identification of the applicant:

Gender: ☐ Male ☐ Female ☐ Transgender

Marital Status: ☐ Single ☐ Married ☐ Others

Citizenship: ☐ Indian ☐ Others

SC ☐ **ST** ☐ **OBC** ☐ **General Category** ☐ **Minority, Others, please specify** ______

Whether Agricultural laborer: ☐ Yes ☐ No

Mobile No.: **Alternate Contact no.**

PAN No.: (Form 60, if PAN not available)

Aadhar No:

Voter ID No.:

MNREGA ID:

Farmer ID:

Land ID:

Current Address:

Address 1:

Address 2:

Village: **City:**

State: **District:**

Sub-District: **Pin Code:**

SBI **Individual- AG-LAF-IND**

Permanent Address:

Address 1:

Address 2:

Village: City:

State: District:

Sub-District: Pin Code:

ASSETS & LIABILITIES

Income details

Sl no.	Source of income (Agriculture/Allied activities/ Business income/Other income)	Brief detail of activity	Amount (Rs Lakhs)

Deposits:

Account type	IFSC code	Bank name	Branch name	Account number

Particulars of assets:

Sl no.	Type of Asset (Movable/Immovable/Liquid)	Brief detail (Address for immovable property)	Quantity of asset	Amount (Rs. Lakhs)

Particulars of Statutory dues outstanding (if any):

Name of the Statutory Body	Dues outstanding (in Rs.)	Year of Pendency	Remarks

Existing loans, if any:

Type of Institution	Bank & Branch name	Facility type	Purpose of loan	Limit sanctioned (Rs. Lakhs)	Present Outstanding (Rs. Lakhs)	Account No.	Instalment repayable (Rs. Lakhs)

What is the arrangement for sales of produce?

Do you have tie-up? ☐ Yes ☐ No Tie-up with ______________________________

Tie-up Code:

Direct sale: ☐ Yes ☐ No If yes, distance from open market ____________ km

Particulars of assets:

Sl no.	Godown/ warehouse	Produce	Quantity (Quintals)	Present Market value/ MSP (Rs. Lakhs)	Total value (Rs. Lakhs)	Margin (Rs. Lakhs)	Advance value (Rs. Lakhs)

SBI

Individual- AG-LAF-IND

LOAN DETAILS

Loan amount (in Rs.):

(in words) Rupees

Purpose of loan:

Sl no.	Facility type	Product / Purpose	Amount	Repayment frequency	For TL - Moratorium	Advance value
	(CC/TL/OD/DOD/Others)		(Rs. Lakhs)	(Rs. Lakhs)	(months)	(months)
Total loan (Rs Lakhs):						

Is this a Takeover Loan ☐ Yes ☐ No (if Yes, Please specify the details below)

Particulars of assets:

Type of Institution	Bank & Branch name	Facility type	Purpose of loan	Limit sanctioned	Present Outstanding	Account No.	Instalment repayable
				(Rs. Lakhs)	(Rs. Lakhs)		(Rs. Lakhs)

Mode of operations: ☐ Single ☐ Jointly

Experience in no. of years of Agri business activity for which loan is requested ☐☐ years.

Consent from borrower beyond mandatory crop insurance: Asset/ Health/PAIS: ☐ Yes ☐ No

Affidavit obtained: (applicable for financing to Tenant farmers, oral lessees and sharecroppers) ☐ Yes ☐ No

SECURITY & GUARANTOR DETAILS

Security (Existing) :

Security	Type of security	Security value	Security description	Owned/ Guarantor's
		(Rs Lakhs)		
Primary				
Collateral 1				
Collateral 2				
Collateral 3				
Collateral 4				

Guarantor details, if any (Existing): (Details to be captured in separate form)

Name	Address

Security (Proposed)

Security	Type of security	Security value	Security description	Owned/ Guarantor's
		(Rs Lakhs)		
Primary				
Collateral 1				
Collateral 2				
Collateral 3				
Collateral 4				

Guarantor details, if any (Existing): (Details to be captured in separate form)

Name	Address

SBI

Individual- AG-LAF-IND

Part of government scheme

I am availing subsidy from Govt sponsored scheme ☐

If yes, Govt. scheme name:

Application ID:

Subsidy type: ☐ Interest subsidy ☐ Capital subsidy ☐ Subvention ☐ Others

Particulars of assets:

Village	Survey / Block No.	Khata No.	Ownership type		Area	Irrigated area	Whether land mortgaged with a bank
			Owned / Leased / Sharecropper	Years of experience (years)	(acres)	(acres)	

Particulars of assets:

Survey / Block No.	Khata No.	Area	Crop grown	Season	Cropping pattern	Source of irrigation
		(acres)		(Rabi/ Kharif/ Zaid)	(Short term/Long term)	

DECLARATIONS

I. I/We authorize the bank to issue/link a debit card with my KCC account in case I am availing a KCC loan.

II. 'I/We further agree that any facility that may be provided to me/us shall be governed by the rules of the Bank that may be in force from time to time. I/We will be bound by the terms and conditions of the credit facility/ies that may be granted to me/us. I/We authorize the Bank to debit my loan account with the Bank for any fees, charges, interest etc. as may be applicable.

III. I/We certify that the information provided by me/us in this application form is true and correct in all respects and State Bank of India is entitled but not bound to verify this directly or through any third-party agent. I/We confirm that the attached copies of financials/Bank Statements/Title/Legal documents etc. are submitted by me/us against my/our loan application and certify that these are true copies. I/We further acknowledge the Bank's right to seek any information from any other source in this regard. I/We understand that all of the above-mentioned information shall form the basis of any facility that the Bank may decide to grant to me/us at its sole discretion.

IV. I/We give my/our consent to the bank to verify my/our Credit Information Reports.

V. I/We understand that the Bank will use the information furnished by me/us in accordance with the applicable laws of India and any other foreign laws to which I/we/the Bank may be subject to. The said information will be used solely for the purpose of processing, sanctioning, updating, maintaining, and operating my/our account/s or facility/ies availed by me/us and processing transactions initiated by me/us in such account/ pursuant to such facility.

VI. It will be in order for the Bank to disqualify me/us from receiving any credit facilities from the Bank in case it is proved that the declaration of my/our outside borrowings made above contain incorrect facts.

VII. I/We confirm that no insolvency proceedings or suits for recovery of outstanding dues or monies whatsoever for attachment of my/our assets or properties and/or any criminal proceedings have been initiated and/or are pending against me/us and that I/we have never been adjudicated insolvent by any court or other authority.

VIII. No action nor other steps have been taken or legal proceedings started by or against me/us in any court of law/other authorities for winding up, dissolution, administration or re-organization or for the appointment of a receiver, administrator, administrative receiver, trustee or similar officer or for my/our assets.

IX. I/We understand and acknowledge that State Bank of India shall have absolute discretion, without assigning any reasons (unless required by applicable law), to reject my/our application and that State Bank of India shall not be responsible/liable in any manner whatsoever to me/us of such rejection and any costs, losses, damages or expenses, or other consequences, caused by reasons of such rejection, or any delay in notifying me/us of such rejection, or our application.

X. I/We certify that none of the immovable properties offered for Equitable Mortgage have been mortgaged/proposed to be mortgaged to any other Banks/Institution. I/We also certify that owner of the property have clear, unambiguous, absolute and mortgageable title over these properties.

Individual- AG-LAF-IND

XI. I/We acknowledge that the existence of this account and details thereof (including details of transactions and any defaults committed by me), will be recorded with credit reference and such information (including processed information) may be shared with banks/financial institutions, credit bureau, statutory/regulatory authorities and other credit grantors for the purposes of assessing further applications for credit by me/us and/or members of my/our household, and for occasional debt tracing and fraud prevention.

XII. I/We agree to receive SMS alerts/Phone calls related to my/our application status and account activity as well as product use message/calls that the Bank will send/make, from time to time, on the mobile/ phone number/ email as mentioned in this application form. I/We undertake to intimate the Bank in the event of any change in the mobile/ phone number/ email and residential address furnished by me/us.

XIII. I/We also acknowledge and agree that the consent hereby provided by me/us shall be legally binding on me/us irrespective of my/our registration with DNDINCPR registries and shall override such registrations.

XIV. I/We hereby declare that I/We have no borrowings/liabilities excepting those mentioned here as on the date of application.

XV. I/We hereby undertake to abide by the terms and conditions that the Bank may stipulate in sanction of this loan and inform Bank in the event of acquiring any other assets during the tenure of the advance.

____________________	____________________	____________________
Signature applicant	Signature co-applicant 1	Signature co-applicant 2
Place:	Place:	Place:
Date:	Date:	Date:

Acknowledgement of receival of form

Received from Mr./Mrs/M/s ______________________________ Loan application received on ______________________, for loan of Rs. ____________________ Applicant would be advised of the Bank's decision on the application as per banks guidelines.
Application received by ____

Date: ____________
Place: ____________

Authorized Signatory

Section 7B: Copy of Sample Land Record for Maharashtra State (Accessed here):

3/27/22, 3:19 PM https://mpbhulekh.gov.in:8092/UniSearch/getKhasraCopyView?dist_id=39&teh_id=01&lgdcode=502209&khasraId=1039010...

Madhya Pradesh Computrised Land Records

Khasra

Format 1 (See rule 6)

The Madhya Pradesh Bhu-Rajsva Sanhita (Bhu-Servekshan Tatha Bhu-Abhilekh Niyam 2020)

Village:करौंदाटोला			Halka:करौंदाटोला			Tehsil: पुष्पराजगढ़			District: अनूपपुर		Year: 2021-2022
Unique ID of part of parcel of Land	Type of land parcel (Survey Number/ Block Number)	Plot Number (In caseof a Block)	1. Area (in ha./sq. metre) 2. Land use for which assessed 3. Land Revenue /Lease Rent (in Rs.)	1.Name of Bhumiswami, his mother's/ Father's/ Husband's name and address of residence 2. Govt. land	Share of each Bhumiswami (in cae of joint holding)	1. Name of Government lessee, his mother's/ father's/ husband's name and address of residence 2. Period of lease 3. Area under lease	Name of occupancy tenant (if any), his mother's/ fathers' husband's name and address of residence	Encumbrances and charges on land 1. Mortgage 2. Hypothecation 3. Ongoing land acquisition process	Crop details Crop 1. Kharif 2. Rabi 3. Jayad 4. Other	Area under crop	1. Irrigation status of land 2.Structures/trees on land 3. Other remarks 4. Orders for correction of entries during year in column no (1) to 9
1	2	3	4	5	6	7	8	9	10	11	12
1525941123	343/1/2 (S)		3.7220 हेक्टेयर रु.0.00	(शासकिय) वन विभाग मध्यप्रदेश शासन शासकीय संस्था	1						न्यायालय कलेक्टर के प्रकरण क्र. 0012/ अ-20(3)/2021-22, आदेश दि. 10/01/2022 के अनुसार भू-अभिलेख अद्यतित

Note :-

1. This form is only for the information of the applicant |
2. It cannot be used as evidence in any court of law |
3. For digitally signed copy please apply through IT Center, LSK, MPOnline or online. |
4. For correction/amendment in entries, contact the concerned district/tehsil office |

Madhya Pradesh Computrised Land Records

Khasra

Format 1 (See rule 6)

The Madhya Pradesh Bhu-Rajsva Sanhita (Bhu-Servekshan Tatha Bhu-Abhilekh Niyam 2020)

Village:करौंदाटोला			Halka:करौंदाटोला			Tehsil: पुष्पराजगढ़			District:अनूपपुर		Year: 2021-2022
Unique ID of part of parcel of Land	Type of land parcel (Survey Number/ Block Number)	Plot Number (In caseof a Block)	1. Area (in ha./sq. metre) 2. Land use for which assessed 3. Land Revenue /Lease Rent (in Rs.)	1.Name of Bhumiswami, his mother's/ Father's/ Husband's name and address of residence 2. Govt. land	Share of each Bhumiswami (in cae of joint holding)	1. Name of Government lessee, his mother's father's husband's name and address of residence 2. Period of lease 3. Area under lease	Name of occupancy tenant (if any), his mother's/ fathers' husband's name and address of residence	Encumbrances and charges on land 1. Mortgage 2. Hypothecation 3. Ongoing land acquisition process	Crop details Crop 1. Kharif 2. Rabi 3. Jayad 4. Other	 Area under crop	1. Irrigation status of land 2.Structures/trees on land 3. Other remarks 4. Orders for correction of entries during year in column no (1) to 9
1	2	3	4	5	6	7	8	9	10	11	12
1326241427	423/1 (S)		10.2830 हेक्टेयर चरागाह 10.283छोटे झाड़ का जंगल 10.2830 रु.0.00	(शासकीय) वन विभाग मध्यप्रदेश शासन शासकीय संस्था	1						चरागाह 10.2830 छोटे झाड़ का जंगल 10.2830 न्यायालय कलेक्टर के प्रकरण क्र. 0012/अ-20(3)/2021-22, आदेश दि. 10/01/2022 के अनुसार भू-अभिलेख अद्यतित

Note :-

1. This form is only for the information of the applicant |
2. It cannot be used as evidence in any court of law |
3. For digitally signed copy please apply through IT Center, LSK, MPOnline or online. |
4. For correction/amendment in entries, contact the concerned district/tehsil office |

Madhya Pradesh Computrised Land Records

Khasra

Format 1 (See rule 6)

The Madhya Pradesh Bhu-Rajsva Sanhita (Bhu-Servekshan Tatha Bhu-Abhilekh Niyam 2020)

Village:करौंदाटोला			Halka:करौंदाटोला			Tehsil: पुष्पराजगढ़			District: अनुपपुर		Year: 2021-2022
Unique ID of part of parcel of Land	Type of land parcel (Survey Number/ Block Number)	Plot Number (In caseof a Block)	1. Area (in ha./sq. metre) 2. Land use for which assessed 3. Land Revenue /Lease Rent (in Rs.)	1.Name of Bhumiswami, his mother's/ Father's/ Husband's name and address of residence 2. Govt. land	Share of each Bhumiswami (in cae of joint holding)	1. Name of Government lessee, his mother's father's husband's name and address of residence 2. Period of lease 3. Area under lease	Name of occupancy tenant (if any), his mother's/ fathers' husband's name and address of residence	Encumbrances and charges on land 1. Mortgage 2. Hypothecation 3. Ongoing land acquisition process	Crop details Crop 1. Kharif 2. Rabi 3. Jayad 4. Other	Crop details Area under crop	1. Irrigation status of land 2.Structures/trees on land 3. Other remarks 4. Orders for correction of entries during year in column no (1) to 9
1	2	3	4	5	6	7	8	9	10	11	12
1290479049	342/1 (S)		6.8450 हेक्टेयर अन्य 6.0450 रु.0.00	(शासकीय) वन विभाग म०प्रदेश शासन शासकीय संस्था	1						अन्य 6.0450 न्यायालय कलेक्टर के प्रकरण क्र. 0012/ अ-20(3)/2021-22, आदेश दि. 10/01/2022 के अनुसार भू-अभिलेख अद्यतित

Note :-

1. This form is only for the information of the applicant |
2.It cannot be used as evidence in any court of law |
3. For digitally signed copy please apply through IT Center, LSK, MPOnline or online. |
4. For correction/amendment in entries, contact the concerned district/tehsil office |

Section 8: Steps for computing τ_i^*

Step 1: Compute the marginal benefit attained by borrowing household i from an additional unit of credit at each quantile $\tau \in \{0.02, 0.04..., 0.98\}$ as $\hat{\hat{\alpha}}_{1i\tau} = \frac{\hat{\alpha}_{1\tau}(\text{area}_i)p_{iy}}{K_i}$, i.e., assuming that each unit of credit is equally productive.

Step 2: For each i at each $\tau \in \{0.02, 0.04..., 0.98\}$, retain $\hat{\hat{\alpha}}_{1i\tau}$ when you reject[37] the null hypothesis: $H_{0,i,\tau} : \tilde{\alpha}_{1i\tau} = 0$ against the alternative that $H_{1,i,\tau} : \tilde{\alpha}_{1i\tau} > 0$.

Step 3: Starting from the smallest τ retained in step 2, take a weighted sum of the marginal benefit from credit at each quantile as $\sum_{\tau} \Pr(\tau - 1 \le Y_i \le \tau) \frac{\hat{\alpha}_{1\tau}(\text{area}_i)p_{iy}}{K_i}$. Stop summing when $\sum_{\tau} \Pr(\tau - 1 \le Y_i \le \tau) \frac{\hat{\alpha}_{1\tau}(\text{area}_i)p_{iy}}{K_i} = (1 + s_i)$ and store the quantile level at the stopping point as τ_i^*.

Note that to account for returns from credit utilized towards multiple cropping, we conduct steps 1-2 for for all crops grown by the household in the same season and year when the household grows sorghum (and borrows), and sum the marginal benefit for all crops (including sorghum) in step 3 to compare with the marginal cost of credit. Whenever $\tau_i^* < 1$, *i*'s shadow value of capital is greater than zero meaning that the farmer will be stopped at lower input application than where optimal returns are achieved in response to *i*'s subjective belief about state probabilities. Importantly the computation of τ_i^* assumes at each quantile, that if the household were in a given conditional quantile without credit, they would retain

---

[37] The 95% confidence interval associated with $\hat{\hat{\alpha}}_{1j\tau}$ is: $\hat{\hat{\alpha}}_{1j\tau} \pm \left(\frac{\text{area}_j \times p_{jy}}{K_j} \right) \times SE(\hat{\alpha}_{1\tau}) \times 1.645$

their ‘rank’, i.e., lie in the same conditional yield quantile with credit. In other words, access to credit shifts the level of yields but not the relative ordering of households within the conditional yield distribution. Moreover each state (represented by a conditional quantile) and its associated subjective probability will contribute towards the ex-ante assessment of marginal benefit from an additional unit of credit.

Section 9: Robustness

Variable Acronym	Average Difference (Control-Treated): Pre-Matching	Average Difference (Treated-Control): Post-Matching
Economic Status ($\vec{Z}^{eco}$)		
$size_{it}$	-2.13***	-0.63
own_{it}	-0.03*	0.02
nf_{it}	-1322**	-1307*
$cons_{it}$	-7847***	-3216***
liv_{it}	-2266***	-995**
Household-Specific Time Invariant Controls ($\vec{Z}^{fe}$)		
age_{i}	0.98	0.44
edu_{i}	-1.59***	-0.83*
$prior_{i}$	-1.26***	-0.93**
$soil_{i}$	5.02	4.87
$caste_{i}$	0.03*	0.02
Institutional Production Environment ($\vec{Z}^{ins}$)		
$kharif_{it}$	-0.21***	-0.14***
$loans_{t}$	-1773	-1149
rs_{dt}	14.21***	8.38*

Table 1 (S9): Assessing Balance Post-Matching (Formal)

Variable Acronym	Average Difference (Treated-Control): Pre-Matching	Average Difference (Treated-Control): Post-Matching
Economic Status ($\vec{Z}^{eco}$)		
$size_{it}$	-1.95***	-1.3*
own_{it}	-0.4*	0.03
nf_{it}	-700	-34
$cons_{it}$	-5950***	-3708***
liv_{it}	-1788***	-770
Household-Specific Time Invariant Controls ($\vec{Z}^{fe}$)		
age_i	1.15	0.43
edu_i	-0.26	0.14
$prior_i$	-1.79***	-0.63**
$soil_i$	3.79	3.33
$caste_i$	0.003	0.02
Institutional Production Environment ($\vec{Z}^{ins}$)		
$kharif_{it}$	-0.30***	-0.18***
$loans_t$	-9957***	-6069**
rs_{dt}	9.87***	4.04

Table 2 (S9): Assessing Balance Post-Matching (Informal)

Table 3 (S9): Marginal Credit Impact (Formal): Balance Adjusted and Base Estimates (0.05 calliper)

Quantile	**Formal Credit**	
	Base	**Balance Adjusted**
0.05	-11.43 (13.01)	-7.85 (12.32)
0.1	4.44 (8.05)	-4.79 (7.97)
0.15	10.68 (20.94)	10.05 (21.51)
0.2	15.2 (16.25)	18.02 (16.34)
0.25	25.82* (14.93)	23.16 (15.18)
0.3	33.99** (16.59)	35.92* (19.1)
0.35	37.51** (16.92)	43.85** (17.73)
0.4	39.61** (16.66)	39.99** (18.25)
0.45	37.7* (22.8)	44.41** (19.47)
0.5	39.64 (33.31)	39.01 (26.48)
0.55	48.73 (40.18)	50.07* (29.55)
0.6	72.13** (29.48)	74.9** (31.56)
0.65	71.6** (28.96)	73.74** (37)
0.7	89.86*** (34.72)	88.03** (37.52)
0.75	90.83*** (28.18)	83.71** (40.77)
0.8	86.42*** (28.18)	90.85*** (20.4)
0.85	97.59* (53.4)	111.46 (79.35)
0.9	126.63** (49.13)	147.02** (67.09)
0.95	100.53*** (29.33)	146.24*** (49.61)

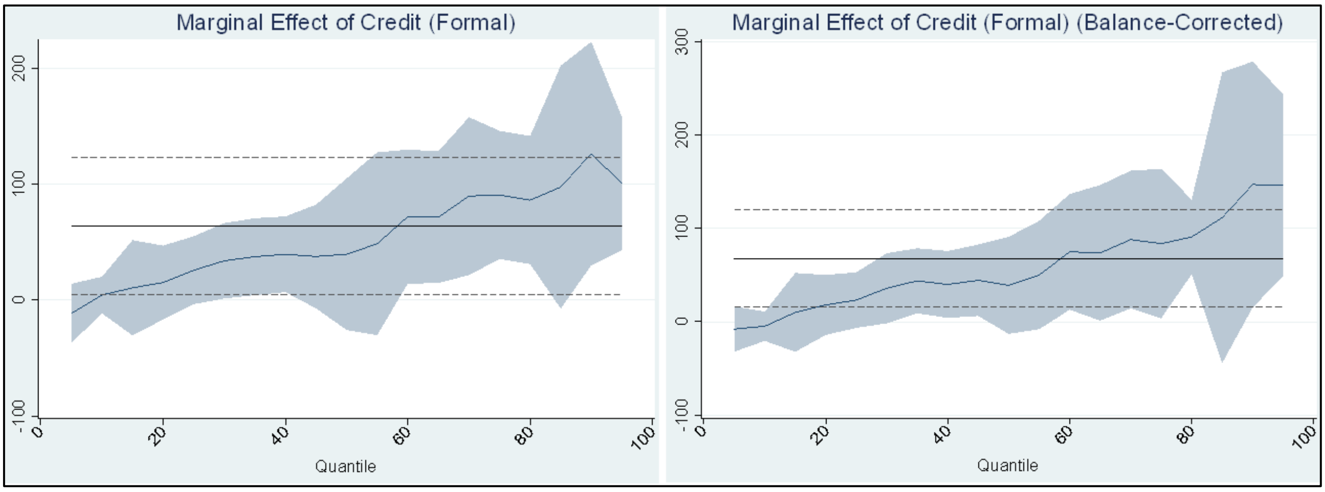


Figure 1(S9): Marginal Effect of Formal Credit (With and Without Balance Correction)

Notes: The grey bands indicate the 95% confidence interval.

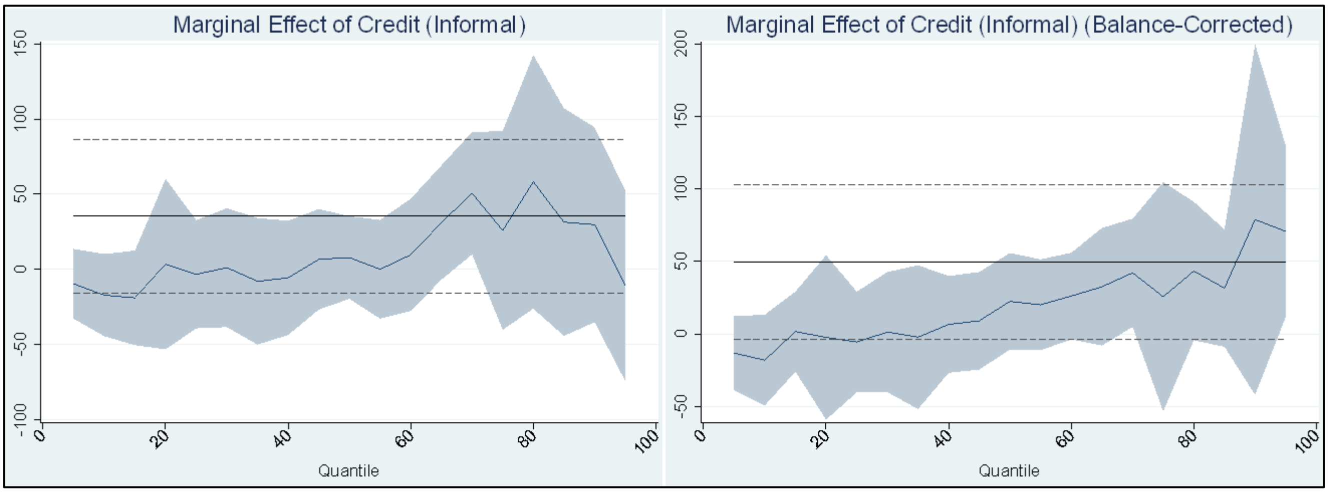


Figure 2(S9): Marginal Effect of Informal Credit (With and Without Balance Correction)

Notes: The grey bands indicate the 95% confidence interval.

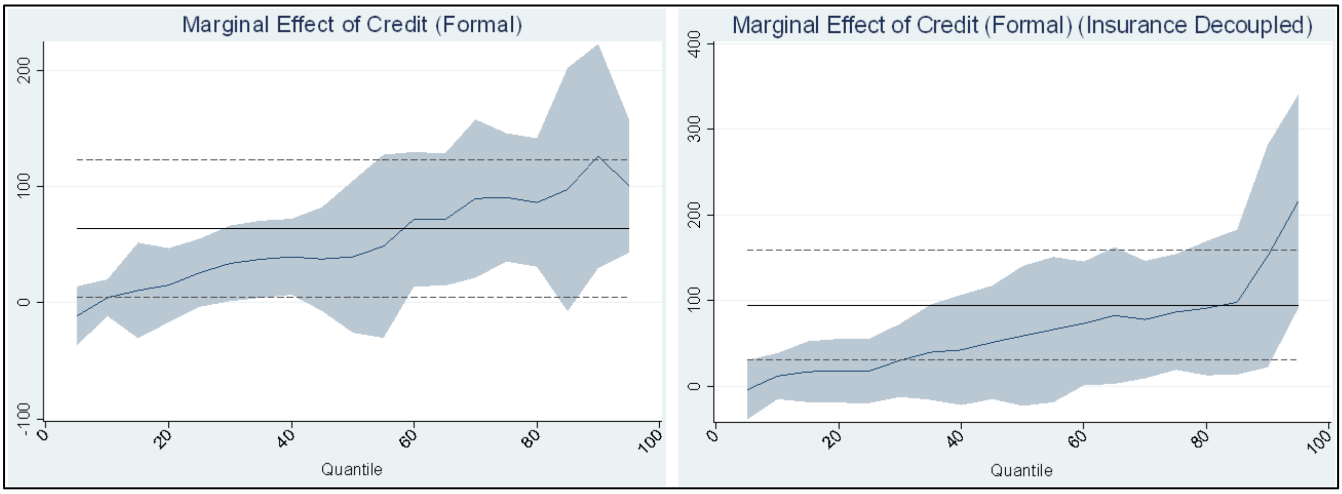


Figure 3(S9): Marginal Effect of Formal Credit (With and Without Insurance Decoupling)

Notes: The grey bands indicate the 95% confidence interval.

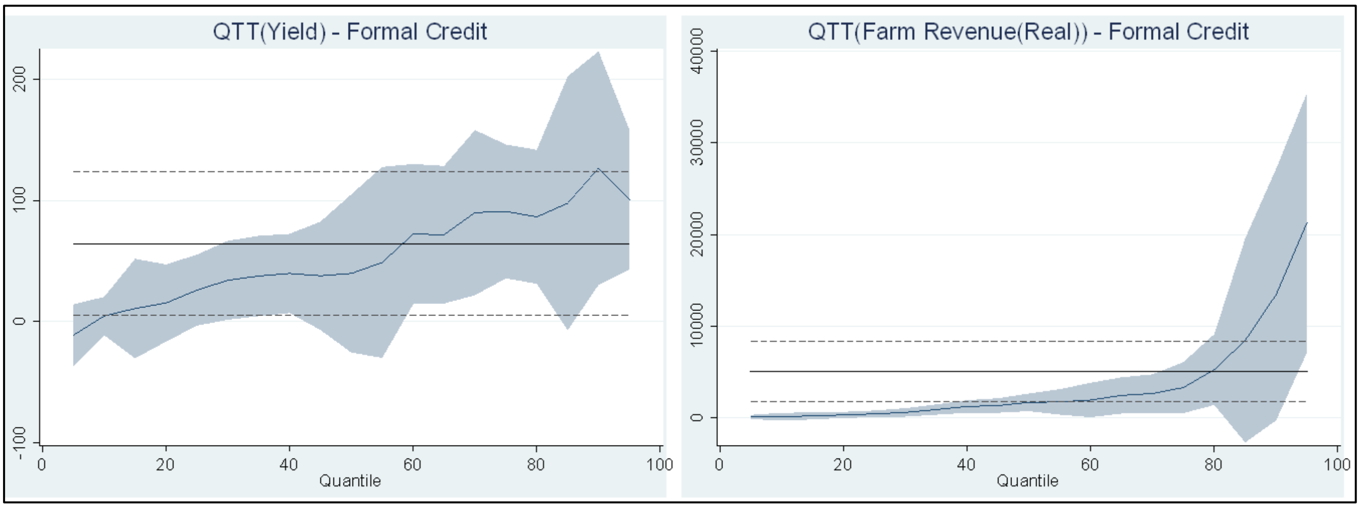


Figure 4(S9): Marginal Effect of Formal Credit on Sorghum Yield and Farm Revenue (in rupees, real)

Notes: The grey bands indicate the 95% confidence interval.

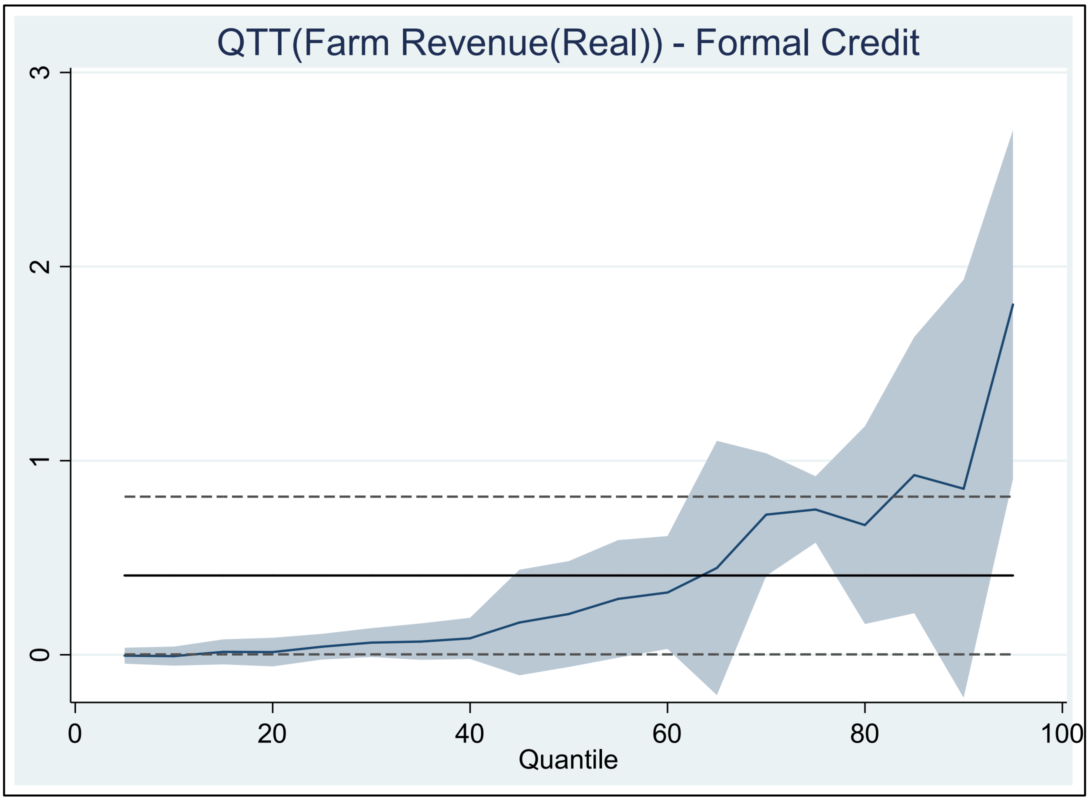


Figure 5(S9): Marginal Effect of Formal Credit (Continuous Treatment, in rupees) on Farm Revenue (in rupees, real)

Notes: The grey bands indicate the 95% confidence interval.

Section 10: Revisions:

(I)

In this section, I outline how I assess credit impacts for Sorghum-based cropping systems. A key motivation for incorporating cropping-system heterogeneity into the matching procedure is that credit access and its productivity effects are unlikely to be uniform across farming systems. Households operating within different cropping systems face distinct input requirements, revenue streams, and exposure to production risks. A matched control household drawn from a different cropping system thus could be a poor counterfactual even if it resembles the treated household on observable demographic, wealth and land characteristics since the baseline production environment and the margin on which credit operates would differ systematically.

In order to classify the cropping patterns in the VDSA dataset, we first begin by categorizing the VDSA crops based on Ministry of Agriculture and Farmers' Welfare classification (see Table 1). We exclude a certain set of crops due to extremely low frequency of instances (and area) and unclear classification. Importantly, these exclusions do not lead to any loss of households and less than 10 percent of overall cultivated area across the years.

Crop Categories	**VDSA Crop Names**
Cereals	Paddy, Wheat, Maize, Barley
Commercial (including Oilseeds)	Cotton, Soybean, Groundnut, Castor, Sunflower, Mustard, Sesamum, Jute, Safflower, Linseed, Mulberry, Niger Seed, Turmeric, Agashi, Rapeseed, Guvar Gum, Black Mustard
Fruits, Vegetables and Spices	Onion, Potato, Chillies, Cluster Bean, Cucumber, Coriander, Ladys Finger, Arecanut, Grapes, Brinjal, Green Pea, Tomato, Lemon, Cauliflower, Coconut, Drumstick, Mango, Pomegranate, Ber, Banana, Citrus, Ridge Gourd, Vegetable, Garlic, Leafy Vegetable, Colocasia, Peas, Watermelon, Papaya, Bitter Gourd, Cumin, Kagzi Lime, Beans, Beans(Bokala_Fawa_Runner Beans), French Bean, Radish, Gourd, Guava, Orange, Sweet Potato, Pumpkin, Carrot, Sapota, Fenugreek, Cumin, Tinda, Bottle Gourd, Cabbage, Green Chillies, Amla, Custard Apple, Ginger, Smooth Gourd, Amaranthus, Red Green Leaf, Tamarind, Capsicum, Green Leaves, Lettuce, Lotus Seed, Red Leaf, Shevaga, Soft Gourd, Spinach.
Millets	Sorghum, Finger Millet, Pearl Millet, Oats, Minor Millet, Small Millet (Marua).
Pulses	Pigeonpea, Chickpea, Greengram, Blackgram, Cowpea, Lentil, Horsegram, Lathyrus, D Lab Lab, Matki, Minor Pulse, Tewada, Phaseolus Trilobus.
Exclusions (Annual/ Perennial Crops, Plantation Crops, Fodder and Flowers)	Sugarcane, Tobacco, Teak, Trees, Variga, Dhaincha, Others, Sajgura, Avestor, Jungle Tree, Acarus Calmus, Rabda, Current Fallow, Fodder, Green Grass, Chrysanthemum, Jasmine, Aster,

	Flower, Marigold, Rose, Gelardia, Billion Dollar Grass (Sawan), Mahua.

Table 1: VDSA Crops Classified

Importantly, Sorghum is the predominant crop in the Millets category accounting for ~84% of cropping instances and 92% of cropping area in the category.

Crop Name	Instances	Instances (Percentage)	Crop Area	Crop Area (Percentage)
Sorghum	3209	84	6479.591	91.722
Pearl Millet	277	7	357.4684	5.060
Finger Millet	338	9	213.7045	3.025
Minor Millet	6	0	13.5	0.191
Small Millet (Marua)	1	0	0.1	0.001

Table 2: Sorghum is the predominant crop in the Millets category

Based on this crop classification, we can also see (Tables 3 and 4) that while millets account for only ~5% of area cultivated in the Kharif season, they account for over a third of the area cultivated in Rabi.

Crop Category	Frequency (Count)	Percentage (Count)	Crop Area (in acres)	Percentage of Area Cultivated
Cereals	9201	35.5	6088.96	19.77
Commercial	6654	25.7	15962.25	51.82
Fruits and Vegetables	1501	5.8	1137.73	3.69
Millets	1425	5.5	1306.71	4.24
Pulses	7132	27.5	6274.78	20.37

Table 3: Frequency and Area Cultivated by Crop Categories (Kharif)

Crop Category	Frequency (Count)	Percentage (Count)	Crop Area (in acres)	Percentage of Area Cultivated
Cereals	5322	42.94	6910.53	39.1
Commercial	966	7.79	1025.6	5.8
Fruits and Vegetables	1339	10.8	789.61	4.5
Millets	2421	19.54	5761.62	32.6
Pulses	2345	18.92	3195.5	18.1

Table 4: Frequency and Area Cultivated by Crop Categories (Rabi)

Based on this crop classification, we arrive at the following cropping systems for our sample, season-wise and for the year (i.e., across Kharif and Rabi) – please see Table 5 – Table 7. As expected, the Millet + (Other Crops) cropping system accounts for a relatively lower share i.e., 16% of cropping instances and 27.3% of area cultivated in the Kharif season. In contrast, it accounts for 32% of cropping instances and ~42% of cultivated area in Rabi season.

Assessed for a yearly cropping system basis, the Millet + (Other Crops) cropping system accounts for 34% of cropping instances and 44.1% of area cultivated.

Cropping Pattern (Kharif)	Count	Percentage (Count)	Area	Percentage (Area)
Commercial + Pulses	1081	14.6	6276.9	20.4
Commercial + Millet + Pulses	623	8.4	4784.0	15.5
Cereals	1950	26.4	3839.7	12.5
Commercial	903	12.2	3711.7	12.0
Cereals + Commercial	312	4.2	1469.9	4.8
Cereals + Commercial + Pulses	200	2.7	1436.4	4.7
Pulses	440	5.9	1300.2	4.2
Commercial + Fruits and Vegetables + Millet + Pulses	81	1.1	1174.8	3.8
Cereals + Pulses	434	5.9	1132.5	3.7
Commercial + Fruits and Vegetables + Pulses	127	1.7	1112.3	3.6
Cereals + Commercial + Millet + Pulses	136	1.8	1086.9	3.5
Cereals + Commercial + Fruits and Vegetables + Pulses	65	0.9	538.6	1.7
Fruits and Vegetables + Pulses	105	1.4	394.4	1.3
Fruits and Vegetables	287	3.9	367.0	1.2
Cereals + Commercial + Fruits and Vegetables + Millet + Pulses	45	0.6	359.7	1.2
Cereals + Fruits and Vegetables	131	1.8	284.8	0.9
Commercial + Millet	47	0.6	279.3	0.9
Cereals + Fruits and Vegetables + Pulses	64	0.9	195.4	0.6
Commercial + Fruits and Vegetables	50	0.7	189.4	0.6
Millet + Pulses	113	1.5	184.3	0.6
Cereals + Commercial + Millet	25	0.3	156.3	0.5
Cereals + Commercial + Fruits and Vegetables	29	0.4	139.4	0.5
Cereals + Millet + Pulses	48	0.6	104.5	0.3
Cereals + Millet	32	0.4	103.5	0.3
Millet	41	0.6	72.3	0.2
Cereals + Fruits and Vegetables + Millet + Pulses	6	0.1	37.0	0.1
Commercial + Fruits and Vegetables + Millet	2	0	20.5	0.1
Fruits and Vegetables + Millet + Pulses	4	0.1	18.5	0.1
Cereals + Fruits and Vegetables + Millet	7	0.1	16.8	0.1
Cereals + Commercial + Fruits and Vegetables + Millet	2	0	12.0	0.0
Fruits and Vegetables + Millet	7	0.1	7.5	0.0

Table 5: Kharif Cropping Systems by Count and Area Cultivated

Cropping Pattern (Rabi)	Count	Percentage (Count)	Area	Percentage (Area)
Cereals	1578	30.1	3235.6	18.3
Millet	939	17.9	3189.5	18.0
Cereals + Pulses	425	8.1	2643.8	15.0

Cereals + Millet	270	5.1	1047.8	5.9
Cereals + Commercial + Pulses	96	1.8	823.1	4.7
Pulses	411	7.8	785.9	4.4
Cereals + Millet + Pulses	89	1.7	727.9	4.1
Millet + Pulses	130	2.5	594.5	3.4
Cereals + Commercial + Millet + Pulses	25	0.5	545.6	3.1
Cereals + Fruits and Vegetables	130	2.5	502.9	2.8
Cereals + Commercial	107	2	381.0	2.2
Cereals + Fruits and Vegetables + Pulses	92	1.8	371.4	2.1
Fruits and Vegetables	317	6	346.0	2.0
Commercial	152	2.9	317.3	1.8
Cereals + Commercial + Fruits and Vegetables + Pulses	111	2.1	294.4	1.7
Fruits and Vegetables + Pulses	51	1	220.7	1.2
Fruits and Vegetables + Millet	41	0.8	182.3	1.0
Commercial + Millet + Pulses	25	0.5	180.8	1.0
Commercial + Millet	35	0.7	176.8	1.0
Cereals + Fruits and Vegetables + Millet + Pulses	22	0.4	173.8	1.0
Cereals + Fruits and Vegetables + Millet	29	0.6	165.5	0.9
Cereals + Commercial + Fruits and Vegetables	27	0.5	109.7	0.6
Cereals + Commercial + Millet	16	0.3	104.3	0.6
Commercial + Pulses	31	0.6	94.8	0.5
Commercial + Fruits and Vegetables + Millet + Pulses	10	0.2	91.5	0.5
Fruits and Vegetables + Millet + Pulses	14	0.3	83.4	0.5
Commercial + Fruits and Vegetables	33	0.6	82.8	0.5
Cereals + Commercial + Fruits and Vegetables + Millet + Pulses	14	0.3	72.6	0.4
Commercial + Fruits and Vegetables + Millet	8	0.2	56.8	0.3
Cereals + Commercial + Fruits and Vegetables + Millet	6	0.1	45.2	0.3
Commercial + Fruits and Vegetables + Pulses	15	0.3	35.3	0.2

Table 6: Rabi Cropping Systems by Count and Area Cultivated

Cropping Pattern (Annual)	Count	Percentage (Count)	Area	Percentage (Area)
Cereals + Commercial + Pulses	669	8.1	8112.5	16.7
Cereals + Commercial + Millet + Pulses	423	5.1	5648.8	11.6
Commercial + Millet + Pulses	523	6.4	3712.2	7.7
Commercial + Pulses	662	8	3101.9	6.4
Cereals	1270	15.4	2939.6	6.1
Cereals + Commercial + Fruits and Vegetables + Millet + Pulses	188	2.3	2827.7	5.8
Cereals + Pulses	735	8.9	2623.3	5.4
Cereals + Commercial	397	4.8	2582.1	5.3
Cereals + Commercial + Fruits and Vegetables + Pulses	235	2.9	2139.5	4.4
Commercial	493	6	1703.7	3.5
Millet + Pulses	331	4	1591.8	3.3

Millet	398	4.8	1221.7	2.5
Commercial + Fruits and Vegetables + Millet + Pulses	102	1.2	1121.8	2.3
Cereals + Millet + Pulses	118	1.4	821.8	1.7
Cereals + Commercial + Millet	108	1.3	803.8	1.7
Cereals + Fruits and Vegetables	215	2.6	691.7	1.4
Cereals + Fruits and Vegetables + Millet + Pulses	75	0.9	684.0	1.4
Cereals + Commercial + Fruits and Vegetables	100	1.2	673.8	1.4
Commercial + Fruits and Vegetables	100	1.2	621.4	1.3
Cereals + Fruits and Vegetables + Millet	110	1.3	613.4	1.3
Pulses	178	2.2	581.6	1.2
Fruits and Vegetables + Millet + Pulses	77	0.9	573.4	1.2
Fruits and Vegetables + Millet	106	1.3	522.5	1.1
Cereals + Fruits and Vegetables + Pulses	152	1.8	514.5	1.1
Commercial + Fruits and Vegetables + Pulses	60	0.7	506.7	1.0
Cereals + Millet	121	1.5	471.7	1.0
Commercial + Millet	73	0.9	392.7	0.8
Cereals + Commercial + Fruits and Vegetables + Millet	31	0.4	225.2	0.5
Fruits and Vegetables + Pulses	40	0.5	193.2	0.4
Fruits and Vegetables	113	1.4	139.0	0.3
Commercial + Fruits and Vegetables + Millet	21	0.3	131.9	0.3

Table 7: Yearly Cropping Systems by Frequency and Area Cultivated

Based on Table 5 – Table 7, we conclude that there are three distinct types of cropping systems where credit impact may be studied: Cereals + (Other Crops), Commercial + (Other Crops) and Millet + (Other Crops). However, whether crop-specific credit impacts can be studied within any of these cropping system categories, remains to be seen. To check this, we first identify the major crops in our data to map them back to these cropping systems. Table 8 presents the top crops by share of cultivated area accounting for ~80% of total cultivated area across the years. Among these, isolating crops with at least 5% of cultivated area across the years leads to a list of 9 major crops with Sorghum being the top crop by cultivated area.

Crop Name	Crop Category	Cultivated Area	Percentage
Sorghum	Millet	6480.6	12.8
Cotton	Commercial	6377	12.6
Paddy	Cereals	5935.9	11.7
Wheat	Cereals	5701.2	11.3
Soybean	Commercial	5269.1	10.4
Pigeonpea	Pulses	4162.2	8.2
Chickpea	Pulses	2567.8	5.1
Groundnut	Commercial	2426.5	4.8
Maize	Cereals	1358	2.7
Sugarcane	Commercial	1164.4	2.3

Blackgram	Pulses	1035.1	2
Greengram	Pulses	1000.3	2
Sunflower	Commercial	828.2	1.6
Others	NA	725.3	12.5

Table 8: Crop-Wise Cultivated Area (Excluding Fallow Land)

So the major crop-wise cropping systems (accounting for at least 5% of cultivated area individually over the years and ~80% of cultivated area in total) are:

Sorghum + (Other Crops)	Soybean + (Other Crops)	Sugarcane + (Other Crops)
Cotton + (Other Crops)	Pigeonpea + (Other Crops)	
Wheat + (Other Crops)	Chickpea + (Other Crops)	
Paddy + (Other Crops)	Groundnut + (Other Crops)	

Now, the question is, for how many of these crops can we potentially study credit impact. In case of propensity score matching, this would involve a comparison of cropping pattern across the treatment and potential controls group. However, before we make this assessment for a propensity score matching (PSM) type of setup, we must rule out the possibility of a panel-data type of analysis. To do this, we first document the total cropping frequency for each crop over the study duration i.e., 2001-2014 (see Table 9A and Table 9B) to make the case that even if panel were possible, it would be a short panel. Then we will try to isolate the sub-sample where both pre-treatment and post-treatment data is available (in at least 1 year) and assess if that sample is significantly different from the panel of households that does not fulfil these criteria (see Table 10).

Cropping Frequency	**Sorghum**	**Cotton**	**Wheat**	**Paddy**	**Soybean**	**Pigeonpea**	**Chickpea**	**Groundnut**	**Sugarcane**
1	125	98	126	76	106	205	129	109	13
2	70	64	105	56	16	112	83	62	20
3	59	51	56	76	14	97	64	36	8
4	47	41	72	102	16	63	39	29	14
5	40	30	128	287	18	85	27	27	13
6	41	27	53	43	27	85	36	61	6
7	35	25	19	12	17	31	.	1	7
8	34	15	14	8	28	20	2	2	3
9	11	6	11	5	9	17	1	.	9
10	17	12	5	6	.	12	.	.	20
11	7	7	1	6	2	14	.	.	4
12	6	11	3	8	5	8	.	.	6
13	14	4	1	5	1	9	.	.	2
14	14	16	.	5	.	18	.	.	3

Table 9A: HH-Level Crop-Wise Cropping Frequency

Cropping Frequency	Sorghum	Cotton	Wheat	Paddy	Soybean	Pigeonpea	Chickpea	Groundnut	Sugarcane
1	24.0	24.1	21.2	10.9	40.9	26.6	33.9	33.3	10.2
2	37.5	39.9	38.9	19.0	47.1	41.1	55.6	52.3	25.8
3	49.8	52.5	48.3	29.9	52.5	53.6	72.4	63.3	32.0
4	57.9	62.6	60.4	44.6	58.7	61.8	82.7	72.2	43.0
5	65.6	70.0	82.0	85.9	65.6	72.8	89.8	80.4	53.1
6	73.5	76.6	90.9	92.1	76.1	83.8	99.2	99.1	57.8
7	80.2	82.8	94.1	93.8	82.6	87.8	99.2	99.4	63.3
8	86.7	86.5	96.5	95.0	93.4	90.4	99.7	100.0	65.6
9	88.8	87.9	98.3	95.7	96.9	92.6	100.0	100.0	72.7
10	92.1	90.9	99.2	96.5	96.9	94.2	100.0	100.0	88.3
11	93.5	92.6	99.3	97.4	97.7	96.0	100.0	100.0	91.4
12	94.6	95.3	99.8	98.6	99.6	97.0	100.0	100.0	96.1
13	97.3	96.3	100.0	99.3	100.0	98.2	100.0	100.0	97.7
14	100.0	100.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 9B: HH-Level Crop-Wise Cropping Frequency (Cumulative Percentage)

From tables 9A and 9B, we establish that the median cropping frequency across most major crops including Sorghum, Cotton, Soybean, Pigeonpea, Chickpea, and Groundnut is three years or fewer over the fourteen-year study period. Paddy and Sugarcane have slightly longer median cropping durations of five years, while Wheat falls at four years. These patterns indicate that only a short panel would be feasible for any of the major crops. However, the more fundamental concern is not panel length per se, but the selection bias that would arise from restricting the sample to households with both pre- and post-treatment observations. Table 10 shows that among formal borrowers, the share of households with at least one year of pre-borrowing and one year of post-borrowing data ranges from approximately 13% for Soybean to 28% for Sugarcane. Retaining only this subsample would yield a group of households that differs systematically from the broader population of borrowers Given both the short panel length and the severity of the resulting selection problem, a panel-based identification strategy is not pursued for any of the major cropping systems.

Crop-Wise Cropping Pattern	Count of Formal Borrowers with at least 1 year of pre-borrowing & post-borrowing data (Percentage of Total Formal Borrowers)	Count of Informal Borrowers with at least 1 year of pre-borrowing & post-borrowing data (Percentage of Total Informal Borrowers)
Sorghum+	89 (22%)	78 (22.4%)
Cotton+	66 (19.5%)	66 (21.4%)
Wheat+	67 (16.8%)	65 (20%)
Paddy+	67 (20.1%)	90 (27.5%)

Soybean+	25 (13%)	19 (10.8%)
Pigeonpea+	112 (21%)	121 (23.5%)
Chickpea+	36 (12.6%)	25 (9.9%)
Groundnut+	32 (13.1%)	37 (15.9%)
Sugarcane+	31 (27.7%)	13 (20.1%)

Table 10: Percentage of borrowers with at least one year of pre- and post- data.

Now that a panel analysis is ruled out, we examine for which crops, a cropping-pattern specific propensity score matching analysis is possible. To assess this, we construct tables with year-wise count of first-time borrowers and potential counterfactuals for each crop (Table 11A – Table 11I). This is in line with the most relaxed matching criterion i.e., year-by-year matching. The rows highlighted indicate years where the count of treated households exceed the potential number of controls. For all crops except Sorghum and Pigeonpea, there are at least 5 years where the number of treated households exceeds the potential number of control households. Importantly, the crops that are ruled out from this step cannot be studied in a *cropping-pattern-specific* manner anyway since this would be an even stricter matching criterion.

Year	Sorghum (Formal Credit)		Sorghum (Informal Credit)	
	Potential Controls	First-Time Borrowers	Potential Controls	First-Time Borrowers
2001	98	59	98	16
2002	68	23	68	21
2003	61	16	61	16
2004	50	13	50	5
2005	77	34	77	7
2006	45	22	45	21
2007	51	20	51	13
2008	49	13	49	14
2009	39	35	39	32
2010	37	18	37	21
2011	27	10	27	22
2012	22	15	22	9
2013	15	7	15	5
2014	17	5	17	1

Table 11A: Count of Borrowers and Potential Counterfactuals (Sorghum)

Year	Cotton (Formal Credit)		Cotton (Informal Credit)	
	Potential Controls	First-Time Borrowers	Potential Controls	First-Time Borrowers
2001	49	39	49	22
2002	31	16	31	8
2003	20	15	20	18
2004	23	13	23	5

2005	24	25	24	7
2006	15	11	15	15
2007	6	11	6	17
2008	8	19	8	28
2009	8	16	8	13
2010	21	17	21	11
2011	24	17	24	10
2012	18	4	18	18
2013	13	18	13	13
2014	10	7	10	8

Table 11B: Count of Borrowers and Potential Counterfactuals (Cotton)

Year	**Wheat (Formal Credit)**		**Wheat (Informal Credit)**	
	Potential Controls	**First-Time Borrowers**	**Potential Controls**	**First-Time Borrowers**
2001	32	37	32	6
2002	7	10	7	6
2003	1	1	1	2
2004	6	1	6	2
2005	20	26	20	6
2006	21	21	21	9
2007	23	17	23	12
2008	9	4	9	1
2009	51	39	51	28
2010	121	65	121	31
2011	104	31	104	30
2012	87	14	87	13
2013	113	13	113	22
2014	85	15	85	27

Table 11C: Count of Borrowers and Potential Counterfactuals (Wheat)

Year	**Paddy (Formal Credit)**		**Paddy (Informal Credit)**	
	Potential Controls	**First-Time Borrowers**	**Potential Controls**	**First-Time Borrowers**
2001	24	7	24	13
2002	15	8	15	8
2003	7	11	7	6
2004	9	10	9	4
2005	9	16	9	8
2006	9	5	9	10
2007	2	3	2	8
2008	3	7	3	8
2009	52	4	52	16
2010	268	97	268	36
2011	240	43	240	49
2012	234	30	234	22
2013	225	12	225	15

2014	203	11	203	28

Table 11D: Count of Borrowers and Potential Counterfactuals (Paddy)

Year	**Soybean (Formal Credit)**		**Soybean (Informal Credit)**	
	Potential Controls	**First-Time Borrowers**	**Potential Controls**	**First-Time Borrowers**
2001	0	2	0	0
2002	0	2	0	0
2003	1	2	1	1
2004	3	3	3	0
2005	0	3	0	1
2006	1	11	1	2
2007	4	19	4	9
2008	32	30	32	19
2009	20	15	20	25
2010	29	6	29	11
2011	22	7	22	7
2012	9	8	9	8
2013	11	7	11	10
2014	13	5	13	5

Table 11E: Count of Borrowers and Potential Counterfactuals (Soybean)

Year	**Pigeonpea (Formal Credit)**		**Pigeonpea (Informal Credit)**	
	Potential Controls	**First-Time Borrowers**	**Potential Controls**	**First-Time Borrowers**
2001	56	46	56	17
2002	74	33	74	21
2003	52	34	52	25
2004	54	19	54	8
2005	48	41	48	16
2006	55	15	55	25
2007	31	18	31	22
2008	33	28	33	26
2009	88	32	88	54
2010	142	33	142	48
2011	80	20	80	22
2012	67	8	67	14
2013	71	16	71	9
2014	51	9	51	12

Table 11F: Count of Borrowers and Potential Counterfactuals (Pigeonpea)

Year	**Chickpea (Formal Credit)**		**Chickpea (Informal Credit)**	
	Potential Controls	**First-Time Borrowers**	**Potential Controls**	**First-Time Borrowers**
2001	13	7	13	0
2002	2	4	2	0
2003	8	4	8	1
2004	11	1	11	1

2005	20	9	20	1
2006	6	7	6	3
2007	4	2	4	1
2008	14	3	14	0
2009	41	31	41	26
2010	56	29	56	29
2011	23	26	23	18
2012	25	27	25	16
2013	35	15	35	11
2014	19	8	19	9

Table 11G: Count of Borrowers and Potential Counterfactuals (Chickpea)

Year	Groundnut (Formal Credit)		Groundnut (Informal Credit)	
	Potential Controls	First-Time Borrowers	Potential Controls	First-Time Borrowers
2001	6	7	6	2
2002	8	6	8	4
2003	7	2	7	0
2004	18	11	18	0
2005	16	7	16	4
2006	15	5	15	1
2007	8	7	8	2
2008	6	5	6	5
2009	43	52	43	24
2010	45	17	45	28
2011	15	16	15	19
2012	16	9	16	16
2013	19	4	19	11
2014	13	8	13	6

Table 11H: Count of Borrowers and Potential Counterfactuals (Groundnut)

Year	Sugarcane		Sugarcane	
	Potential Controls	First-Time Borrowers	Potential Controls	First-Time Borrowers
2001	7	17	7	1
2002	4	8	4	0
2003	0	2	0	0
2004	6	4	6	2
2005	18	16	18	4
2006	22	6	22	2
2007	8	16	8	5
2008	3	2	3	0
2009	3	2	3	3
2010	2	7	2	9
2011	7	5	7	0
2012	6	1	6	2
2013	7	0	7	3

2014	5	3	5	2

Table 11I: Count of Borrowers and Potential Counterfactuals (Sugarcane)

Among Sorghum and Pigeonpea, we assess whether a household that cultivates each of these crops is a principal grower/cultivator (i.e., allocates a majority operational holding share) of that crop. Only if a majority of households are principal growers (or cultivators) of a crop, does it make sense to study that crop in isolation. Because if not, then defining a household's portfolio based on that crop makes no sense. To operationalise the notion of a principal cultivator, we define a household as a principal cultivator of a crop if that crop accounts for at least 50% of its total operational holding in a given season-year. This threshold is chosen to reflect a farming portfolio in which the crop in question constitutes the dominant land use, such that farm-level credit and input decisions are meaningfully organized around it.

To contextualise this threshold, we compute the average share of cultivated area allocated to each of the 125 crops in the VDSA data. The median share across all crop-household-year observations is 0.2 and the 90th percentile is 0.5 suggesting that a 50% threshold is a demanding criterion in our multiple cropping context, that most crops in the data do not meet. For Sorghum, the average share of cultivated area across sorghum-growing households and years is 0.5, indicating that the typical sorghum-growing household allocates the majority of its operational holding to sorghum and can therefore be regarded as a principal cultivator. For Pigeonpea, the corresponding average share is below 0.2, falling at the median of the full crop distribution, which implies that most pigeonpea-growing households cultivate it as a secondary crop rather than as the primary basis of their farming system. This distinction is consequential for the analysis: defining a household's farming portfolio around a crop that accounts for a minor share of its land would produce a misleading characterisation of the household's production environment and credit needs. On this basis, we proceed with Sorghum.

Credit Impact Analysis for Sorghum-Based Cropping Systems:

In the first step, I simplify the 21 distinct cropping patterns among sorghum growers to the three categories detailed in major revision 1 summary i.e., Sorghum (Rabi/Kharif) + Pulses/Commercial (Kharif), Sorghum Only (Rabi) / Sorghum (Rabi) + Cereals (Rabi) and Sorghum (Overall). None of the years are a major cause of concern in terms of potential counterfactual availability. So we conduct this matching exercise by cropping system, define the outcome variables as the area-weighted yield for each cropping system and rerun the yield regressions with credit to assess credit impact. Figures 1 – 6 provide a snapshot of the preliminary results.

Preliminary Results:

Figure 1: Formal Credit Impact on Sorghum Yields for Sorghum-Growing HHs. N = 731

($N_{\text{Borrowers}} = 226$ | $N_{\text{Non-Borrowers}} = 505$)

Interpretation: $\hat{\beta}_{\tau}$ is the is the estimated difference in the τ-th quantile of sorghum yields between HHs that took credit and (matched) HHs that did not take credit, among households that principally grew sorghum — holding other covariates (inputs, soil type, seed type, weather and farmer experience) constant.

Sorghum Only/ Sorghum + Cereals (Wheat):

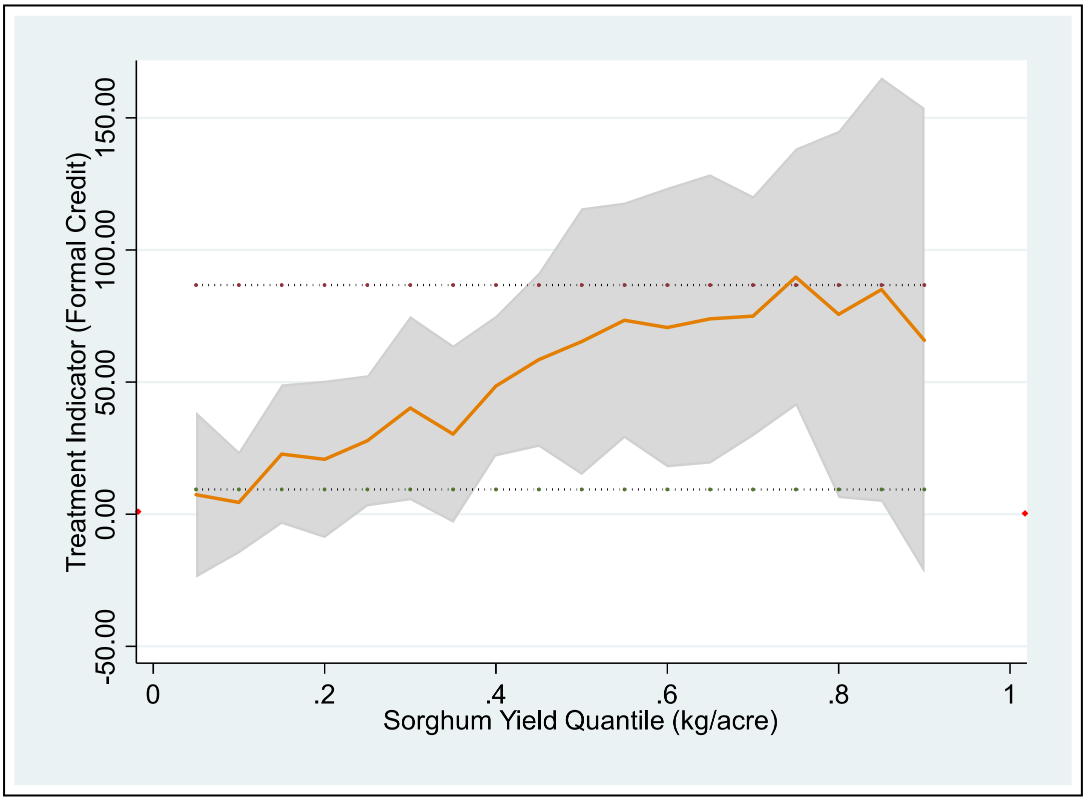


Figure 2: Formal Credit Impact on Sorghum Yields for Sorghum + Cereal Growing Households. N = 486

($N_{\text{Borrowers}} = 126$ | $N_{\text{Non-Borrowers}} = 360$)

Interpretation: $\hat{\beta}_{\tau}$ is the is the estimated difference in the τ-th quantile of sorghum yields between HHs that took credit and (matched) HHs that did not take credit, among households that cultivate sorghum and intra-seasonally crop it with cereals — holding other covariates (inputs, soil type, seed type, weather and farmer experience) constant.

Average Impact on Cereal Yields: 97.97** (46.54)

Quantile Effects: See Figure 4.

Sorghum + Pulses/Commercial (Kharif):

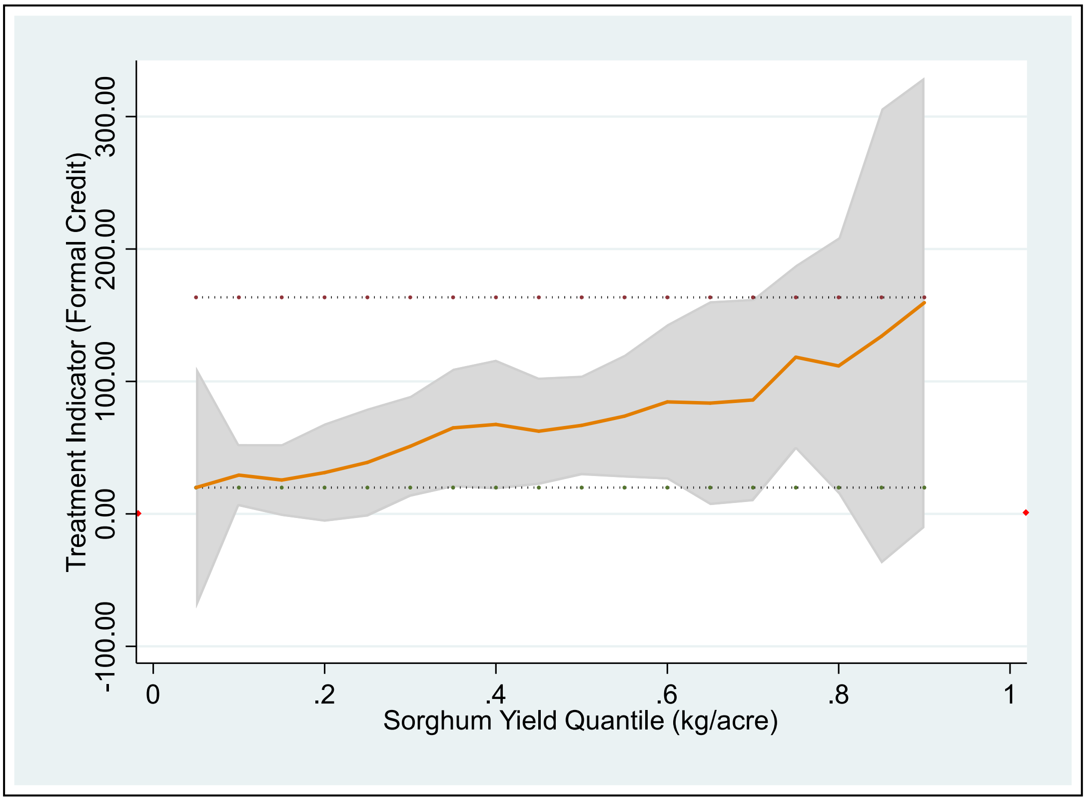


Figure 3: Formal Credit Impact on Sorghum Yields for Sorghum + Pulses/ Commercial Crop Growing HHs. N = 452

(N $_{\text{Borrowers}}$ = 157 | N $_{\text{Non-Borrowers}}$ = 295)

Interpretation: $\hat{\beta}_{\tau}$ is the is the estimated difference in the τ-th quantile of sorghum yields between HHs that took credit and (matched) HHs that did not take credit, among households that grew sorghum and intra/inter-seasonally crop it with pulses and/or commercial (oilseed) crops — holding other covariates (inputs, soil type, seed type, weather and farmer experience) constant.

Average Impact on Pulses Yields: 74.91** (40.41)

Quantile Effects Pulses Yields: See Figure 5.

Average Impact on Commercial Crops Yields: 42.69* (24.1)

Quantile Effects Commercial Yields: See Figure 6.

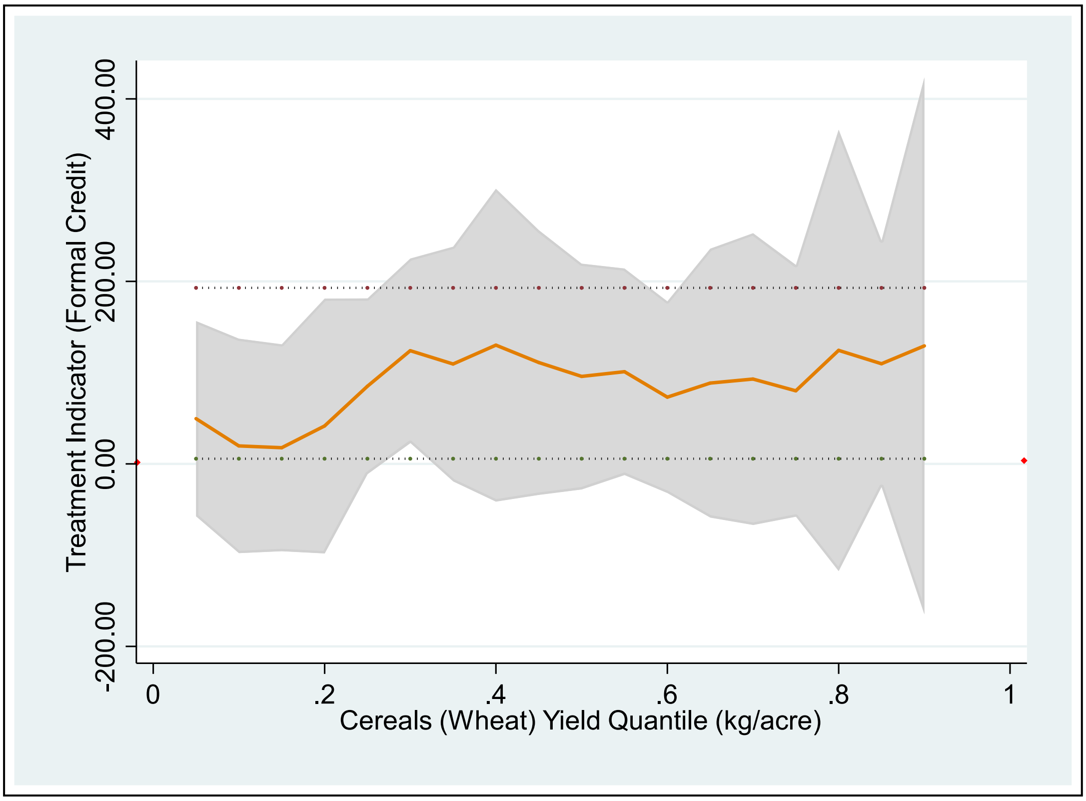


Figure 4: Formal Credit Impact on Cereal Yields for Sorghum + Cereal Growing HHs. N = 172 (Rest are Sorghum Only HHs)

(N $_{\text{Borrowers}}$ = 64 | N $_{\text{Non-Borrowers}}$ = 108)

Interpretation: $\hat{\beta}_{\tau}$ is the is the estimated difference in the τ-th quantile of cereal (wheat) yields between HHs that took credit and (matched) HHs that did not take credit, among households that cultivate sorghum and intra-seasonally crop it with cereals — holding other covariates (inputs, soil type, seed type, weather and farmer experience) constant.

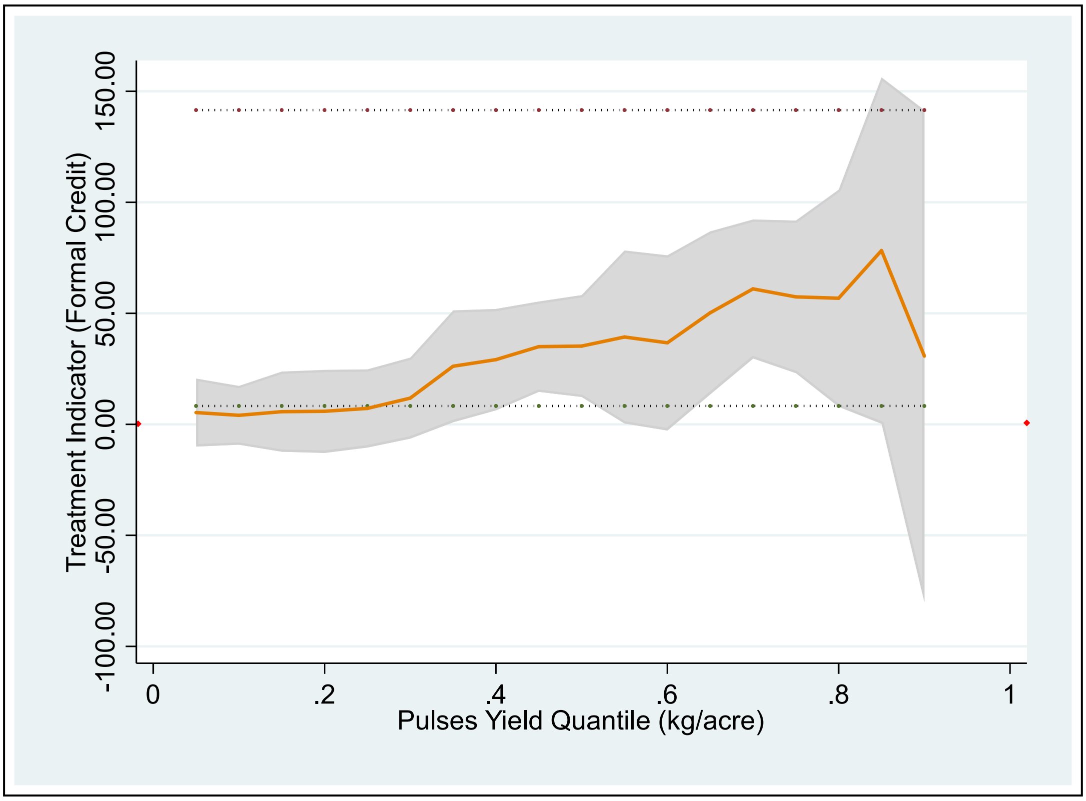


Figure 5: Formal Credit Impact on Pulses Yields for Sorghum + Commercial/Pulse Growing HHs. N = 407 (Rest are Commercial Only HHs) | (N $_{\text{Borrowers}}$ = 136 | N $_{\text{Non-Borrowers}}$ = 271)

Interpretation: $\hat{\beta}_{\tau}$ is the is the estimated difference in the τ-th quantile of pulses (pigeonpea, greengram, blackgram, horsegram) yields between HHs that took credit and (matched) HHs that did not take credit, among households that cultivate sorghum and intra/inter-seasonally crop it with pulses and other commercial crops (e.g., cotton and oilseeds) — holding other covariates (inputs, soil type, seed type, weather and farmer experience) constant.

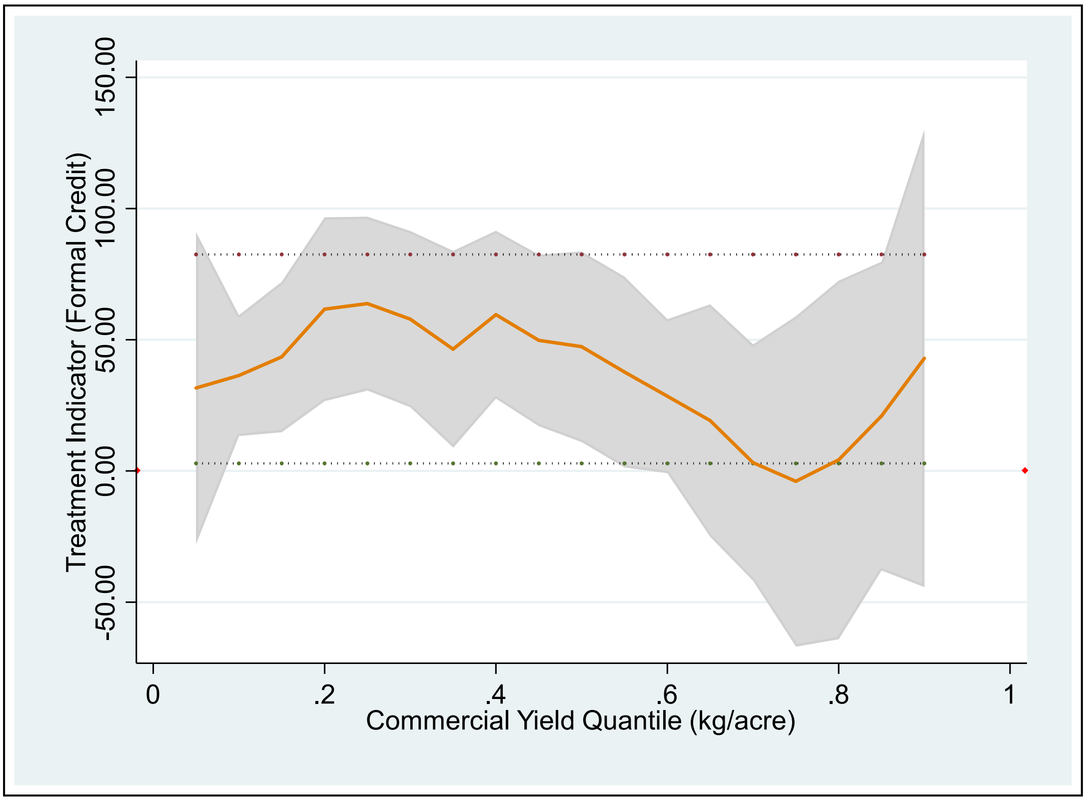


Figure 6: Formal Credit Impact on Commercial (i.e., cotton and oilseeds) Yield for Sorghum + Commercial/Pulse Growing HHs. N = 333 (Rest are Pulse Only HHs) | ($N_{\text{Borrowers}}$ = 120 | $N_{\text{Non-Borrowers}}$ = 213)

Interpretation: $\hat{\beta}_{\tau}$ is the is the estimated difference in the τ-th quantile of pulses (pigeonpea, greengram, blackgram, horsegram) yields between HHs that took credit and (matched) HHs that did not take credit, among households that cultivate sorghum and intra/inter-seasonally crop it with pulses and other commercial crops (e.g., cotton and oilseeds) — holding other covariates (inputs, soil type, seed type, weather and farmer experience) constant.

Appendix 1

(II)

Mandi Data Scraping Details:

- The total number of mandis in India (as available on the AgMarknet website) as on 22nd April, 2026 is 4299.
- The mandi-locations are available under the Market Profile tab.
    - The market profile tab has a list of 4284 mandis. So there is a discrepancy of 15 mandis which are there in the master list but for which we do not have a market profile available.
- Among the 4284 mandis for which a market profile is available, the location is not available for 1829 mandis.
    - We did this exercise for ~300 mandis and figured out that relying on any other APIs to automate this process would not work out since the 'top' match (as per Google Earth) is generally not the correct one. And filtering among top and non-top matches would typically be more cumbersome than just looking the Mandi up on Google Earth.
    - Moreover, looking up the Mandi location is not straightforward and must account for vernacular inconsistencies etc. The process of manual geocoding we followed in using Google Earth and the rationale for this exercise is detailed below.
- For the list of VDSA districts i.e., a total of 15 districts, we generated a set of contiguous districts. This led to a total of 98 unique districts across 11 states. These districts contain a total of 759 mandis for the period 2001-2026. Of these 759 mandis, we were able to find the locations (from the Market Profile) for 576 mandis. 173 more were in the Market Profile list but the latitude, longitude was missing. 10 were not in the Market Profile list at all. So we need to manually find geolocations for 183 mandis.
- So finally, we have a list of 759 geocoded mandis with the type of geocoding as follows:

Geocoding Type	Number of Mandis	Comments
Market List (as available on AgMarkNet)	573	• Technically, the records were available for 576 mandis but one of them was (0,0) so it was done manually instead.

		• 2 more entries needed correction for latitude and longitude values being identical. So again, these were done manually instead. • Additionally, 1 of the 573 entries needed correction for a missing decimal. • Another entry had a latitude value replaced by the longitude value and vice versa.
Manual Geocoding (using Google Earth)	179	• Protocol described below.
Manual Geocoding (using Google Earth) attempted but unsuccessful.	7	Could not be located.

One thing to consider is that this list of mandis was generated based on the entire data available from 2001 to 2026. However, it is possible that some of these mandis were only instituted post 2014. In that case, it is best to include only those mandis in the master list of potential matches that have 'some' data between 2001 to 2014. Based on this, we exclude 152 mandis. Figure 4 below provides the comparison of the 2001-2026 mandis list with the 2001-14 mandis list. As expected, the maximum changes are seen in places where the mandi density was relatively lower to begin with.

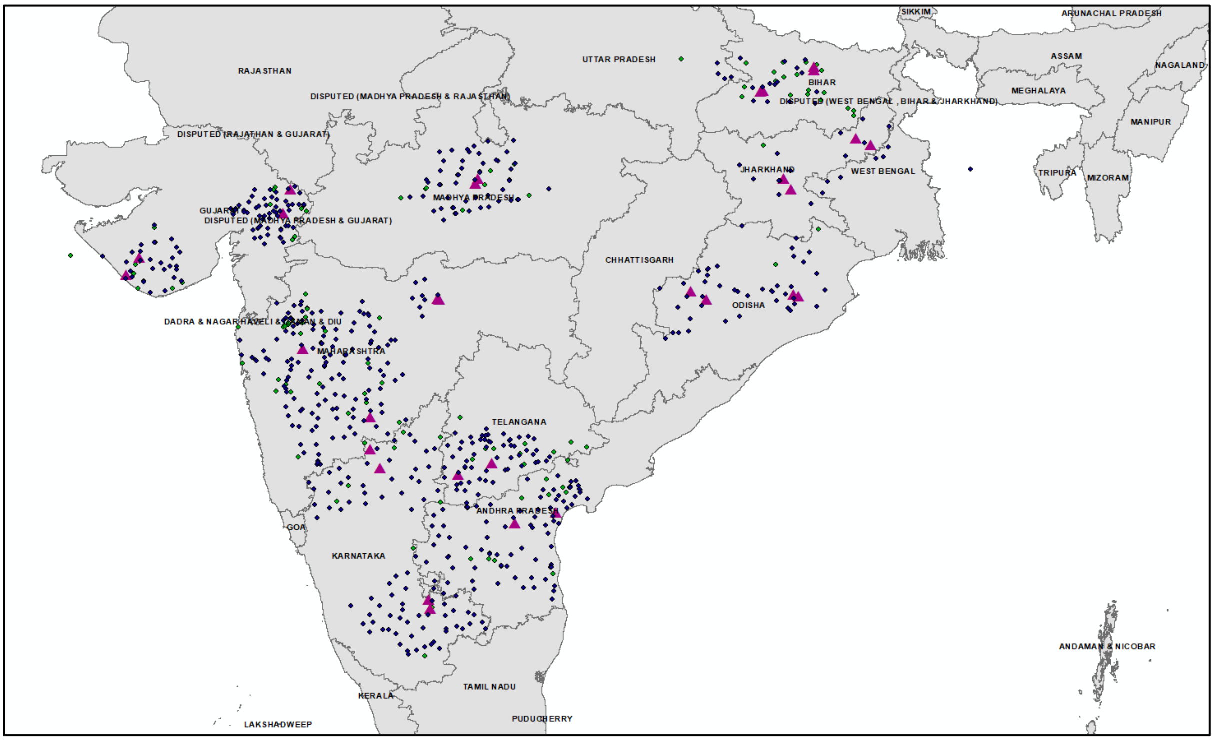


Figure 4: VDSA Villages with Mandi Locations. The pink triangles are VDSA villages, the green points are mandi locations across all years. The dark blue points are mandi locations where prices data is available for 2001-14. **Note:** Shapefiles sourced from Survey of India.

Prices Data:

The total count of rows for the 2001-2014 period is 31,439,936. If you collapse this by year and commodity for each Mandi – 368,714 rows. There are 336 distinct commodities across the 14 years period. Cleaning in accordance with VDSA commodities and removal of irrelevant ones leaves us with 210 commodities. For example, see Cowpea local names detailed in Table 13.

Commodity Name	Cleaned Crop Name
Alasande Gram	Cowpea
Cow	Cowpea
Cowpea (Lobia/Karamani)	Cowpea
Cowpea (Veg)	Cowpea
Karamani	Cowpea

Table 13: Cowpea local commodity names in APMC data

The full list of crops and associated cleaned names are given in Table 14.

Commodity Name (Prices Data)	Cleaned Name (per VDSA)
Alasande Gram	Cowpea
Alsandikai	Beans
Amaranthus	Amaranthus
Amla (Nelli Kai)	Amla
Arecanut (Betelnut/Supari)	Arecanut
Arhar (Tur/Red Gram)(Whole)	Pigeonpea
Arhar Dal (Tur Dal)	Pigeonpea
Ashgourd	Gourd
Astera	Aster
Avare Dal	Avestor
Bajra (Pearl Millet/Cumbu)	Pearl Millet
Balekai	Banana
Banana	Banana
Banana – Green	Banana
Barley (Jau)	Barley
Beans	Beans
Beaten Rice	Paddy
Bengal Gram (Gram)(Whole)	Chickpea
Bengal Gram Dal (Chana Dal)	Chickpea
Ber (Zizyphus/Borehannu)	Ber
Betal Leaves	Arecanut
Bhindi (Ladies Finger)	Ladys Finger

Big Gram	Chickpea
Bitter gourd	Bitter Gourd
Black Gram (Urd Beans)(Whole)	Blackgram
Black Gram Dal (Urd Dal)	Blackgram
Borehannu	Ber
Bottle gourd	Bottle Gourd
Brinjal	Brinjal
Broken Rice	Paddy
Bunch Beans	Cluster Bean
Cabbage	Cabbage
Cane	Sugarcane
Capsicum	Capsicum
Carrot	Carrot
Castor Oil	Castor
Castor Seed	Castor
Cauliflower	Cauliflower
Chapparad Avare	Beans(Bokala_Fawa_Runner Beans)
Chennangi (Whole)	Lentil
Chennangi Dal	Lentil
Chikoos (Sapota)	Sapota
Chili Red	Chillies
Chilly Capsicum	Capsicum
Chrysanthemum	Chrysanthemum
Chrysanthemum (Loose)	Chrysanthemum
Cluster beans	Cluster Bean
Coconut	Coconut
Coconut Oil	Coconut
Coconut Seed	Coconut
Colacasia	Colocasia
Copra	Coconut
Coriander (Leaves)	Coriander
Corriander seed	Coriander
Cotton	Cotton
Cotton Seed	Cotton
Cow	Cowpea
Cowpea (Lobia/Karamani)	Cowpea
Cowpea (Veg)	Cowpea
Cucumbar (Kheera)	Cucumber
Cummin Seed (Jeera)	Cumin
Custard Apple (Sharifa)	Custard Apple
Dal (Avare)	Lentil
Dhaincha	Dhaincha
Drumstick	Drumstick
Dry Chillies	Chillies
Dry Fodder	Fodder

Dry Grapes	Grapes
Duster Beans	Beans(Bokala_Fawa_Runner Beans)
Field Pea	Lentil
Foxtail Millet (Navane)	Minor Millet
French Beans (Frasbean)	Beans(Bokala_Fawa_Runner Beans)
Galgal (Lemon)	Lemon
Garlic	Garlic
Gingelly Oil	Sesamum
Ginger (Dry)	Ginger
Ginger (Green)	Ginger
Gram Raw (Chholia)	Chickpea
Gramflour	Chickpea
Grapes	Grapes
Green Avare (W)	Beans(Bokala_Fawa_Runner Beans)
Green Chilli	Chillies
Green Fodder	Fodder
Green Gram (Moong)(Whole)	Greengram
Green Gram Dal (Moong Dal)	Greengram
Green Peas	Green Pea
Ground Nut Oil	Groundnut
Ground Nut Seed	Groundnut
Groundnut	Groundnut
Groundnut (Split)	Groundnut
Groundnut pods (raw)	Groundnut
Guar	Cluster Bean
Guar Seed (Cluster Beans Seed)	Cluster Bean
Guava	Guava
Gurellu	Niger Seed
Hippe Seed	Mahua
Hybrid Cumbu	Green Grass
Indian Beans (Seam)	Beans(Bokala_Fawa_Runner Beans)
Indian Colza (Sarson)	Mustard
Jaffri	Marigold
Jamamkhan	Fodder
Jasmine	Jasmine
Javi	Fodder
Jowar (Sorghum)	Sorghum
Jute	Jute
Jute Seed	Jute
Kabuli Chana (Chickpeas-White)	Chickpea
Kacholam	Ginger
Karamani	Cowpea
Kartali (Kantola)	Soft Gourd

Kharif Mash	Blackgram
Knool Khol	Cabbage
Kodo Millet (Varagu)	Minor Millet
Kulthi (Horse Gram)	Horsegram
Leafy Vegetable	Leafy Vegetable
Lemon	Lemon
Lentil (Masur)(Whole)	Lentil
Lime	Lemon
Linseed	Linseed
Lint	Cotton
Little gourd (Kundru)	Soft Gourd
Long Melon (Kakri)	Smooth Gourd
Lotus	Lotus Seed
Lotus Sticks	Lotus Seed
Lukad	Fodder
Mahua	Mahua
Mahua Seed (Hippe seed)	Mahua
Maize	Maize
Mango	Mango
Mango (Raw-Ripe)	Mango
Marigold (Calcutta)	Marigold
Marigold (loose)	Marigold
Mash	Blackgram
Masur Dal	Lentil
Mataki	Matki
Methi (Leaves)	Fenugreek
Methi Seeds	Fenugreek
Millets	Minor Millet
Moath Dal	Matki
Mustard	Mustard
Mustard Oil	Mustard
Nelli Kai	Amla
Niger Seed (Ramtil)	Niger Seed
Onion	Onion
Onion Green	Onion
Orange	Orange
Other Pulses	Minor Pulse
Paddy (Dhan)(Common)	Paddy
Papaya	Papaya
Papaya (Raw)	Papaya
Patti Calcutta	Lemon
Peas (Dry)	Peas
Peas Wet	Peas
Peas cod	Peas
Pegeon Pea (Arhar Fali)	Pigeonpea
Pepper garbled	Capsicum

Pepper ungarbled	Capsicum
Pointed gourd (Parval)	Soft Gourd
Pomegranate	Pomegranate
Potato	Potato
Pumpkin	Pumpkin
Raddish	Radish
Ragi (Finger Millet)	Finger Millet
Rajgir	Minor Millet
Rat Tail Radish (Mogari)	Radish
Red Gram	Lentil
Rice	Paddy
Ridgeguard (Tori)	Ridge Gourd
Rose (Local)	Rose
Rose (Loose))	Rose
Rose (Tata)	Rose
Round gourd	Smooth Gourd
Safflower	Safflower
Sajje	Pearl Millet
Same/Savi	Cluster Bean
Sarasum	Sesamum
Season Leaves	Mustard
Sesamum (Sesame,Gingelly,Til)	Sesamum
Snakeguard	Smooth Gourd
Soyabean	Soybean
Spinach	Spinach
Sponge gourd	Soft Gourd
Squash (Chappal Kadoo)	Pumpkin
Sugar	Sugarcane
Sugarcane	Sugarcane
Sunflower	Sunflower
Sunflower Seed	Sunflower
Sweet Potato	Sweet Potato
Sweet Pumpkin	Pumpkin
T.V. Cumbu	Pearl Millet
Tamarind Fruit	Tamarind
Tamarind Seed	Tamarind
Tender Coconut	Coconut
Thinai (Italian Millet)	Minor Millet
Thogrikai	Pigeonpea
Thondekai	Smooth Gourd
Tinda	Tinda
Tobacco	Tobacco
Tomato	Tomato
Toria	Soft Gourd
Tube Rose (Double)	Rose
Tube Rose (Loose)	Rose

Tube Rose (Single)	Rose
Turmeric	Turmeric
Turmeric (raw)	Turmeric
Water Melon	Watermelon
Wheat	Wheat
Wheat Atta	Wheat
White Peas	Chickpea
White Pumpkin	Pumpkin

Table 14: Mapping APMC Commodities to VDSA Crops

We select from the set of Mandis plotted in Figure 4, the relevant mandis, i.e., those nearest to the VDSA villages. Then we summarize the crop-category wise prices. Importantly, this is a dynamic set in the sense that only mandis that have some data in a given year are a potential nearest-neighbour match. Figure 5, Figure 6 and Figure 7 plot the crop category wise prices data for Aurepalle village in Andhra Pradesh, Kapanimbargi village in Karnataka and Kanzara village in Maharashtra displaying heterogenieity in crop output prices across the three main states in our sample of hosueholds cultivating sorghum-based cropping systems.

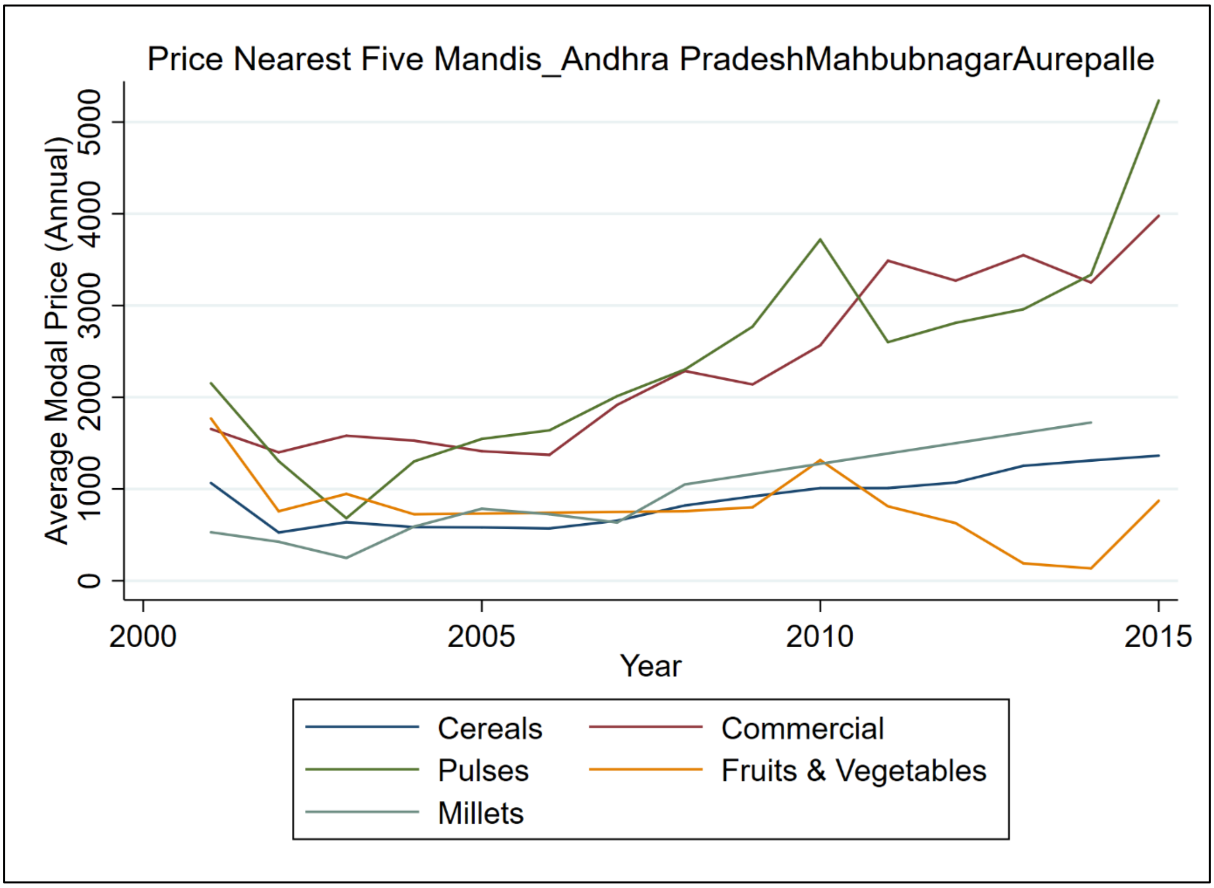


Figure 5: Commodity-Village-Year Wise Prices, Aurepalle Village, Andhra Pradesh

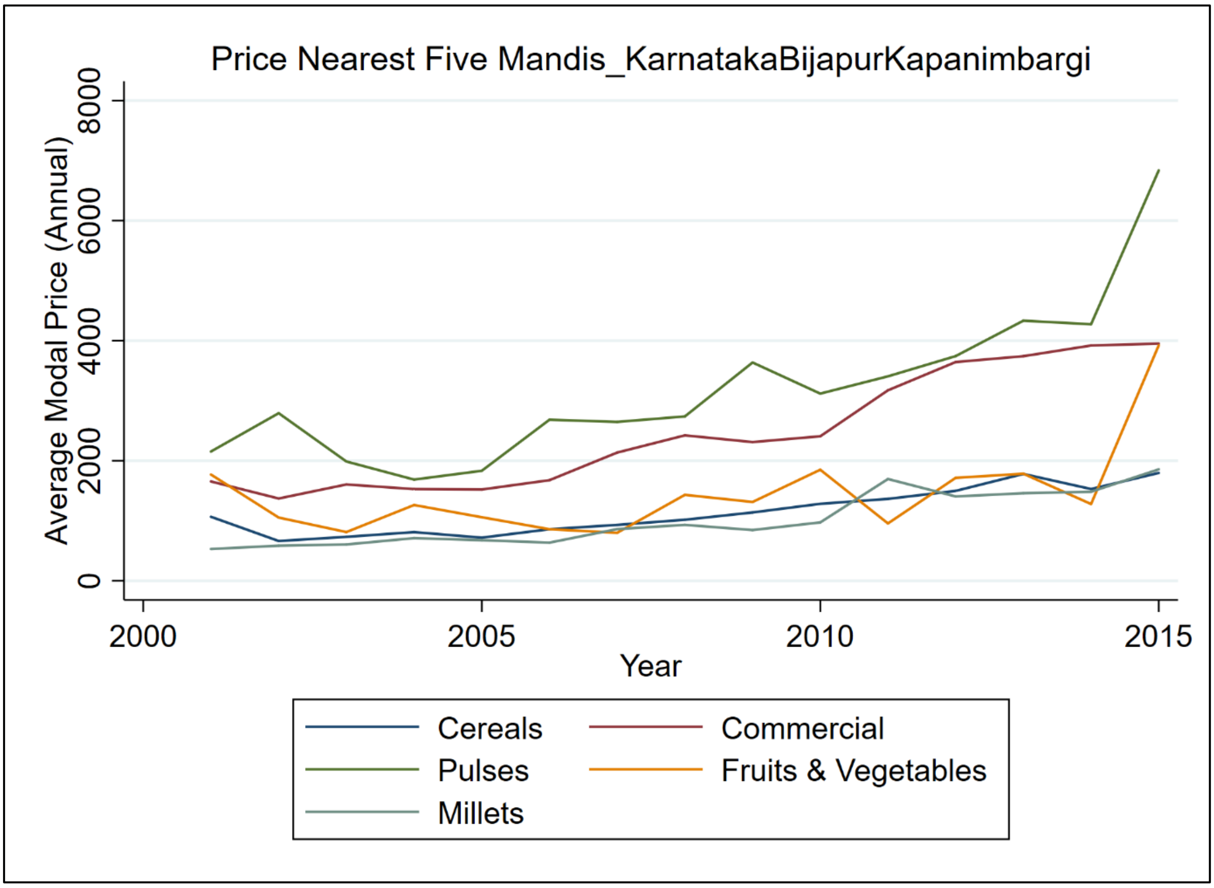


Figure 6: Commodity-Village-Year Wise Prices, Kapanimbargi Village, Karnataka

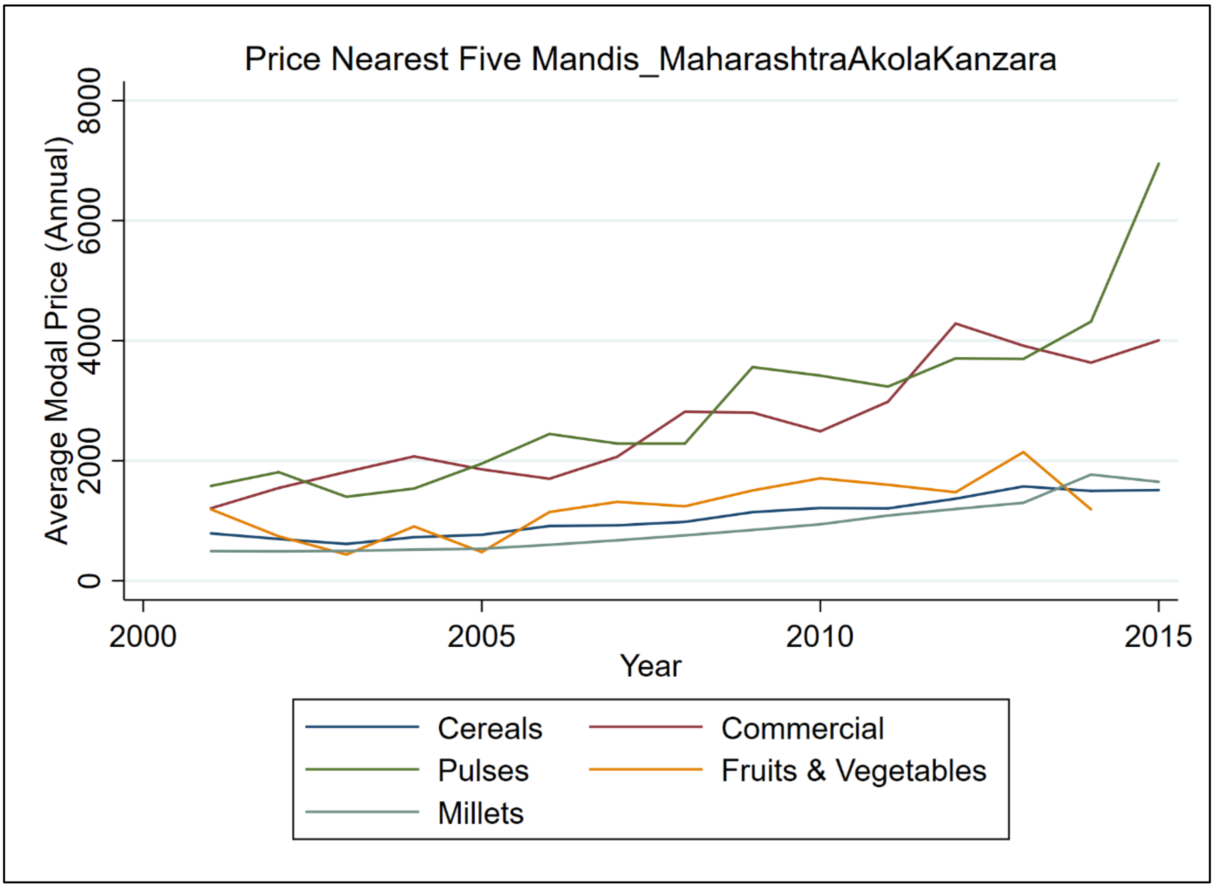


Figure 7: Commodity-Village-Year-Wise Prices, Akola Village, Maharashtra

Chapter 3: Input-conditioned crop yield density estimation for smallholder farms in India: An application of the BetaIV framework

Sumedha Shukla

(IIIT-Delhi)

sumedhas@iiitd.ac.in

Gaurav Arora

(IIIT-Delhi)

gaurav@iiitd.ac.in

Sandip Kumar Agarwal

(Indian Institute of Science, Education and Research (IISER), Bhopal)

sandipa@iiserb.ac.in

Working Paper Link: https://ageconsearch.umn.edu/record/322580?ln=en&v=pdf

DOI: https://doi.org/10.22004/AG.ECON.322580

Input-conditioned crop yield density estimation for smallholder farms in India: An application of the *BetaIV* framework

Abstract

Agricultural production is inherently risky, whereby output realized from a given choice of inputs is uncertain. Climate change is expected to exacerbate production uncertainty, leading to global food security challenges, especially for developing countries like India. Characterizing the shape of crop yield distributions provides an empirical measure of production risk conditional on input application which is fundamental to the understanding of farm-level risk management. However, most existing evidence on crop yield densities either relies on data aggregated at coarse spatial scales or at high spatial granularity focuses on relatively developed, resource-rich regions. In this study we estimate crop yield densities for wheat, rice, cotton and pigeonpea conditional on key inputs like nitrogen, phosphorus, and irrigation using farm-level, observational data for smallholder (i.e., ~2-5 acres) farms in the semi-arid regions of India. Our empirical setting characterized by limited access to institutionalized risk management strategies, and use of farm-level observational instead of agronomic trial data provides novel insights into farmers' use of inputs for managing production-risks. Our estimates provide a first piece of evidence on the yield risk-properties of crop inputs and present a realistic stepping stone for agricultural policy pertaining to production risk management in developing countries. As an illustration, we present a case-study examining the welfare effects of cotton-pigeonpea mixed-cropping (relative to monocropping) treating the share of land allocated to mixed cropping as a farmer's endogenous risk-management decision. While mixed cropping has been shown to enhance farm incomes, improve dietary diversity and lower production costs, the production risk effects mediating these outcomes remain ambiguous. Existing empirical work on cropping diversity and production risk has focused primarily on crop-specific (intra-species) varietal diversification. We contribute by examining the risk-reducing potential of a widely practiced inter-species mixed cropping system in semi-arid regions. Welfare implications are evaluated using certainty equivalent defined as expected income net of cost of risk exposure, where risk exposure is quantified using the variance and skewness of yield distributions following an adaptation of Di Falco and Chavas (2006; 2009). We find that mixed cropping reduces the variability of returns and increases skewness, leading to an overall reduction in downside risk exposure for the representative farmer.

JEL Codes: Q12, O13, C14, C36

Keywords: Yield Density Estimation; Risk Influencing Inputs; Beta Regression; Irrigation and Nitrogen Application; Instrumental Variables

Input-conditioned crop yield density estimation for smallholder farms in India: An application of the *BetaIV* framework

3.1 Introduction

"Until a science of yield probabilities can be developed, correct decisions in agriculture are virtually impossible."

Day (1965)

Agricultural production is inherently risky whereby the amount and quality of output resulting from a bundle of inputs is not known with certainty, i.e., the production function is stochastic (Moschini & Hennessy 2001). Climate change and the resultant increase in weather variability (World Bank 2010; Antón, et al.2012; Calzadilla et al. 2013; Stevanović et al. 2016; Altieri & Nicholls 2017) is expected to exacerbate crop yield variability (Schmidhuber & Tubiello 2007; Ray et al. 2015; Matiu et al. 2017; Arora et al. 2020) intensifying global food insecurity (Malthus 1798; Paddock & Paddock 1968; Evans 1998; Fischer et al. 2014). At greatest threat are the developing countries where crop yields remain far below potential and a relatively high share of the population derives their income from agriculture (Tilman et al. 2011; George 2014; Barrett & Bevis 2015; Vogel & Meyer 2018).

Characterizing the shape of crop yield distributions provides an empirical measure of production risk conditional on input application, and is therefore a critical intermediate step for analyzing farm management decisions (Day 1965; Hennessey 2009). This link between input use and production risk is especially important in developing countries, where access to institutionalized risk-mitigation tools (e.g., crop insurance) is limited and often skewed away from poorer, subsistence farmers who are most vulnerable to shocks (Byerlee et al. 2009; Eriksen and Brown 2011; FAO 2015; Fisher et al. 2019). Because marginal changes in input use shift the entire distribution of yields, these input-risk relationships underpin decisions ranging from crop insurance adoption to the management of environmental externalities. Consequently, estimating input-conditional yield densities is a foundational empirical exercise for agricultural policy that seeks to understand and influence decision-making under uncertainty. Reflecting this importance, yield-density estimation has a long and rich tradition in the economics literature (Day 1965; Just & Pope 1978; Antle & Goodger 1984; Buccola 1986; Gallagher 1986; Gallagher 1987; Nelson and Preckel 1989; Taylor 1990; Just and Weninger 1999; Moss and Shonkwiler 1993; Ramirez 1997; Ker & Coble 2003; Atwood et al. 2003; Du et al. 2012; Zheng et al. 2014; Ker and Tolhurst 2019). Please see Supplementary Information (SI), Section 1 (S1), Table 1 for details.

Existing evidence however is primarily derived from either agronomic experiments in relatively developed, resource-rich countries or from data at coarse spatial scales (e.g., counties in the United States) (see Table 1(S1)). A mere extension of this evidence for smallholder farmers in developing countries is likely to be unsuitable because the production and risk environment may be fundamentally different across the two settings. In fact, both the risk-attitudes as well as strategies available for risk mitigation can differ, given the high levels of exposure to weak institutions and extreme resource constraints in developing countries (Cervantes-Godoy et al. 2013; Mani at al. 2013; Schilbach et al. 2016; Waldman et al. 2020). At the same time, the synthesis of such evidence for developing countries can (often) be prohibitively costly – given the lack of consistent, disaggregated production data and consequent gaps in knowledge of the distribution of yields at even a coarse granularity - for a significant portion of the developing world (Leff et al. 2004; You et al. 2009; Van Ittersum et al. 2013; Ravisankar, et al. 2022).

In this study, we estimate input-conditioned crop yield densities using plot-level, repeated cross-sectional, *observational* data for rice, wheat, cotton and pigeonpea grown on smallholder (i.e., ~2-5 acres) farms in the semi-arid regions of India. The semi-arid regions characterized by a natural prevalence of abiotic stressors such as long dry seasons, high radiation, inconsistent rainfall and nutrient-poor soils are considered "hot-spots" for climate change (Ahlstorm et al. 2015). The concentration of a majority of rural-poor, food-insecure, and vulnerable populations in these areas presents a distinctive empirical setting that is particularly relevant for agricultural risk management policy across the developing world (Singh and Hazell 1993; Evans 1998; Krishnamurthy, et al. 2011; FAO 2016; Garcia-Franco, et al. 2018; Ravisankar et al. 2022). However, the estimation of this production structure is not straightforward and must carefully consider the implications of utilizing observational data and the choice of strategy for modelling yield densities.

Traditionally the estimation of agricultural yields and the analysis of the stochastic aspects of production response, particularly how inputs affect yield risk at fine spatial scales, have relied on agronomic data generated in controlled experimental settings. Such on-farm trials are invaluable for understanding phenological and physiological crop growth processes and for evaluating the performance of inputs at different stages of the growth cycle (Franzel and Coe 2002; Coe et al. 2017). However, evidence from these experiments cannot be directly applied to policy-relevant contexts involving economic agents (i.e., farmers) who differ widely in resource endowments, production constraints, and risk attitudes. This is because farmers' input decisions depend not only on biological crop response but also on how this response interacts with features of their production environment such as imperfect labor or credit markets, and with behavioural considerations. For example, a technology that raises yields but lengthens crop-standing time may be unattractive to a farmer facing acute labour shortages. Recent findings on persistent gaps between expected and realized yields following introduction of new technologies underscore this point (Kravchenko 2016; Laajaj et al. 2020).

Using observational data makes the isolation of causal effects of inputs difficult because *observed input application* decisions are the *optimal choices/decisions* made by a farmer given the resource constraints faced by them wherein a shift in the latter would likely change both input application as well as crop yields (Arrow 1958; Binswanger 1981; Buschena and Zilberman 1994; Mas-Colell, Whinston and Green 1995; Moschini and Henessey 2001). Consequently, the econometric estimation strategy must account for the endogeneity between a farmer's observed input application decisions and their agricultural outcomes, particularly the channels that operate through land quality, wealth, socioeconomic status, farming experience, and demographic characteristics (McArthur and McCord 2017; Cunningham 2021).

We address this endogeneity by combining instrumental-variable (IV) Tobit regressions for input choices with Beta regressions for yield estimation (Ferrari & Cribari-Neto 2004; Smithson & Verkuilen 2006), following the BetaIV framework introduced by Arora et al. (2021). This approach allows us to model yields conditional on observed, endogenous input decisions, specifically nitrogen, phosphorus, and irrigation application. The non-linearity in the IV framework accounts for the corner solutions in input application choices. In view of this non-linearity, a two-stage-residual-inclusion (2SRI) estimation method (Terza, et al. 2008) is employed rather than the standard two-stage predictor substitution (2SPS) method.

To our best knowledge, this paper provides the first systematic evidence on the risk properties of crop inputs for major crops in semi-arid regions of India. As an illustration, in the final section of this paper we utilize the estimated densities to examine the welfare effects of cotton-pigeonpea mixed-cropping[38] by incorporating the share of land allocated to mixed cropping as an additional input in the analysis (details will be discussed hereafter).

Beta regressions are appropriate since (a) yields are bounded between zero (crop failure) and a maximum attainable value (yield potential), and (b) yield distributions exhibit

[38] Mixed cropping refers to simultaneous cultivation of multiple crops on a single plot.

skewness. Theoretically, yields can be argued to be non-normal or skewed based on the resource-theoretic view grounded in the law of minimum technology articulated in Hennessey (2009). In this framework, tail attributes of resource availability distributions are key in determining yield skew. Specifically, with uniform resource availability, yield skewness tends to be positive. Making the production environment more "reliable" (e.g., through better infrastructure, irrigation) leads to extremely low yields becoming rarer while the upper yield bound remains essentially unchanged constrained by agronomy and technology. The resulting yield distribution therefore shifts toward more negative skewness.

Early evidence on yield skewness includes Day (1965) who used a parametric method of moments approach applied to the Pearson family of distributions to find non-normality and skewness in corn, cotton and oats yields, thus providing early support for beta yield densities. The need for skewness flexibility is further reflected in evidence of negative yield skewness for corn (Nelson and Preckel 1989; Moss and Shonkweiler 1993; Babcock and Hennessey 1996; Harri et al. 2009), positive yield skewness for cotton and sorghum (Goodwin and Ker 1998; Ramirez et al. 2003), positive or negative skewness for soybean and barley yields depending on production environment (Gallagher 1987; Ramirez 1997; Ramirez et al. 2003; Harri et al. 2009), and negligible skewness for wheat yields (Ramirez 1997; Ramirez et al. 2003). SI Table 1(S1) provides details on the geographical scope and spatial granularity of these studies.

Beyond allowing for unconditional flexibility in yield skewness, the stochastic specification of the input-yield relationship also requires careful theoretical grounding with regard to the role of inputs in yield risk. Just and Pope (1978) articulated eight "reasonable" postulates that a stochastic production function should satisfy in order to reflect *apriori* economic and technical considerations about how inputs influence both output levels and output risk (measured as output variance). These included: (a) positive expected output; (b) positive marginal expected product of inputs within the observed range; (c) diminishing marginal expected product of inputs within the observed range; (d) the possibility that inputs shift output variance without affecting expected output; (e) the possibility that inputs may increase, decrease, or leave unchanged the variance of output; (f) local risk invariance in the optimal choice of a risk-neutral farmer; (g) no sign restrictions on the variance of marginal products; and (h) the possibility of constant returns to scale.

Just and Pope's framework formalized the idea that inputs play a direct role in production risk management and demonstrated that conventional models based on multiplicative stochastic specification of yields i.e., those modelling only mean output as a function of inputs were insufficient[39]. Following this they proposed a more general *parametric* stochastic specification (Just and Pope 1979) that followed the postulates outlined above by modelling expected yields and yield variance as a function of inputs based on the log-linear distributional form. However, the Just and Pope definition of risk was limited to variance and hence could not capture how inputs influence the probability density of lower yields.

More fundamentally, Antle (1983) argued that conventional parametric, econometric production models (including Just and Pope (1979)) were inadequate representations of yield due to the arbitrary restriction imposed on the moments of output resulting from an “ad-hoc appending of additive or multiplicative random errors to deterministic production functions” (Antle 1983, p. 192) for representing production uncertainty. Specifically, he showed that multiplicative error models (e.g., Cobb-Douglas, CES, transcendental log) imply that elasticity of any higher order moment with respect to an input is directly proportional to the elasticity of *mean* production with respect to that input. While Just and Pope’s specification relaxed this marginally by modelling expected yield and variance separately, the proposed functional form still restricted variance and higher order moments in a similar manner by theoretically imposing that elasticity of any higher order moments (except variance) with respect to an input is directly proportional to elasticity of *variance* with respect to that input.

To circumvent the issue Antle (1983) proposed a *non-parametric* approach that expressed moments of the probability distribution of output (including higher order moments[40] such as skewness and kurtosis) as functions of inputs. This moment-based specification also incorporated the role of inputs in influencing downside risk, i.e., the probability of low yields which was empirically documented by Day (1965), Just and Pope (1979), Antle and Goodger (1984) and others. However, Nelson and Preckel (1989) argued that such a nonparametric approach is a strength only if a parametric probability distribution is not needed and “contingent on availability of information to specify a parametric probability distribution, the estimation of a parametric function is likely to be more efficient than the estimation of a non-parametric model”. Moreover, while non-parametric methods are impervious to specification

[39] Popular production specifications (e.g., Cobb-Douglas, constant elasticity of substitution (CES), translog and transcendental functions) specified in the multiplicative stochastic form as $y = f(x)e^{\varepsilon}$ where y is yield, x is a vector of inputs and ε is random disturbance such that $E(\varepsilon) = 0$, could satisfy postulates (a),(b) and (h). However, given (a) and (b), this specification would not admit a negative sign in (e). Moreover, (d) cannot not hold for any distribution of ε that satisfied (a) i.e., elasticity of any higher order moment with respect to an input is directly proportional to the elasticity of mean production with respect to that input.

[40] The count of moments required is based on the Stieltjes Condition (Antle 1983).

errors and therefore (relatively) more accurate and robust, they do not perform as well with multiple variables and smaller samples (Taylor 1984; Nelson and Preckel 1989).

Instead, Nelson and Preckel propose the use of the parametric *Beta* regression model for crop yield density estimation which captures the bounded nature of crop yields (between zero (crop failure) and maximum yield potential), as well as non-normality in yields by allowing for flexibility in yield skewness. Modelling both the mean yields as well as higher order moments as a function of input application also allows an examination of the role of inputs in managing downside yield risk (Nelson and Preckel 1989; Nelson 1990; Babcock and Hennessy 1996). With this grounding for the beta regressions, the next section (Section 3.2) details the empirical methodology followed by the mixed-cropping case study (Section 3.3). Section 3.4 presents the data summaries and model diagnostics while Section 3.5 presents the main results with discussion and Section 3.6 concludes.

3.2 Empirical Framework

3.2.1 Yield Model

Consider a random sample of farming households cropping one of the following crops $\mathrm{c} \in \{\mathrm{cotton,paddy,wheat,pigeonpea}\}$ on one or more plots indexed by i in either of kharif (rainy) or rabi (dry) season[41]. Farm-level yields (y) for a crop c are random and are modelled as a function of (a) physical inputs including seed type, fertilizer and electric irrigation application; (b) weather, i.e., temperature and rainfall; (c) farming ability as reflected in the age and education levels of the farmer; (d) land quality indicated by soil type, soil depth, and plot slope; and (e) plot-specific controls reflecting soil nitrogen levels - intercropping and crop-rotation indicators. In line with Nelson and Preckel (1989) and Babcock and Hennessy (1996); we use a reparametrized[42] version of the beta distribution (McDonald and Xu 1995; Ferrari & Cribari-Neto, 2004; Smithson and Verkuilen 2006) to model crop yield density for a crop as a function of the $k \times 1$ input vector $\vec{x}$, where k is the total number of inputs.

Obviously, all inputs that impact yields need not influence yield risk as well which can be captured by splitting the input vector accordingly. For now, we take $\vec{x}$ to be the set of both yield-influencing *and* risk characterizing inputs. Define the beta yield density for a crop as in

[41] Cotton and Pigeonpea may be intercropped i.e., grown on the same plot in the same season-year. Section 3.3 explores this further.

[42] Required for convenience in interpretation (Smithson and Verkuilen 2006)

equation (1) where $y \in [a,b]$, $\Gamma(.)$ is the gamma function, and $\omega(\vec{x}) > 0$; $\tau(\vec{x}) > 0$ are shape parameters such that a higher ω with a constant τ pulls the density towards the lower bound a - i.e., increases the rightward skew, whereas a higher τ keeping ω fixed, pulls the density towards the upper bound b - i.e., increases leftward skew.

$$f(y \mid \vec{x}) = \frac{\Gamma(\omega(\vec{x}) + \tau(\vec{x}))}{\Gamma(\omega(\vec{x}))\Gamma(\tau(\vec{x}))}(y-a)^{(\omega(\vec{x})-1)}(b-y)^{(\tau(\vec{x})-1)} \qquad (1)$$

The shape parameters are modelled as functions of input application since a change in input choices can influence the shape of the yield distribution. A linear transformation is required to rescale yields such that the they are bounded strictly between 0 and 1 (excluding the boundaries). In line with Smithson and Verkuilen (2006), without loss of generality, we define $\tilde{y} = \frac{1}{N}\left(\frac{y-a}{b-a}(N-1)+0.5\right)$ such that $\tilde{y} \in (0,1)$ and the density of $\tilde{y}$ is given by $f(\tilde{y} \mid \vec{x}) = \frac{\Gamma(\omega(\vec{x}) + \tau(\vec{x}))}{\Gamma(\omega(\vec{x}))\Gamma(\tau(\vec{x}))}(\tilde{y}-a)^{(\omega(\vec{x})-1)}(b-\tilde{y})^{(\tau(\vec{x})-1)}$. By definition the mean, variance and skewness (γ) of $\tilde{y}$ is given by:

$$E(\tilde{y} \mid \vec{x}) = \mu(\vec{x}) = \frac{\omega(\vec{x})}{\omega(\vec{x}) + \tau(\vec{x})} \qquad (2)$$

$$V(\tilde{y} \mid \vec{x}) = \sigma^2(\vec{x}) = \frac{\omega(\vec{x})\tau(\vec{x})}{(\omega(\vec{x}) + \tau(\vec{x}))^2(\omega(\vec{x}) + \tau(\vec{x}) + 1)} \qquad (3)$$

$$\gamma(\tilde{y} \mid \vec{x}) = \frac{2(\tau(\vec{x}) - \omega(\vec{x}))\sqrt{1 + \tau(\vec{x}) + \omega(\vec{x})}}{(2 + \omega(\vec{x}) + \tau(\vec{x}))\sqrt{\tau(\vec{x}) \times \omega(\vec{x})}} \qquad (4)$$

We can rewrite the variance as $V(\tilde{y} \mid \vec{x}) = \sigma^2(\vec{x}) = \left(\frac{\mu(\vec{x})(1-\mu(\vec{x}))}{(\omega(\vec{x}) + \tau(\vec{x}) + 1)}\right)$. Further, in order to lend a convenient interpretation to $\omega(\vec{x})$ and $\tau(\vec{x})$ as the location and dispersion parameters, we follow Smithson and Verkuilen (2006) and define $\phi(\vec{x}) = (\omega(\vec{x}) + \tau(\vec{x}))$ such that:

$$\omega(\vec{x}) = \mu(\vec{x})\phi(\vec{x}) \qquad (5)$$

$$\tau(\vec{x}) = \phi(\vec{x})(1 - \mu(\vec{x})) \qquad (6)$$

Here, $\mu(\vec{x})$ is termed as the "location" parameter and $\phi(\vec{x})$ is called the "dispersion/precision" parameter. Substituting the definition of $\phi(\vec{x})$, we get:

$$V(\tilde{y} \mid \vec{x}) = \sigma^2(\vec{x}) = \left(\frac{\mu(\vec{x})(1-\mu(\vec{x}))}{(\phi(\vec{x}) + 1)}\right) \qquad (7)$$

So, for a given value of $\mu(\vec{x})$ - the location parameter, a reduction in $\phi(\vec{x})$ - the dispersion parameter, implies that the variance in yield, i.e., $V(\tilde{y} \mid \vec{x})$ goes up. This is reflective of the classic feature of bounded response models whereby the dispersion in outcomes is partially dependent on the location parameter. At the same time, since $\mu(\vec{x})$ and $\phi(\vec{x})$ do not place any restriction on each other – they can be modelled separately under the generalized linear model (GLM) paradigm as detailed below.

3.2.2 Location Submodel

The location submodel must be specified such that the expected yields lie between 0 and 1. To ensure this, we use a logistic link function where $\vec{\beta}$ is a $k \times 1$ vector of coefficients.

$$\mu(\vec{x}) = \left(\frac{\exp(\vec{x}\,'\vec{\beta})}{1+\exp(\vec{x}\,'\vec{\beta})} \right) \qquad (8)$$

3.2.3 Dispersion Submodel

The dispersion submodel must be specified such that the variance in yields is strictly non-negative. To ensure this, we use a log link function where $\vec{w}$ and $\vec{\eta}$ are $k \times 1$ vectors of inputs and the associated coefficients respectively.

$$\ln(\phi(\vec{x})) = -\vec{w}\,'\vec{\eta} \text{ , i.e., } \phi(\vec{x}) = \exp(-\vec{w}\,'\vec{\eta}) \qquad (9)$$

Based on the above link functions, the log likelihood maximization problem for estimating the parameters in $\vec{\beta}$ and $\vec{\eta}$ can be written as in equation (10) - see Smithson and Verkuilen (2006) for details.

$$\begin{aligned}\max_{\vec{\beta},\vec{\eta}} \ln L(\vec{\beta},\vec{\eta}; y, X, W) = {} & \ln\Gamma\left(\frac{e^{X\vec{\beta}-W\vec{\eta}} + e^{-W\vec{\eta}}}{1+e^{X\vec{\beta}}} \right) - \ln\Gamma\left(\frac{e^{X\vec{\beta}-W\vec{\eta}}}{1+e^{X\vec{\beta}}} \right) - \ln\Gamma\left(\frac{e^{-W\vec{\eta}}}{1+e^{X\vec{\beta}}} \right) \\ & + \ln y \left(\frac{e^{X\vec{\beta}-W\vec{\eta}}}{1+e^{X\vec{\beta}}} - 1 \right) + \ln(1-y) \left(\frac{e^{-W\vec{\eta}}}{1+e^{X\vec{\beta}}} - 1 \right)\end{aligned} \qquad (10)$$

3.2.4 Auxiliary Input Regressions

An issue to be considered when using observational data for crop yield density estimation is that the isolation of causal effects in equations (8) and (9) is not straightforward since input application choices is endogenous to crop yields whereby higher yields (both potential and realized) can facilitate better input application and changes in input application can also influence both potential and realized yields. For instance, a farm-planner may allocate a higher share of resources (i.e., inputs like nitrogen, irrigation, labour hours, bullock hours etc.) to a plot with higher yield potential (due to (say) better soils), and vice versa. Yield potential incorporates not only the mean yield, but also the potential for lower/higher yields, i.e., the higher order moments of the yield distribution. Hence, both the location and dispersion model can suffer from endogeneity. To correct for this, we utilize the instrumental variable regression framework and, in the first stage, model inputs as a function of both input prices, output prices and plot-level cropping histories as instruments, and other farmer characteristics, land quality indicators and weather controls (detailed after equation (11)).

Now, it is important to emphasize that the *Beta* yield regression model specified in equations (8) to (10) is a non-linear model, i.e., the outcome variable is a non-linear (link) function of the linear combination of predictors in each of equations (8) and (9). Hence, a reproduction of the traditional two-stage-predictor-substitution (2SPS) technique (more commonly, the two-stage least square or 2SLS model) for estimation using instrumental variables is inconsistent (Terza et al. 2008). Instead, a consistent two-stage-residual-inclusion (2SRI) strategy is proposed for the case of non-linear parent regression models like the *Beta* regression model (Terza et al. 2008; Arora et al. 2021). In this framework, the first stage of regressions when using instrumental variables remains the same as in the standard 2SPS model – i.e., we model each endogenous input using instruments that are (a) are sufficiently correlated with the input application decision; and (b) uncorrelated with the error term in the parent regressions (i.e., the location and dispersion submodel(s) for yield).

In the second stage of 2SRI however, instead of substituting a predicted value of the endogenous variable from the first stage (as in 2SPS), the predicted residuals from the first stage model are included as additional regressors besides the endogenous variables (i.e., input application levels). Arora et al. (2021) adapt the 2SRI strategy for IV regressions in case of non-linear parent models to the beta regression model terming it the *BetaIV* framework. In this case, the regression for the location submodel can be written as:

$$\mu(\vec{x}) = \left(\frac{\exp(\vec{x}^{'e}\vec{\beta}_e + \vec{x}^{'o}\vec{\beta}_o + \vec{x}^{'u}\vec{\beta}_u)}{1+\exp(\vec{x}^{'e}\vec{\beta}_e + \vec{x}^{'o}\vec{\beta}_o + \vec{x}^{'u}\vec{\beta}_u)} \right) + v \quad ; \; E(v \mid \vec{x}^{e}, \vec{x}^{o}, \vec{x}^{u}) = 0 \qquad (11)$$

Here $\vec{x}^{'e}$ are the input application choices which are endogenous to yields through $\vec{x}^{'u}$ - an unobserved/latent component that affects mean yields and is correlated with $\vec{x}^{'e}$, and $\vec{x}^{'o}$ represents the observed, exogenous, control variables which influence average yields. To correct for endogeneity, we run auxiliary regressions modelling endogenous input choices as a function of (a) own price of the input; (b) price of other inputs which act as substitutes or complements to this input; (c) lagged price of output[43] which would determine resource allocation to the crop itself; (d) land quality ; (e) scale of operations; (f) farming ability as reflected in the age and education levels of the farmer; (g) weather and (h) plot-specific cropping history controls reflecting soil nitrogen levels and soil moisture content.

The relevance of the instruments is grounded in standard input demand theory. Input prices, specifically the district-year level prices of fertilizer, agricultural labour, and irrigation, directly affect the cost of applying each input and thereby shift farmers' optimal input choices independently of contemporaneous yield outcomes. Labour prices enter through the cost of cultivating and managing a more complex mixed-cropping system, making the extent of intercropping sensitive to prevailing wage rates. Higher wages increase the labour cost of managing two crops simultaneously and are therefore expected to reduce mixed-cropping intensity. We use lagged output prices because planting and input decisions are based on information available prior to production. For intercropping share specifically, both cotton and pigeonpea lagged output prices are included. The rationale is that the land allocation decision between sole cropping and mixed cropping is governed by relative expected returns across crops: higher cotton prices raise the opportunity cost of allocating land to the lower-value legume companion and may discourage mixed cropping, while higher pigeonpea prices improve its relative attractiveness and incentivize greater intercropping intensity.

The exclusion restriction requires that the instruments affect crop yields only through their influence on input application decisions and not through any independent direct channel. We therefore use district-level prices rather than self-reported farm-specific prices to mitigate the risk that individual farmers strategically misreport prices to justify their input decisions. The use of lagged rather than contemporaneous output prices is also important because current-period yield realizations cannot influence past prices. Conditional on the inclusion of weather variables, plot characteristics, farmer characteristics, cropping history, district fixed effects, and time trends, there are limited reasons to expect these lagged prices to affect current-period yields other than through their influence on production decisions. A potential concern is that expected output prices may affect yields directly if higher prices induce farmers to exert greater effort or invest in better-quality seeds, channels that may not be fully captured through the observed input variables included in the yield model. However, the rich set of

[43] For the case study in Section 3.3, we include as instruments, lagged prices of both crops (i.e., Cotton and Pigeonpea).

plot, farmer, and weather controls, together with district fixed effects, substantially reduces the scope for such omitted channels.

The plot-level data on endogenous input application decisions is contained in $Z = \{F_N, F_P, I\}$[44] where F_N, F_P and I are the levels of nitrogen, phosphorus and electric irrigation application per unit area on the plot. Given the extreme resource constraints in semi-arid regions, the optimal input application is often zero, i.e., no input is applied (reflecting a corner solution). A multivariate tobit model is used to account for this situation where a clustering of the observed input application choices is observed at zero. Here, let Z^* be the latent input application variable with an additive error that is normally distributed and homoscedastic. Thus,

$$Z^* = V\delta + \kappa \text{ where } \kappa \mid V \sim N(0, \sigma_\kappa^2) \tag{12}$$

The observed input application is defined as Z - a set of variables censored from below:

$$Z = \begin{cases} Z^* \text{if } Z^* > 0 \\ 0 \text{ if } Z^* \leq 0 \end{cases}$$

The piece-wise density function for each input (i.e., N and I) is then given as:

$$g(Z \mid V) = \begin{cases} 1 - \Phi(V\delta / \sigma_\kappa) \text{ if } Z = 0 \\ \dfrac{1}{\sqrt{2\pi\sigma_\kappa^2}} \exp\left(-\dfrac{1}{2\sigma_\kappa^2} (Z - V\delta)^2 \right) \text{ if } Z > 0 \end{cases} \tag{13}$$

Finally, consistent estimators for δ and σ_κ can be obtained by maximizing the log-likelihood function given by equation (14):

$$\ln L(\delta, \sigma_\kappa; Z, V) = 1_{Z=0} \ln(1 - \Phi(V\delta / \sigma_\kappa)) + 1_{Z>0} \ln\left(\frac{1}{\sqrt{2\pi\sigma_\kappa^2}} \exp\left(-\frac{1}{2\sigma_\kappa^2} (Z - V\delta)^2 \right) \right) \tag{14}$$

After estimating the input IV regressions, we recover the predicted residuals from (12) as $\hat{\kappa} = Z - V\hat{\delta}$ which are then used to replace $\vec{x}^{'u}$ in the second stage model, i.e., the residuals from the first stage instrumental variable regressions are included alongside the endogenous regressors in the second stage model. By replacing $\vec{x}^{'u}$ with its consistent estimate in the form of $\hat{\kappa}$, the 2SRI model removes the root cause of endogeneity in the yield model. This gives the *BetaIV* location submodel as:

---

[44] For the case study in Section 3.3, the input set also includes M defined as the share of land mix-cropped. Note that mixed-cropping is not a concern for wheat and paddy since less than 15% of wheat and paddy plots are mix-cropped.

$$\mu(\vec{x}) = \left(\frac{\exp(\vec{x}'^{e}\vec{\beta}_e + \vec{x}'^{o}\vec{\beta}_o + \hat{\kappa}\vec{\beta}_u)}{1+\exp(\vec{x}'^{e}\vec{\beta}_e + \vec{x}'^{o}\vec{\beta}_o + \hat{\kappa}\vec{\beta}_u)} \right) + \nu \quad (15)$$

Similarly, the *BetaIV* dispersion submodel is given by:

$$\phi(\vec{x}) = \exp[-(\vec{x}'^{e}\vec{\beta}_e + \vec{x}'^{o}\vec{\beta}_o + \hat{\kappa}\vec{\beta}_u)] + \nu^{\phi} \quad (16)$$

3.3 Case Study: Evaluating the role of mixed-cropping for managing production risks on small farms

3.3.1 Background

Multiple cropping, i.e., growing more than one crop species on the same land plot is an important ex-ante strategy to mitigate production or income losses specific to a particular crop (Just and Candler 1985; Dercon 1996; Di Falco and Zoupanido 2016; Khanal and Mishra 2017). Multiple cropping can be practiced through inter-seasonal crop rotation, which involves planting different crops sequentially on the same plot, or through intra-seasonal mixed cropping, where multiple crops are cultivated simultaneously on a plot. We focus on mixed cropping and attempt to understand its role in managing production risks. Agronomic experimental evidence suggests that mixed cropping can stabilize yields, increase plant biomass, reduce pest incidence and thereby improve overall cropping system resilience with the impacts being mediated by geo-physical conditions like temperature and soil type (Holling 1973; Lehman and Tilman 2000; Blaise et al. 2005; Tilman et al. 2005; Lin 2011; Raseduzzaman and Jensen 2017; Raza et al. 2025). Farmers' utilization of mixed cropping as a risk management strategy however, depends on their utility and risk exposure from mixed cropping (Chavas and Holt 1990; Di Falco and Perrings 2003; Lemken et al. 2017).

Since different crops respond differently to environmental and market risks, a typical risk-averse farmer would control their risk exposure by choosing crop-wise acreage and extent of mixed-cropping or diversification. Specifically, a farmer will choose a crop mix characterised by greater mixed cropping if mixed cropping is positively related to productivity and negatively (positively) correlated with production and income variability (skewness) (Just and Pope 1978; Di Falco and Perrings 2003; Di Falco and Perrings 2005; Di Falco and Chavas 2009). Higher order moments of production and income also matter since farmers in developing countries typically exhibit decreasing absolute risk aversion and downside risk aversion (Binswanger 1980; Binswanger and Sillers 1983; Antle 1987; Chavas and Holt 1996). Hence farmers have an incentive to develop cropping strategies that increase yield skewness, i.e., reduce downside yield risk (Di Falco et al. 2007; Di Falco and Chavas 2009;

Di Falco et al. 2010). If allocating land to different crop species is perceived as a risk-reducing strategy, then a risk-averse farmer would allocate a higher land share to mixed cropping compared to a risk-neutral farmer (Chavas and Holt 1990; Chambers and Quiggin 2000). Hence mixed cropping can substitute for formal insurance in hedging against the impact of production risk exposure.

Empirical evidence on the relation between multiple cropping, farm productivity and production risk exposure are mixed (Himmelstein et al. 2016). While crop diversification through multiple cropping is typically reported to increase agricultural production (DiFalco et al. 2007; Donfouet et al. 2017; Yin et al. 2024), conserve soil nutrients (Lin 2011; Raseduzzaman and Jensen 2017), and enhance food security and dietary diversity (Michler and Josephson 2017; Tesfaye and Tirivayi 2020), the production risk impacts mediating these effects are often ambiguous. Early explorations of the mixed cropping - production risk linkage posited crop diversity as an input in the production process and measured production risk using the mean and the variance of crop yields. The endogeneity in cropping diversity was typically addressed using an instrumental variable approach where commonly used instruments included gender, land tenure security, distance between farm and plot (Di Falco et al. 2010), access to credit (Singh and Rani 2025), distance between household and input supplier (Di Falco et al. 2007), and other measures of farmer's transaction cost of engaging in mixed cropping (Grilli et al. 2024).

Increments in crop-specific genetic diversity through mixed cropping have typically been found to improve mean yields and reduce yield variation (e.g., Di Falco and Perrings 2003). However, these effects are heterogeneous across levels of diversification and farmer-specific characteristics such as farm size, farmer wealth and access to irrigation (Maggio and Sitok 2020; Fujimoto and Suzuki 2025). For example, Di Falco et al. (2007) found that increasing durum wheat variety richness raised both farm productivity and yield variability, though the effect was significant only at high levels of genetic diversity in Ethiopia. Fafchamps (1992) observed that larger farms, having already secured food self-sufficiency, tended to diversify more into cash crops, while smaller farms prioritized staple crops to meet subsistence needs. Similarly, Smale et al. (1998) showed that wheat diversity improved mean yields and reduced yield variability in unirrigated areas, but was associated with lower expected yields in irrigated regions.

Assessing the influence of mixed cropping on farm-level risk exposure requires moving beyond a simple yield variance analysis to also account for downside risk - particularly through the assessment of yield skewness (Chavas and Holt 1996; Di Falco and Chavas 2006;

Di Falco and Chavas 2009; Di Falco et al. 2010). Studies adopting this perspective typically employ a flexible, moment-based approach to model the first three moments of the stochastic production function - mean, variance, and skewness - as functions of farm-level inputs, including mixed cropping diversity. Potential endogeneity in cropping diversity is typically addressed using instrumental variable techniques, as discussed above. This framework, pioneered by Di Falco and Chavas (2006), extends the mean-variance approach proposed by Just and Pope (1978) to incorporate higher-order moments, in line with Antle's (1983) theoretical foundation. By examining the impact of mixed cropping on skewness, this method captures the effects of diversification on downside risk exposure. The estimated moments are subsequently used to simulate the effects of crop varietal diversification on farmers' welfare within an expected utility framework, by quantifying expected net revenue and the associated cost of risk bearing (i.e., the risk premium) (Di Falco and Chavas 2006).

Incorporating the effects of downside risk exposure strengthens the case for mixed cropping as a risk-reducing strategy. For instance, Di Falco and Chavas (2006) report that crop genetic diversity can increase average wheat yields and reduce yield variance, but only under low pesticide use. However, the mitigation of downside risk exposure due to crop varietal diversification overweighs the risk-increase at high levels of pesticide use. Similarly, Di Falco and Chavas (2009) find that the impact of varietal diversification on barley yield skewness outweighs its positive effect on yield variability. A unifying feature of the empirical literature on cropping diversity and production risk is its predominant focus on *crop-specific (i.e., intra-species) varietal* diversification. In contrast, evidence on the risk-reducing potential of inter-species mixed cropping systems (e.g., cereal-legume, rice-cereal, rice-fodder) - which are widely practiced in developing regions (Waha et al. 2020) - remains limited. Studying risk exposure within a multi-crop framework that incorporates cross-commodity interactions is crucial. Chavas and Holt (1990), for example, demonstrate that cross-commodity risk reduction in a soybean-corn mixed cropping system implies that an increase in the support price for corn can lead to higher acreage and output for soybeans - an effect that would be overlooked in a single-crop analysis.

In this application we investigate the effects of a *cotton-pigeonpea* traditional mixed cropping system – prevalent in Indian SAT regions - on farm productivity and production risk. We employ beta regressions to model the mean, variance and other higher order moments of cotton and pigeonpea yields as a function of crop inputs - specifically nitrogen, phosphorus, electric irrigation and percentage of land allocated to mixed-cropping. Our empirical strategy accounts for the endogeneity between observed input choices and crop yields using a non-linear instrumental variable framework utilizing tobit regressions to account for corner solutions in input application choices (Arora et al. 2021). The welfare implications of mixed cropping are measured using certainty equivalent - the expected income minus the cost of risk exposure. To estimate risk exposure, we rely on the estimates for higher-order moments of the yield distributions, adapting the framework developed by Di Falco and Chavas (2006) to our multiple cropping context.

This analysis contributes to the existing literature in two dimensions. First, we use farm-level observational data rather than agronomic trial data to estimate input-conditioned yield densities which can provide insights for farmers' decision-making processes - particularly the use of inputs to manage production risk. Second, we focus on smallholder farms in a semi-arid, developing country context, where understanding the role of input use - including the extent of mixed cropping - in risk management is especially important given the limited availability of formal risk mitigation mechanisms. Our estimates offer the first empirical evidence on the yield risk properties diversification through mixed cropping.

3.3.2 Conceptual Framework (Based on DiFalco and Chavas (2006; 2009))

Consider a farm household h endowed with wealth w that chooses to cultivate one or more crops denoted by $j = \{\text{Cotton (1), Pigeonpea (2)}\}$ in a risky environment (e.g., due to weather conditions, pests) in the rainy season in a year on a land plot of size $\bar{l}$ acres. The crop yields (output per acre) for cotton and pigeonpea are denoted by y_1 and y_2 respectively. The farmer can influence the distribution of crop yields by choosing physical inputs (e.g., nitrogen) denoted by $\vec{x}$ and the share of land cultivated for each crop (i.e., l_1 and l_2). Note that inputs are not crop-specific since all cultivation is specified on the same plot of land, i.e., when the farmer grows both cotton and pigeonpea, cropping is mixed. Crop production is specified as a function of inputs, land allocation, and a random component u that captures uncertainty in environmental conditions like weather and pest incidence and affects yields for both crops. Specifically we have $y_1 = q_1(\vec{x}, l_1) + g_1(u)$ and $y_2 = q_2(\vec{x}, l_2) + g_2(u)$ where $g_j(u)$ captures how weather variability affects crop j. $g_1(u)$ can be different from $g_2(u)$ since crop-responses to the same weather conditions can be different – e.g., high rainfall can increase cotton yields while harming pigeonpea cultivation.

The cost of production is given by $C(\vec{x}, l_1, l_2) = \vec{p}_x \cdot \vec{x} + s_1 l_1 + s_2 l_2$ where s_j represents crop-specific land preparation cost per unit of land and $s_1 \neq s_2$ - e.g., cotton requires deeper ploughing compared to pigeonpea. The farmer's income is $m = p_1 y_1 + p_2 y_2 - C(\vec{x}, l_1, l_2)$ where p_1 and p_2 are the market prices for y_1 and y_2 respectively.

The farmer is non-satiated and risk-averse with respect to income and chooses inputs and land allocation to maximize the expected utility from their total earnings, i.e., $E(u(w+m))$ where $u(.)$ is an increasing and concave Bernoulli utility function. The farmer's problem in each year can therefore be outlined as follows:

$$\begin{aligned}
&\max_{\vec{x}, l_1, l_2} E(u(w+m)) \\
&\text{s.t.} \\
&m = p_1 y_1 + p_2 y_2 - C(\vec{x}, l_1, l_2) \\
&l_1 + l_2 = \overline{l}
\end{aligned} \tag{17}$$

Maximizing the expected utility is equivalent to maximizing the certainty equivalent (DiFalco and Chavas 2006; DiFalco and Chavas 2009) such that we can rewrite (17) as (18) where CE is the certainty equivalent and R is the risk premium (i.e., implicit cost of bearing risk or risk exposure) associated with land allocation and input choices.

$$\begin{aligned}
&\max_{\vec{x}, l_1, l_2} CE = E(w+m) - R \\
&\text{s.t.} \\
&m = p_1 y_1 + p_2 y_2 - C(\vec{x}, l_1, l_2) \\
&l_1 + l_2 = \overline{l}
\end{aligned} \tag{18}$$

From (18), we observe that as a risk-averse decision maker, the household has an incentive to reduce R which depends on the *distribution* of earnings ($\pi = w + m$) and the farmer's degree of risk aversion. We measure risk aversion using the Arrow-Pratt measure for absolute risk aversion at a given π, i.e., $r_{2a} \equiv -\left(\partial^2 u / \partial \pi^2 \middle/ \partial u / \partial \pi \right)$. Moreover risk aversion can vary by the level of wealth and income. For example smallholder farmers are likely to display decreasing absolute risk aversion (DARA) as income increases (Pratt 1964; Binswanger 1980; Binswanger and Sillers 1983; Chavas and Holt 1996). We capture these effects and their influence on optimal input and land allocation decisions by examining higher order moments (e.g., skewness) of the earnings distribution. Examining skewness enables us to account for phenomena like DARA because $\frac{\partial r_{2a}}{\partial \pi} = -\left(\partial^3 u / \partial \pi_i^3 \middle/ \partial u / \partial \pi \right) + r_{2a}^2 \equiv r_{3a} + r_{2a}^2 < 0$.

We articulate the role of higher-order yield moments in the optimal input use and land allocation decisions by first rewriting the certainty equivalent as:

$$CE = w + p_1\mu_1(\vec{x},l_1) + p_2\mu_2(\vec{x},l_2) - C(\vec{x},l_1,l_2) - R(r_{2a}, r_{3a}, \sigma_\pi^2, \gamma_\pi) \tag{19}$$

Here μ_j denotes the expected yield of crop j, and σ_π^2 and γ_π represent the variance and skewness[45] of farm income given as $\sigma_\pi^2 = p_1^2\sigma_{y_1}^2 + p_2^2\sigma_{y_2}^2 + 2p_1p_2\operatorname{cov}(y_1, y_2)$ and

$\gamma_\pi = \dfrac{p_1^3\gamma_{y_1}\sigma_{y_1}^3 + p_2^3\gamma_{y_2}\sigma_{y_2}^3}{\left(p_1\sigma_{y_1}^2 + p_2\sigma_{y_2}^2 + 2p_1p_2\operatorname{cov}(y_1,y_2)\right)^{3/2}}$ respectively.

R is a function of the variance and skewness of farm income and we approximate R from the problem outlined in (19) as:

$$R \approx \frac{1}{2}r_{2a}\sigma_\pi^2 + \frac{1}{6}r_{3a}\gamma_\pi \tag{20}$$

To see this, consider that the expected utility from π is equivalent to the utility from the certainty equivalent of π net of the cost of risk, i.e.,

$$\begin{aligned} &u(w + p_1E(y_1) + p_2E(y_2) - C(\vec{x},l_1,l_2) - R) = \\ &E(u(w + p_1y_1 + p_2y_2 - C(\vec{x},l_1,l_2))) \equiv E(u(\pi)) \end{aligned} \tag{21}$$

Writing a Taylor series expansion of $u(\pi)$ around $E(\pi)$ and further taking an expectation gives us the following expression for the RHS in (21).

$$E(u(\pi)) \approx u(E(\pi)) + \frac{1}{2}u''(E(\pi))\sigma_\pi^2 + \frac{1}{6}u'''(E(\pi))\gamma_\pi \tag{22}$$

Similarly, expanding the LHS for a small value of R, we get:

$$u(E(\pi) - R) \approx u(E(\pi)) - Ru'(E(\pi)) \tag{23}$$

Finally, equating (22) and (23) gives us the approximation for R as a function of $r_{2a}, r_{3a}, \sigma_\pi^2$ and γ_π. Based on this, we can write the optimal land allocation for each crop (j) as:

[45] $\sigma_\pi^2 = p_1^2\sigma_{y_1}^2 + p_2^2\sigma_{y_2}^2 + 2p_1p_2\operatorname{cov}(y_1, y_2)$ and $\gamma_\pi = \dfrac{p_1^3\gamma_{y_1}\sigma_{y_1}^3 + p_2^3\gamma_{y_2}\sigma_{y_2}^3}{\left(p_1\sigma_{y_1}^2 + p_2\sigma_{y_2}^2 + 2p_1p_2\operatorname{cov}(y_1,y_2)\right)^{3/2}}$

$$p_j \frac{\partial \mu_j}{\partial l_j} = \underbrace{\underset{>0}{\frac{\partial C}{\partial l_j}}}_{\text{Direct Cost}} + \underbrace{\underset{>0}{\frac{\partial R}{\partial \sigma_\pi^2}} \underset{?}{\frac{\partial \sigma_\pi^2}{\partial l_j}} + \underset{<0}{\frac{\partial R}{\partial \gamma_\pi}} \underset{?}{\frac{\partial \gamma_\pi}{\partial l_j}}}_{\text{Cost of Risk Exposure}} \tag{24}$$

At the optimal l_j^*, the expected marginal product of the land cultivated with crop j is equal to the cost of allocating additional land to j. This includes a direct cost of land preparation – for example expenses incurred to build ridges for drainage in Pigeonpea cultivation, and the implicit cost of additional risk exposure associated with cultivating more j (pigeonpea).

While $\frac{\partial C}{\partial l_j} > 0$, $\frac{\partial R}{\partial \sigma_\pi^2} > 0$, and $\frac{\partial R}{\partial \gamma_\pi} < 0$, the sign of $\frac{\partial \sigma_\pi^2}{\partial l_j}$ depends on the effect of an additional unit of land allocation to j on the variation in yields of cotton and pigeonpea, the extent to which the yields of these crops covaries and the market price of each crop. Similarly, the sign of $\frac{\partial \gamma_\pi}{\partial l_j}$ depends on the effect of an additional unit of land allocation to j on the skewness and variation in cotton and Pigeonpea yields variation, the extent covariation, and the prices of each crop.

In this manner we can assess the risk-properties of mixed cropping for a risk-averse farmer who is also an expected utility maximizer with equation (20) providing a break-down of the farmer's risk exposure into a skewness and variance component. We will now transition from an expected utility framework to a prospect theory framework which can provide a more nuanced understanding of how farmers, particularly those prone to loss aversion, assess mixed cropping as a risk management strategy. While expected utility theory provides a foundational understanding of rational decision-making under uncertainty by evaluating choices based on the average utility of outcomes, it does not account for empirical phenomenon such as a disproportionate aversion to losses compared to equivalent gains.

3.4 Data Description

We utilize a plot-level repeated cross-sectional dataset made available by the International Crops Research Institute for the Semi-Arid Tropics under the Village Dynamics Studies in South Asia (VDSA) program. Our sample comprises of 688 paddy-cultivating households operating 3185 plots in Kharif, 124 paddy-cultivating households operating 534 plots in Rabi, 593 wheat-cultivating households operating on 2237 plots, 406 cotton-growing households operating on 1723 plots and 520 pigeonpea growing households operating on 1520 plots, with majority smallholder farmers (median plot size ~2 acres). Among the cotton cultivators, 172 households operating 1034 plots engage in mixed cropping of cotton and pigeonpea. Table 3.1 presents a list of the variables and their descriptions and Tables 3.2 – 3.4 present the crop-wise summary statistics. The input variables are plot-wise aggregate quantities for the whole growing-season divided by the area of the plot under that crop.

3.5 Results and Discussion

3.5.1 Input Regressions

Tables 3.5-3.8 present the input regression results. Fertilizer prices are generally negatively associated with fertilizer utilization and mixed-cropping intensity while results are mixed with irrigation application. Input response to an increase in labour prices differs significantly across crops. In kharif paddy, irrigation can substitute for labour by suppressing weeds and stabilizing crop growth under flooded conditions, while fertilizer application remains labour-complementary due to timing and management requirements. In contrast, for other crops, fertilizer substitutes for labour while irrigation is typically unaffected. Moreover as expected, increase in labour prices is associated with a reduction in mixed cropping. An increase in irrigation charges is typically associated with a reduction in the use of all other inputs including land share allocated to mixed cropping suggesting that irrigation acts as a complement to all farm operations in semi-arid regions. Higher water costs therefore raise the effective cost of cultivation more broadly, leading farmers to scale back overall production intensity rather than substitute away from irrigation alone. This finding is consistent with irrigation relaxing multiple constraints simultaneously in semi-arid regions. An exception to this result is kharif paddy where increase in irrigation charges is associated with an increase in irrigation which likely reflects the limited substitutability of irrigation for a water-sensitive crop like paddy even in the rainy season.

Land owners have on-average higher fertilizer and irrigation application across crops. Older farmers tend to apply lower amounts of fertilizer while there is no significant association between farmer’s age and irrigation. Operating on problem soils is associated with lower input use reflecting lower scale of operations and higher mixed cropping consistent with the latter’s role in reducing production risks through diversification. Deeper soils are associated with higher nitrogen application and lower irrigation hours per acre while the association with phosphorus application is insignificant. Education is positively associated with input use but the result is not consistent across crops. Forward caste households have higher irrigation application relative to scheduled caste and scheduled tribe households, a finding consistent across all crops. Finally, output prices are generally associated negatively with irrigation application across crops – this finding needs further investigation.

Cultivation of nitrogen-fixing crops in the previous season is associated with higher extent of mixed cropping, lower fertilizer use and higher irrigation application. On the contrary, cultivation of water intensive crops in the previous season is associated with lower extent of mixed cropping in the current season, and higher fertilizer and electric irrigation utilization.

3.5.2 Yield Regressions

Tables 3.9-3.13 summarize the yield regression results. Paddy yields are negatively skewed across the rainy ($\omega(x)$=2.99 and $\hat{\tau}(\tilde{x})$=1.76) and dry ($\omega(x)$=4.62 and $\hat{\tau}(\tilde{x})$=2.95) seasons. Interestingly, although the average yield is higher in the dry season, the probability of lower yields is lower in the rainy season, i.e., cultivation is less-risky. A marginal change in nitrogen significantly increases the mean yield at a decreasing rate and also reduces the probability of lower yields. A marginal change in phosphorus significantly increases the mean yield but also increases the probability of lower yields. Irrigation is revealed as a risk-reducing input across seasons albeit more pronounced in the rainy season. However, irrigation does not change mean yields significantly. Cotton (mono-cropped) yields are positively skewed ($\omega(x)$=3.02 and $\hat{\tau}(\tilde{x})$= 17.11) and a marginal change in nitrogen significantly increases the mean yield but increases the probability of lower yields. In contrast a marginal increase in phosphorus application increased mean yields but the impact on yield variability was insignificant. Irrigation emerges as a risk-reducing input even though the impact on mean yield was insignificant.

Pigeonpea (mono-cropped) yields are positively skewed ($\omega(x)$=1.92 and $\hat{\tau}(\tilde{x})$= 10.99) and a marginal change in nitrogen input significantly reduces yield variability even though the impact on mean yield was insignificant. In contrast, a marginal increase in phosphorus application increases mean yields and reduces yield variability significantly. Irrigation does not have a significant association with yields. Wheat yields display negligible skew ($\omega(x)$ =6.35 and $\hat{\tau}(\tilde{x})$= 6.88) even though the tails are thin and a marginal change in nitrogen has a positive impact on mean yields while the impact on yield variability is negative. In contrast phosphorus application reduces mean yields but increases the odds of higher yields i.e., yield variability. Irrigation application has the expected effect i.e.; it increases mean yields and reduces the odds of lower yields. Mixed cropping increases mean yields for both cotton and pigoenpea. The impact on yield variability is insignificant for cotton but highly significant and negative for Pigeonpea suggesting that mixed cropping is a risk-reducing input for production. Across all crop categories, we find that HYV seeds are associated higher yields on-average but the variability in yields is also higher. Plot level bunding is associated with higher yields on-average. Deeper soils are associated with higher yields on-average for cotton and paddy in the rainy season.

3.5.3 Simulations

To analyse the economic and welfare implications of our econometric results, we present two simulation exercises. First, in order to further characterize the impact of nitrogen, phosphorus and irrigation on crop yields, we first plot the crop yield density for an average farmer in our sample, i.e., at average nitrogen, phosphorus and irrigation levels and then assess the impact of a marginal shift in each input on crop yields (on-average, i.e., for the representative farmer in our sample). We find $\hat{\omega}(\vec{x}) = 2.99$ and $\hat{\tau}(\vec{x}) = 1.76$ for the representative farmer cultivating paddy in kharif season implying that paddy yields are negatively skewed, i.e., the density of high yields is relatively higher (see Figure 3.1, Panel A). Figure 3.1, Panel B compares the change in paddy yield density upon a marginal change in nitrogen ($\hat{\omega}(\vec{x}) = 3.64$ and $\hat{\tau}(\vec{x}) = 1.87$). Clearly, not only the mean yields are higher upon additional nitrogen application, but we find higher probability of low yields in case of lower nitrogen application. Similarly Figure 3.1, Panel C compares the change in paddy yield density upon a marginal change in irrigation ($\hat{\omega}(\vec{x}) = 2.87$ and $\hat{\tau}(\vec{x}) = 1.68$). While the impact of irrigation on expected yields is negligible (expected given rainy season), it still lowers the risk of low yields. Finally Figure 3.1, Panel D compares the change in paddy yield density upon a marginal change in phosphorus ($\hat{\omega}(\vec{x}) = 2.36$ and $\hat{\tau}(\vec{x}) = 1.44$). While the impact of phosphorus on mean yields is negative, its effect on yield variance is also negative, i.e., the density of both lower and higher yields is higher with phosphorus thus, the net welfare effect is ambiguous.

We find $\hat{\omega}(\vec{x}) = 4.62$ and $\hat{\tau}(\vec{x}) = 2.95$ for the representative farmer cultivating paddy in rabi season implying that paddy yields are negatively skewed, i.e., the density of high yields is relatively higher (see Figure 3.2, Panel A). Figure 3.2, Panel B compares the change in paddy yield density upon a marginal change in nitrogen ($\hat{\omega}(\vec{x}) = 3.02$ and $\hat{\tau}(\vec{x}) = 2.67$). Clearly, not only the mean yields lower upon additional nitrogen application, but we find higher probability of low yields in case of lower nitrogen application. This results however is not likely to be a causal effect since the instrument tests fail for this case (explained in the next section). Figure 3.2, Panel C compares the change in paddy yield density upon a marginal change in irrigation ($\hat{\omega}(\vec{x}) = 4.92$ and $\hat{\tau}(\vec{x}) = 3.01$). While the impact of irrigation on expected yields is negligible, it still lowers the risk of low yields albeit only marginally. Finally Figure 3.2, Panel D compares the change in paddy yield density upon a marginal change in phosphorus ($\hat{\omega}(\vec{x}) = 2.41$ and $\hat{\tau}(\vec{x}) = 1.1$). Clearly the impact of phosphorus on mean yields is positive while its effect on yield variance is also negative, i.e., the density of higher yields is higher with phosphorus application.

We find $\hat{\omega}(\vec{x}) = 6.35$ and $\hat{\tau}(\vec{x}) = 6.88$ for the representative farmer cultivating wheat in rabi season implying that wheat yields are negligibly skewed (see Figure 3.3, Panel A). Figure 3.3, Panel B compares the change in wheat yield density upon a marginal change in nitrogen ($\hat{\omega}(\vec{x}) = 7.9$ and $\hat{\tau}(\vec{x}) = 6.4$). Clearly, not only the mean yields are higher upon additional nitrogen application, but we find lower probability of low yields in case of lower nitrogen application. Figure 3.3, Panel C compares the change in wheat yield density upon a marginal change in irrigation ($\hat{\omega}(\vec{x}) = 5.1$ and $\hat{\tau}(\vec{x}) = 4.7$). We find that impact of irrigation on expected yields is positive and it lowers the risk of low yields. Finally Figure 3.3, Panel D compares the change in wheat yield density upon a marginal change in phosphorus ($\hat{\omega}(\vec{x}) = 8.1$ and $\hat{\tau}(\vec{x}) = 16.4$). While the impact of phosphorus on mean yields is positive, its effect on yield variance is also positive thus, the net welfare effect is ambiguous.

We find $\hat{\omega}(\vec{x}) = 3.02$ and $\hat{\tau}(\vec{x}) = 17.11$ for the representative farmer cultivating cotton in kharif season implying that cotton yields are positively skewed, i.e., the density of low yields is relatively higher (see Figure 3.4, Panel A). Figure 3.4, Panel B compares the change in cotton yield density upon a marginal change in nitrogen ($\hat{\omega}(\vec{x}) = 4.13$ and $\hat{\tau}(\vec{x}) = 27.6$). Clearly, while the mean yields are higher upon additional nitrogen application, the probability of relatively lower yields is also higher due to nitrogen application. Similarly Figure 3.4, Panel C compares the change in cotton yield density upon a marginal change in irrigation ($\hat{\omega}(\vec{x}) = 3.2$ and $\hat{\tau}(\vec{x}) = 18.2$) and we find that the impact of irrigation on expected yields and yield skewness is negligible. Finally Figure 3.4, Panel D compares the change in cotton yield density upon a marginal change in phosphorus ($\hat{\omega}(\vec{x}) = 1.8$ and $\hat{\tau}(\vec{x}) = 4.5$). While the impact of phosphorus on mean yields is negative, its effect on yield variance is also negative, i.e., the density of both lower is lower with phosphorus.

We find $\hat{\omega}(\vec{x}) = 1.92$ and $\hat{\tau}(\vec{x}) = 10.99$ for the representative farmer cultivating pigeonpea in kharif season implying that pigeonpea yields are positively skewed, i.e., the density of low yields is relatively higher (see Figure 3.5, Panel A). Figure 3.5, Panel B compares the change in pigeonpea yield density upon a marginal change in nitrogen ($\hat{\omega}(\vec{x}) = 2.51$ and $\hat{\tau}(\vec{x}) =$ 12.42). Clearly, not only the mean yields are higher upon additional nitrogen application, but we find lower probability of low yields in case of nitrogen application. Similarly Figure 3.5, Panel C compares the change in pigeonpea yield density upon a marginal change in irrigation ($\hat{\omega}(\vec{x}) = 2.1$ and $\hat{\tau}(\vec{x}) = 11.4$). While the impact of irrigation on expected yields is negligible (expected given rainy season), it still lowers the risk of low yields albeit only marginally. Finally Figure 3.5, Panel D compares the change in pigeonpea yield density upon a marginal change in phosphorus ($\hat{\omega}(\vec{x}) = 2.38$ and $\hat{\tau}(\vec{x}) = 11.31$). While the impact of phosphorus on mean yields is positive, its effect on yield variance is also positive, i.e., the density of lower yields is lower with phosphorus.

Next, we simulate the effect of increasing mixed cropping share for an average (representative) cotton and pigeonpea farmer in our sample and plot their yield density. Figures 3.6 and 3.7 display these results. Clearly, for the average farmer growing cotton and piegeopea, mixed cropping emerges as a mean-increasing, risk-reducing input. However the impact on pigeopea yield variability is lower at higher level of mixed cropping. We further measure the influence of mixed cropping on risk exposure by evaluating the risk-premium outlined in section 3.3.2. We assume that the representative farmer in our sample is moderately risk averse with a relative risk aversion coefficient of 2.5. Figure 3.8 displays the results from this simulation. We see that cropping diversity through mixed cropping reduces the variability of returns and increases skewness, leading to an overall reduction in downside risk exposure despite its slight negative effect on mean returns.

3.6 Model Selection and Diagnostics:

The following section outlines key challenges in implementing the *BetaIV* framework and explains the rationale behind specific modelling choices. First, we use district-level input and output prices instead of farmer-reported values to mitigate potential biases arising from strategic misreporting of prices to justify input application decisions. Second, we include indicators for whether water-intensive or nitrogen-fixing crops were cultivated in the previous season on the same plot, given that current soil nutrient demands are often influenced by past cropping patterns (Du et al. 2012). Third, we test for weak instruments and over-identifying restrictions. Following Stock and Yogo (2005), we use a threshold F-statistic of 20.27 in the first-stage regression to detect weak instruments; F-statistic below this threshold suggests instrument weakness. The corresponding first-stage F-statistics are reported in Appendix Table A2. To assess the validity of our instruments, we conduct the Sargan test for over-identifying restrictions, following the procedure outlined in Wooldridge (2002, p. 123), with results reported in Table A3. All standard errors are clustered at the village-level.

Fourth since cotton and pigeonpea can be intercropped on the same plot, the second-stage beta regressions need to be estimated within a seemingly unrelated regression (SUR) framework to account for potential correlation in unobserved shocks across the two crops. The joint estimation of two beta regressions, however, requires fixing either the location or dispersion parameter for identification[46]. Accordingly, we implement *a joint means model* to obtain the empirical cross-crop covariance and report the joint mean-regression results as a robustness check. Nonetheless, our primary specification estimates the beta regressions for cotton and pigeonpea separately, which allows us to specify both the location and dispersion submodels as functions of input use - particularly important for assessing the effect of inputs on yield variability and skewness.

Fifth, in some cases, we face the issue that higher input levels are associated with higher input price (in particular, irrigation) or higher output prices are associated with lower input levels even after accounting for the high correlation between input and output prices and dropping them one-at-a-time for estimating the instrumental variable regressions. The issue persists regardless of the use of farmer reported prices or prices reported by Input Surveys (multiple rounds MoAFW). To address this I am planning on incorporating mandi-level output prices to see if the issue persists. Finally, in case of Paddy in the Rabi season, we achieve a poor model fit for the yield model and the instruments turn out to be weak in explaining phosphorus application while the exclusion restrictions are also not satisfied. This needs further investigation. To improve instrument quality, I aim to bring district-level input quantities instead of prices since regulated supply implies they can be an instrument for input application which will not affect plot-level yields except through plot-level application amount.

[46] See GSEM in Stata or BRMS in R.

3.7 Conclusions and Future Work

This study presents the first estimates of yield risk-properties of crop inputs for major crops in semi-arid Indian villages using farm-level data on crop yields and input choices that underpin all agricultural policy. As an illustration, I utilize these estimates to assesses the role of mixed cropping - specifically cotton intercropped with legumes—as a farm-level risk management strategy. Our findings show that greater adoption of mixed cropping significantly reduces the risk of crop failure, particularly by lowering yield variability, with the effect on variance outweighing that on skewness. These results suggest that crop diversity not only provides important environmental benefits (such as reduced input use) but also enables farmers to buffer against the adverse effects of harsh agroecological conditions. Thus, promoting in situ conservation of cropping diversity can enhance agricultural resilience and food security in semi-arid regions.

A limitation of our analysis is that the yield regressions do not fully account for the potential correlation between cotton and pigeonpea yields, especially in their higher-order moments. In addition, concerns remain regarding the validity of the instruments used. Future work will aim to address these limitations by refining the instrumental variable framework and conducting additional robustness checks. We also plan to explore heterogeneity in risk exposure effects across farmer groups differentiated by landholding size, tenure status, and soil quality.

3.8 Figures

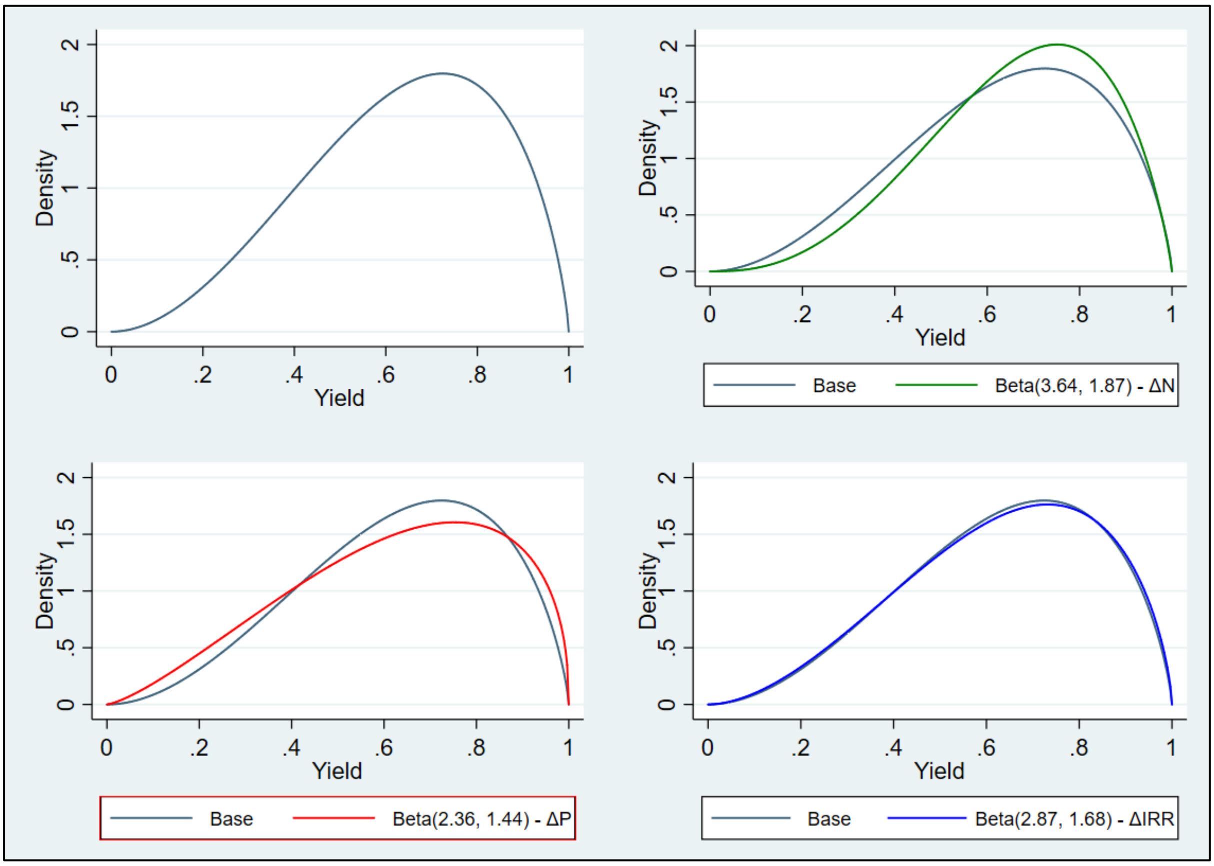


Figure 3.1: Input-conditioned Paddy (Kharif) yields estimated from the BetaIV regression framework. Top-left to right: **Panel A:** Paddy yields for an average farmer in the sample; **Panel B:** Change in yield density upon a marginal shift in nitrogen; **Panel C:** Change in yield density upon a marginal shift in irrigation; **Panel D:** Change in yield density upon a marginal shift in phosphorus.

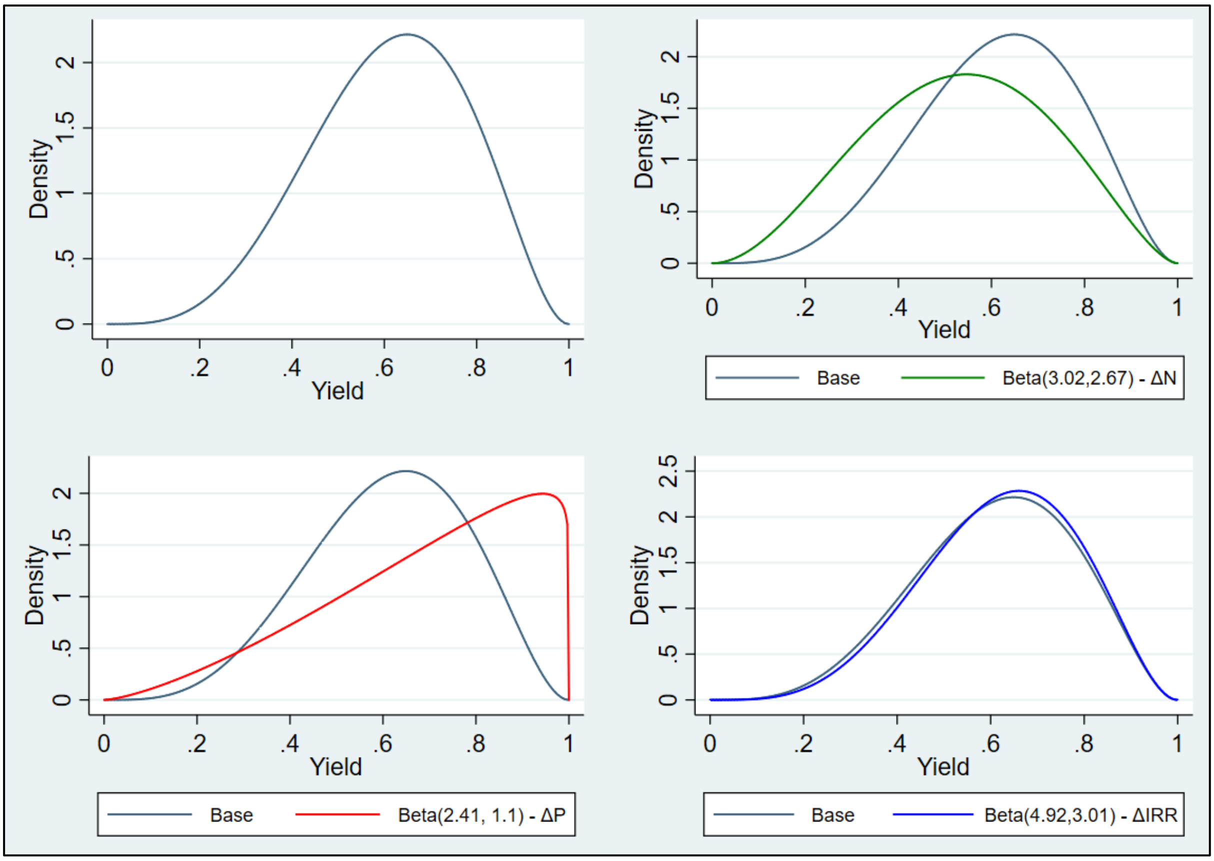


Figure 3.2: Input-conditioned Paddy (Rabi) yields estimated from the BetaIV regression framework. Top-left to right: **Panel A:** Paddy yields for an average farmer in the sample; **Panel B:** Change in yield density upon a marginal shift in nitrogen; **Panel C:** Change in yield density upon a marginal shift in irrigation; **Panel D:** Change in yield density upon a marginal shift in phosphorus.

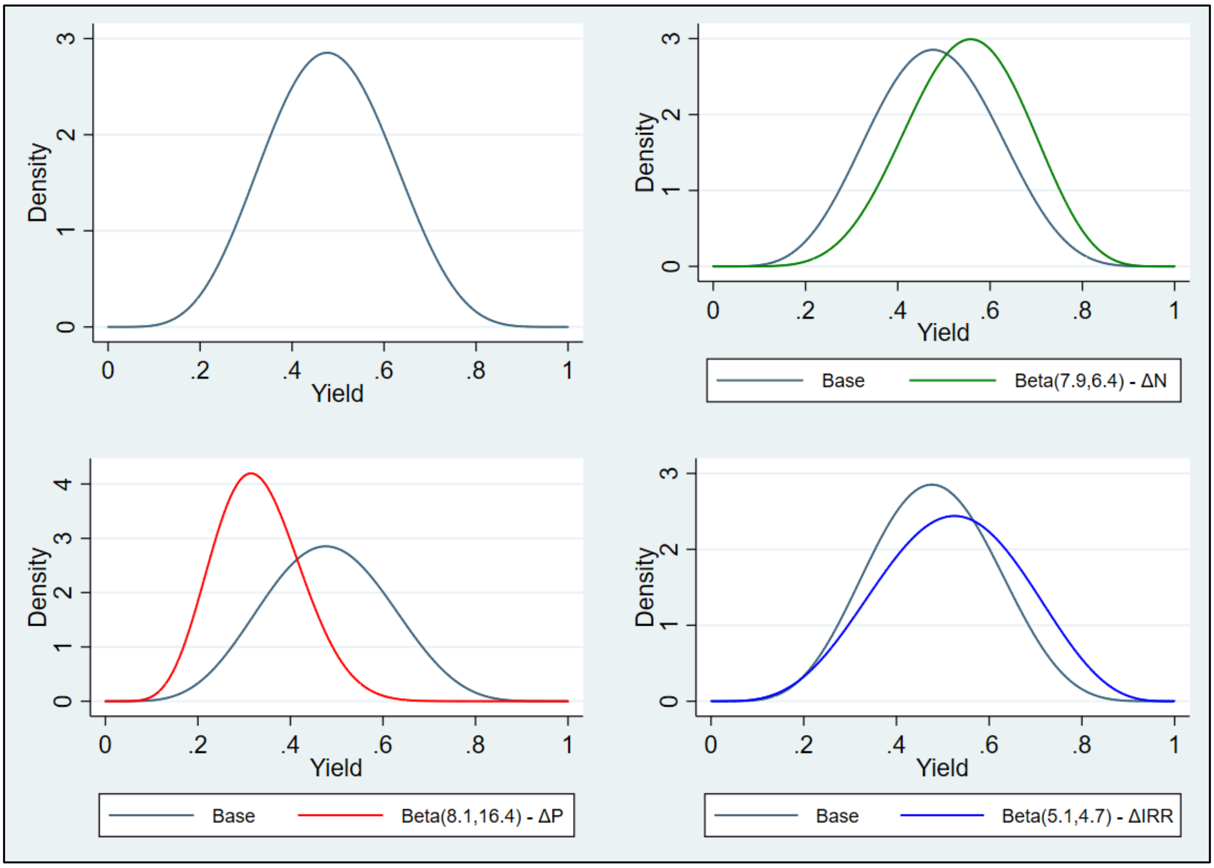


Figure 3.3: Input-conditioned Wheat (Rabi) yields estimated from the BetaIV regression framework. Top-left to right: **Panel A:** Wheat yields for an average farmer in the sample; **Panel B:** Change in yield density upon a marginal shift in nitrogen; **Panel C:** Change in yield density upon a marginal shift in irrigation; **Panel D:** Change in yield density upon a marginal shift in phosphorus.

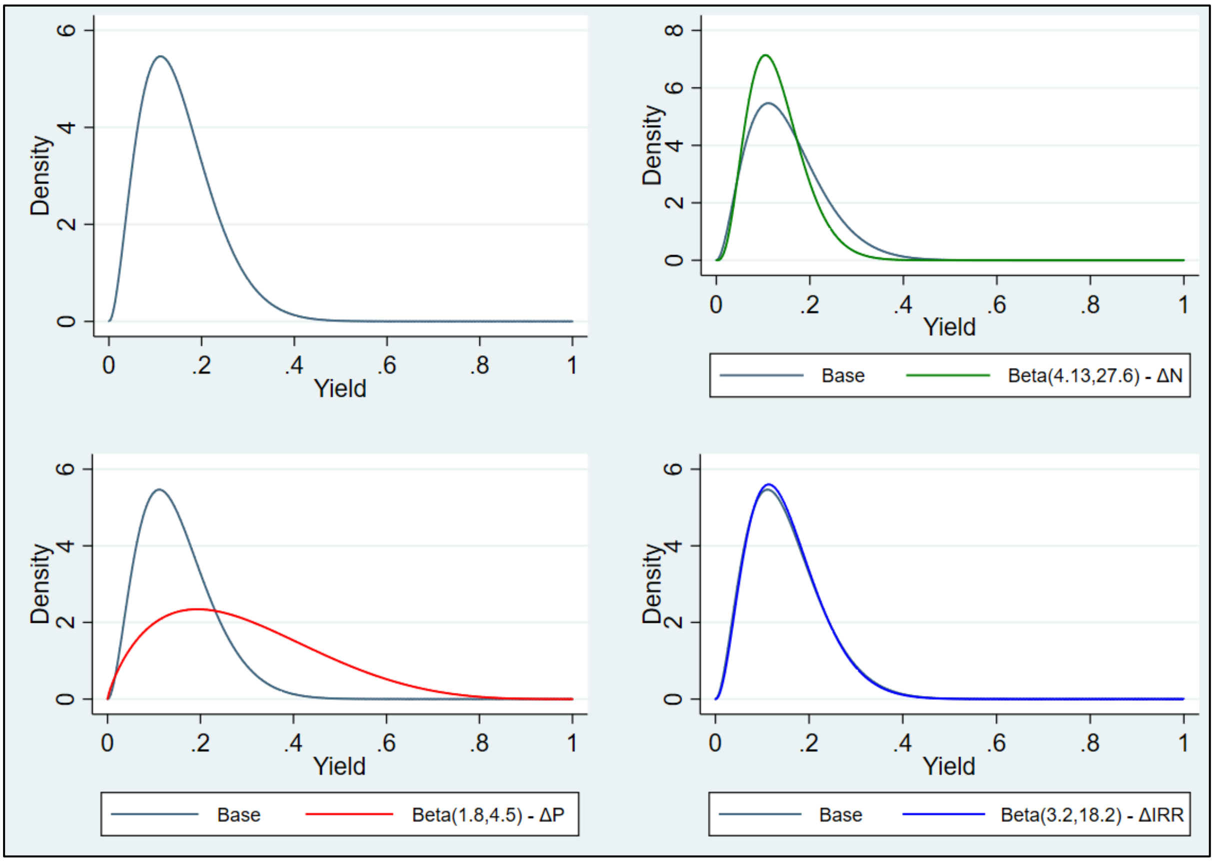


Figure 3.4: Input-conditioned Cotton (Kharif) yields estimated from the BetaIV regression framework. Top-left to right: **Panel A:** Wheat yields for an average farmer in the sample; **Panel B:** Change in yield density upon a marginal shift in nitrogen; **Panel C:** Change in yield density upon a marginal shift in irrigation; **Panel D:** Change in yield density upon a marginal shift in phosphorus.

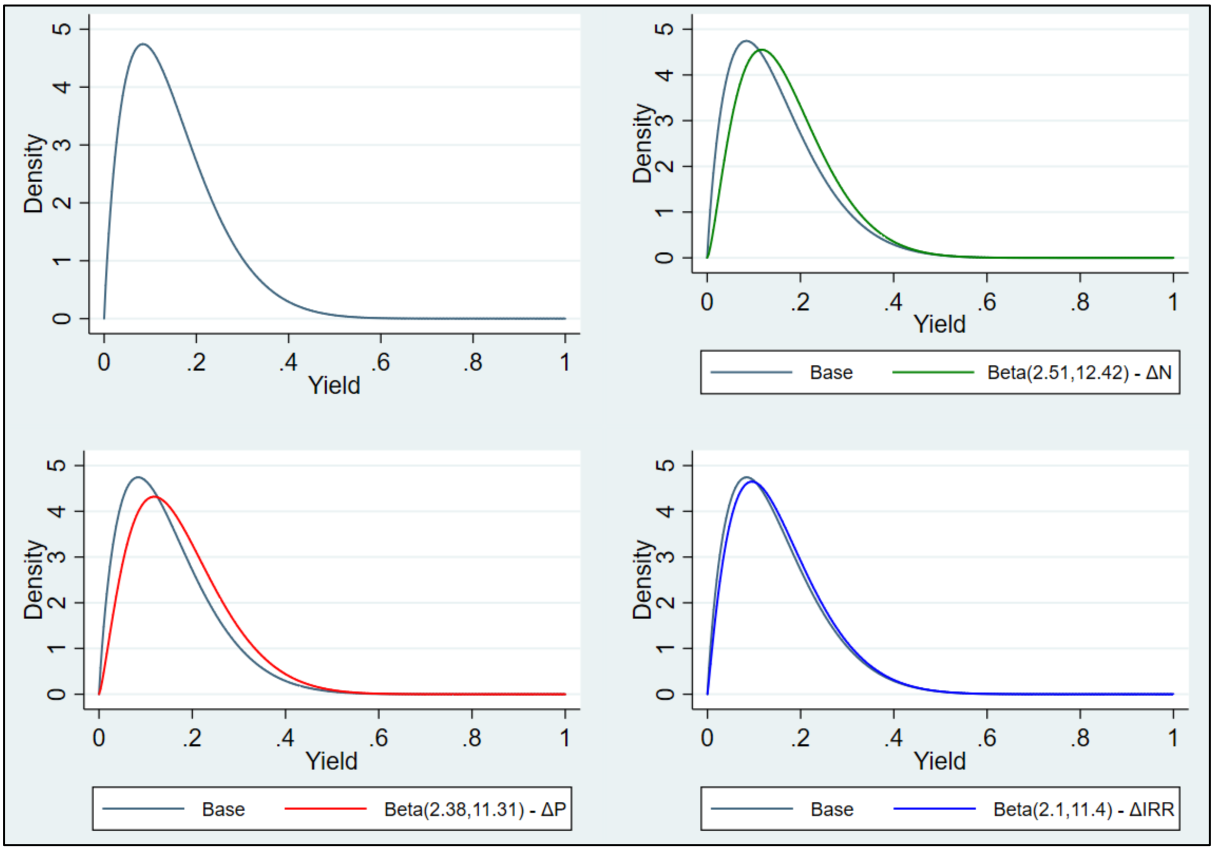


Figure 3.5: Input-conditioned Pigeonpea (Kharif) yields estimated from the BetaIV regression framework. Top-left to right: **Panel A:** Wheat yields for an average farmer in the sample; **Panel B:** Change in yield density upon a marginal shift in nitrogen; **Panel C:** Change in yield density upon a marginal shift in irrigation; **Panel D:** Change in yield density upon a marginal shift in phosphorus.

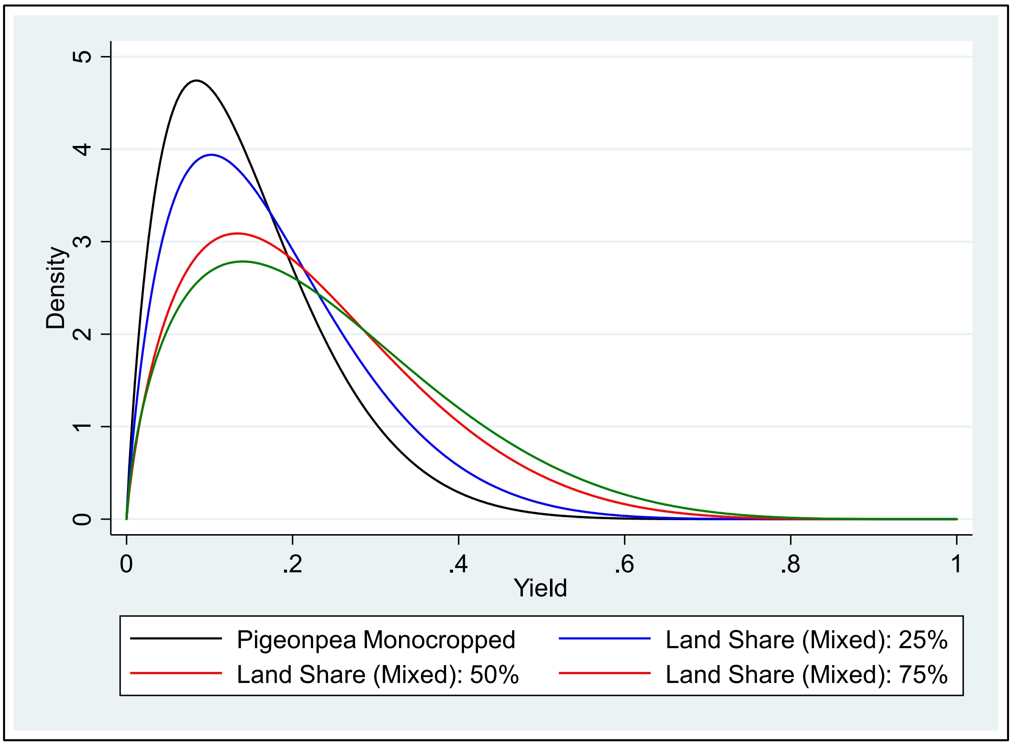


Figure 3.6: Beta Pigeonpea Yield PDF (in black). Other colours represent change in yield density due to a marginal shift in mix-cropped area | Plot(s) for a representative farmer.

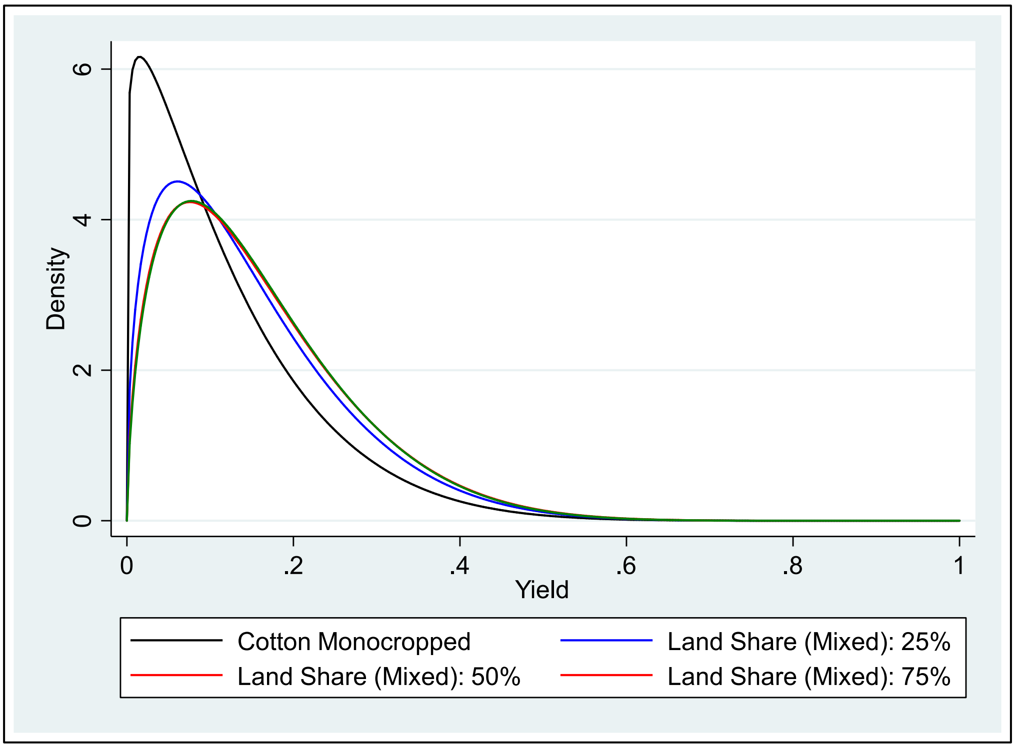


Figure 3.7: Beta Cotton Yield PDF (in black). Other colours represent change in yield density due to a marginal shift in mix-cropped area | Plot(s) for a representative farmer.

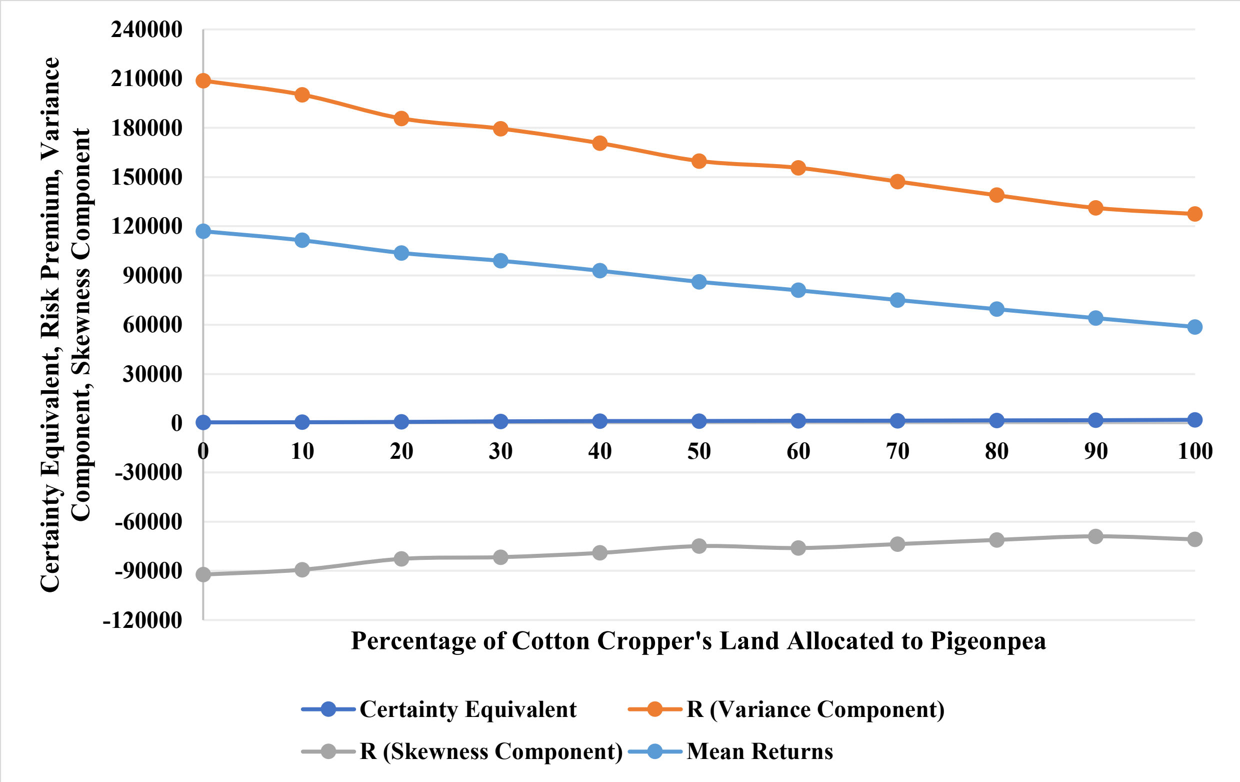


Figure 3.8: Certainty Equivalent, Mean Revenue and Risk Premium (Variance and Skewness Components. Values are in INR.

3.9 Tables

Acronym	Variable
Dependent Variable(s) $Y_{j,it}$	Yield (kg/acre) of crop $j \in$ {Paddy(Kharif), Paddy(Rabi), Wheat(Rabi), Cotton(Kharif), Pigeonpea(Kharif)} cultivated on plot i in year t.
Inputs	
$nitro_{it}$	Nitrogen applied (kg/acre)
$phos_{it}$	Phosphorus applied (kg/acre)
irr_{it}	Electric irrigation applied (hrs/acre)
Input Prices	
$P_{F,dt}$	Fertilizer price (rupees/kg) in district d in season-year t.
$P_{L,dt}$	Agricultural labour price (rupees/hr) in district d in season-year t.
$P_{I,dt}$	Irrigation charges (rupees/acre) in district d in season-year t.
Input Application Costs	
$house\_km_i$	Distance between house and plot (in kilometres)
irr_km_i	Distance between source of irrigation and plot (in kilometres)
Output Prices	
$P_{j,d,t-1}$	Farm harvest price of crop j in district d in season-year t.
Farmer Ability	
own_{ht}	Ratio of household's owned land to total operational holding in t.
age_h	Age of household head
edu_h	Years of education of household head
fwd_h	=1 if h is an upper-caste household, 0 otherwise.
Plot-Level Fixed Effects	
$problem_i$	=1 if i has saline/erosive/alkaline/poor fertility soils, 0 otherwise.
$deep_i$	=1 if i has deep (>1.1 m.) soils, 0 otherwise.
$bund_i$	=1 if plot i has bunding, 0 otherwise.
Cropping History	
$water_{i,t-1}$	=1 if water-intensive crop(s) cultivated on i in last season, 0 otherwise.
$nfix_{t-1}$	=1 if nitrogen-fixing crop(s) cultivated on i in last season, 0 otherwise.
Weather	
$temp_{vt}$	Average temperature in growing season in village v ($^{\circ}$ celsius)
rf_{vt}	Total rainfall in season $s \in \{kharif, rabi\}$ in village v (mm.)
Additional Variables for Cotton-Pigeonpea Mixed-Cropping	
M_{it}	=1 if household cultivates more than one crop on plot i in the growing season in year t, 0 otherwise.
$mix_{j,ht}$	Percentage of land mix-cropped by household h in season-year t.
$area_{ht}$; $area\_mix_{j,ht}$	Total operational holding; Area under mixed cropping
$P_{j_1,d,t-1}; P_{j_2,d,t-1}$	Lagged farm harvest price of Pigeonpea (crop j_1) and Cotton (crop j_2) in d in t.

Table 3.1: Variable Descriptions. **Note:** Details of water intensive and nitrogen fixing crops are provided in appendix (Table A4).

Variable	Paddy (Kharif) (N = 7824)					Paddy (Rabi) (N = 692)				
	Mean	SD	Min	Median	Max	Mean	SD	Min	Median	Max
$Y_{j,it}$	1254.86	744.07	0	1300	3200	1956.52	573.42	0	2000	3200
$nitro_{it}$	1.93	1.44	0	1.77	9.6	2.90	1.38	0	2.66	9.15
$phos_{it}$	0.66	0.96	0	0.48	23	1.61	0.95	0	1.58	6.27
irr_{it}	1.02	6.57	0	0	17	1.37	4.57	0	0	50
M_{it}	0.01	0.10	0	0	1	10.12	7.26	0	10.35	62.5
$P_{F,dt}$	22	7	7	24	32	9.20	3.89	6.85	7.36	31.04
$P_{L,dt}$	17.5	5	5	18	33	11.45	6.37	5.005	8.45	26.37
$P_{I,dt}$	285	339	1	126	1247	186.68	64.92	32.29	151.78	689.02
$house\_km_i$	0.83	0.73	0	0.66	30	1.28	0.84	0	1	5
irr_km_i	285.17	339.22	0	126.95	1247.52	0.16	0.33	0	0	3
$P_{paddy,d,t-1}$	9.5	1.65	5	9	18	8.37	2.95	4.98	6.57	14.97
own_{ht}	0.88	0.27	0	1	1	0.74	0.41	0	1	1
age_h	51.58	13.11	18	51	88	47.08	12.32	21	45	82
edu_h	6.08	4.96	0	6	19	3.08	3.83	0	0	15
fwd_h	0.27	0.44	0	0	1	0.30	0.46	0	0	1
$problem_i$	0.39	0.48	0	0	1	0.21	0.40	0	0	1
$deep_i$	0.47	0.49	0	0	1	0.02	0.15	0	0	1
$bund_i$	0.91	0.27	0	1	1	0.08	0.27	0	0	1
$water_{i,t-1}$	0.19	0.39	0	0	1	0.17	0.38	0	0	1
$nfix_{t-1}$	0.24	0.43	0	0	1	27.70	0.70	24.80	27.77	29.86
$temp_{vt}$	28.62	0.83	24.44	28.87	30.30	673.70	181.93	265.37	643.84	1510.03
rf_{vt}	946.84	267.73	265.37	911.32	2000.58	0.83	0.37	0	1	1

Table 3.2: Summary Statistics (Paddy (Kharif and Rabi))

Variable	Wheat (Rabi) (N = 4108)				
	Mean	SD	Min	Median	Max
$Y_{j,it}$	907.73	389.11	0	900	2000
$nitro_{it}$	2.93	2.07	0	2.55	9.93
$phos_{it}$	1.30	1.08	0	1.06	8.73
irr_{it}	1.37	4.57	0	0	50
M_{it}	0.48	1.07	0	0	20
$P_{F,dt}$	23.31	5.33	13.13	21.71	36.01
$P_{L,dt}$	19.65	4.31	10.52	20.27	29.72
$P_{I,dt}$	1326.10	405.39	333.26	1220.53	2076.30
$house\_km_i$	0.98	0.82	0	0.8	15
irr_km_i	0.28	2.38	0	0.1	100
$P_{wheat,d,t-1}$	11.60	2.62	6.61	12.12	20.77
own_{ht}	0.84	0.31	0	1	1
age_h	51.75	12.94	23	51	90
edu_h	7.42	4.95	0	9	19
fwd_h	0.33	0.47	0	0	1
$problem_i$	0.37	0.48	0	0	1
$deep_i$	0.43	0.49	0	0	1
$bund_i$	0.25	0.37	0	1	1
$water_{i,t-1}$	0.54	0.43	0	0	1
$nfix_{t-1}$	28.18	0.49	0	1	1
$temp_{vt}$	897.64	1.21	25.52	28.87	29.74
rf_{vt}	0.82	269.63	337.48	849.36	2000.58

Table 3.3: Summary Statistics (Wheat (Rabi))

Variable	Cotton (N = 2857)					Pigeonpea (N=2156)				
	Mean	**SD**	**Min**	**Median**	**Max**	**Mean**	**SD**	**Min**	**Median**	**Max**
$Y_{j,it}$	426.24	324.11	0	335.97	2800	352.55	404.26	0	241	4000
$nitro_{it}$	0.67	0.63	0	0.5	6.98	0.31	0.41	0	0.20	4.18
$phos_{it}$	0.53	0.56	0	0.38	5.83	0.25	0.37	0	0.16	4
irr_{it}	0.14	0.53	0	0	6.31	0.06	0.32	0	0	7.62
M_{it}	0.41	0.49	0	0	1	0.48	0.50	0	0	1
$mix_{j,ht}$	0.31	0.21	0	0.28	1	0.87	0.09	0	0.87	1
$P_{F,dt}$	13.53	5.80	4.2	14.14	35.79	16.94	7.09	7.27	14.71	39.96
$P_{L,dt}$	11.34	6.60	3.75	8.91	27.15	10.73	6.19	3.58	7.25	26.47
$P_{I,dt}$	168.02	260.23	0	66.63	1358.81	52.49	62.71	0	34.47	626.12
$house\_km_i$	1.44	0.87	0	1.25	15.5	1.70	1.46	0	1.5	18
irr_km_i	0.65	5.93	0	0.19	150	0.59	3.27	0	0.2	100
$P_{1,dt}$	28.14	6.64	16.06	27.87	43.5	28.84	4.82	16.06	27.88	47.83
$P_{2,dt}$	24.62	8.34	11.83636	20.20	43.08	24.99	7.93	11.84	20.20	45.95
own_{ht}	0.83	0.30	0	1	1	0.86	0.30	0	1	1
age_h	47.87	11.44	19	47	82	49.74	11.86	17	49	87
edu_h	4.98	4.44	0	4	18	5.67	4.52	0	5	18
fwd_h	0.16	0.37	0	0	1	0.28	0.45	0	0	1
$problem_i$	0.55	0.50	0	1	1	0.57	0.50	0	1	1
$deep_i$	0.14	0.35	0	0	1	0.26	0.44	0	0	1
$bund_i$	0.60	0.49	0	1	1	0.64	0.48	0	1	1
$water_{i,t-1}$	0.33	0.47	0	0	1	0.73	0.44	0	1	1
$nfix_{t-1}$	0.70	0.46	0	1	1	0.56	0.50	0	1	1
$temp_{vt}$	27.74	0.79	25.20	27.51	30.54	27.10	0.70	24.45	27.11	30.54
rf_{vt}	749.94	256.80	340.21	752.97	2000.59	718.46	215.30	283.34	710.87	2000.59

Table 3.4: Summary Statistics (Cotton and Pigeonpea)

Notes: Mixed cropping percentage is reported for mixed-croppers, i.e., $N_{\text{cotton}} = 1163$ and $N_{\text{pigeonpea}} = 1034$.

Paddy (Kharif) Inputs **Explanatory Variables**	**(1)** $nitro_{it}$	**(2)** $phos_{it}$	**(3)** irr_{it}
$P_{F,dt}$	-0.02** (0.009)	-0.02*** (0.007)	-2.01*** (0.10)
$P_{L,dt}$	-0.06*** (0.01)	-0.07*** (0.01)	1.32*** (0.15)
$P_{I,dt}$	-0.001*** (0.00)	0.002 (0.002)	0.02*** (0.001)
irr_km_i	-0.01* (0.008)	0.009 (0.006)	0.10*** (0.03)
$P_{paddy,d,t-1}$	0.05 (0.03)	0.10*** (0.02)	-2.64*** (0.23)
own_{ht}	0.02 (0.07)	0.28*** (0.06)	-0.97* (0.58)
age_h	0.007*** (0.002)	0.001 (0.001)	-0.06*** (0.01)
edu_h	0.01** (0.00)	0.005 (0.004)	0.07 (0.04)
fwd_h	-0.32*** (0.05)	0.08* (0.04)	2.81*** (0.48)
$problem_i$	-0.06 (0.04)	-0.28*** (0.03)	-3.42*** (0.42)
$deep_i$	0.41*** (0.05)	-0.31*** (0.04)	-5.10*** (0.62)
$water_{i,t-1}$	0.23*** (0.06)	0.35*** (0.05)	4.86*** (0.51)
$nfix_{t-1}$	-0.10* (0.05)	-0.40*** (0.04)	2.62*** (0.45)
$temp_{vt}$	-0.10 (0.07)	-0.17*** (0.05)	-3.83*** (0.37)
rf_{vt}	-0.00* (0.00)	-0.00 (0.00)	-0.003* (0.001)
Trend	0.12*** (0.04)	0.04 (0.03)	2.77*** (0.23)
Constant	2.55 (2.26)	4.37** (1.75)	115.93*** (10.45)
District Fixed Effect		Yes	
Observations		7824	
Log Likelihood		-28250.87	

Table 3.5: Input Regression Results (Paddy (Kharif))

Paddy (Rabi) Inputs Explanatory Variables	**(1)** $nitro_{it}$	**(2)** $phos_{it}$	**(3)** irr_{it}
$P_{F,dt}$	-0.23***	-0.081	0.007
	(0.04)	(0.03)	(0.16)
$P_{L,dt}$	0.25***	0.126**	-1.50***
	(0.06)	(0.05)	(0.42)
$P_{I,dt}$	-0.006***	-0.003**	-0.04***
	(0.002)	(0.001)	(0.007)
irr_km_i	-0.22	-0.003**	-0.85***
	(0.15)	(0.001)	(0.27)
$P_{paddy,d,t-1}$	0.05	0.10*	-1.65
	(0.03)	(0.02)	(0.65)
own_{ht}	0.21*	-0.05	-0.42
	(0.12)	(0.05)	(0.48)
age_h	-0.007	-0.001	0.006
	(0.004)	(0.003)	(0.01)
edu_h	0.001	-0.01	-0.05
	(0.01)	(0.01)	(0.06)
fwd_h	-0.45***	-0.16	-0.68
	(0.14)	(0.11)	(0.58)
$problem_i$	0.01	-0.06	0.02
	(0.12)	(0.10)	(0.50)
$deep_i$	-0.25	-0.17	-2.50
	(0.36)	(0.28)	(3.62)
$water_{i,t-1}$	-0.02	-0.03	0.99
	(0.19)	(0.15)	(0.71)
$nfix_{t-1}$	-0.03	-0.01	-0.31
	(0.14)	(0.11)	(0.56)
$temp_{vt}$	0.48***	0.15	4.74***
	(0.18)	(0.14)	(0.69)
rf_{vt}	-0.001	-0.001*	-0.02***
	(0.001)	(0.001)	(0.004)
Trend	-0.10	-0.11	6.53***
	(0.09)	(0.06)	(0.68)
Constant	-2.16	-0.46	-112.25***
	(4.39)	(3.36)	(18.71)
District Fixed Effect		Yes	
Observations		692	
Log Likelihood		-3703.2	

Table 3.6: Input Regression Results (Paddy (Rabi))

Wheat (Rabi) Inputs Explanatory Variables	**(1)** $nitro_{it}$	**(2)** $phos_{it}$	**(3)** irr_{it}
$P_{F,dt}$	0.03** (0.01)	0.01 (0.01)	0.018 (0.02)
$P_{L,dt}$	0.02 (0.01)	-0.04*** (0.01)	-0.03 (0.03)
$P_{I,dt}$	-0.001*** (0.000)	0.00** (0.00)	-0.00 (0.00)
irr_km_i	-0.007 (0.01)	0.002 (0.007)	0.08 (0.05)
$P_{wheat,d,t-1}$	0.05** (0.01)	0.10 (0.20)	1.68 (0.65)
own_{ht}	-0.12 (0.08)	0.00 (0.06)	0.33** (0.14)
age_h	0.002 (0.002)	-0.003** (0.001)	0.005 (0.004)
edu_h	0.02*** (0.006)	0.020*** (0.004)	0.03*** (0.01)
fwd_h	-0.08 (0.06)	-0.02 (0.04)	0.82*** (0.11)
$problem_i$	-0.11** (0.05)	-0.032 (0.042)	-0.17* (0.10)
$deep_i$	0.01 (0.06)	-0.41*** (0.05)	-0.26** (0.12)
$water_{i,t-1}$	0.19*** (0.06)	0.07 (0.04)	0.59*** (0.11)
$nfix_{t-1}$	-0.001 (0.05)	0.02 (0.04)	0.41*** (0.09)
$temp_{vt}$	0.09 (0.08)	0.28*** (0.06)	0.94*** (0.142)
rf_{vt}	0.00 (0.001)	0.00 (0.00)	-0.005*** (0.001)
Trend	-0.07** (0.03)	0.02 (0.02)	0.94*** (0.07)
Constant	-0.77 (1.96)	-6.80*** (1.499)	-25.98*** (3.63)
District Fixed Effect		Yes	
Observations		4108	
Log Likelihood		-15046.46	

Table 3.7: Input Regression Results (Wheat (Rabi))

Cotton Inputs Explanatory Variables	(1) $nitro_{it}$	(2) $phos_{it}$	(3) irr_{it}	(4) $area\_mix_{j,ht}$
$area_{ht}$				0.284^{***} (0.010)
$P_{F,dt}$	-0.037^{***} (0.009)	-0.028^{***} (0.009)	-0.071^{***} (0.021)	-0.047^{**} (0.020)
$P_{L,dt}$	0.110^{***} (0.016)	0.069^{***} (0.017)	0.002 (0.040)	-0.170^{***} (0.047)
$P_{I,dt}$	-0.0003 (0.0004)	0.0004 (0.0003)	-0.002^{***} (0.0002)	-0.003^{***} (0.001)
irr_km_i	-0.0001 (0.003)	0.001 (0.004)	-1.060^{***} (0.183)	0.001 (0.007)
$P_{1,d,t-1}$	0.027^{***} (0.007)	0.030^{***} (0.007)	-0.022 (0.014)	-0.006 (0.024)
$P_{2,d,t-1}$	-0.002 (0.007)	-0.003 (0.008)	-0.016 (0.061)	0.003 (0.021)
own_{ht}	0.077 (0.073)	0.071 (0.078)	0.244 (0.190)	-0.168 (0.214)
age_h	-0.002 (0.002)	-0.003 (0.002)	0.001 (0.005)	-0.000 (0.005)
edu_h	-0.004 (0.006)	-0.013^{**} (0.006)	0.014 (0.015)	-0.053^{***} (0.017)
fwd_h	-0.092 (0.064)	0.059 (0.069)	0.259^{*} (0.156)	-0.055 (0.163)
$problem_i$	-0.025 (0.043)	-0.046 (0.046)	-0.425^{***} (0.113)	0.438^{***} (0.121)
$deep_i$	0.236^{***} (0.072)	0.227^{***} (0.077)	-0.462^{*} (0.255)	-0.690^{***} (0.183)
$water_{i,t-1}$	-0.047 (0.051)	-0.065 (0.055)	1.126^{***} (0.133)	-0.413^{**} (0.166)
$nfix_{t-1}$	-0.112^{*} (0.063)	-0.082 (0.067)	0.099 (0.146)	4.010^{***} (0.385)
$temp_{vt}$	-0.308^{***} (0.068)	-0.327^{***} (0.073)	-0.327^{**} (0.164)	-0.037 (0.174)
rf_{vt}	-0.0001^{***} (0.00001)	-0.0006^{*} (0.0003)	-0.0007 (0.0009)	-0.001^{**} (0.0005)
Trend	-0.133^{***} (0.017)	-0.089^{***} (0.019)	0.349^{***} (0.066)	-0.234^{***} (0.045)
Constant	10.455^{***} (1.887)	10.693^{***} (2.036)	5.957 (4.501)	-1.028 (4.764)
District Fixed Effect	Yes			
Observations	2783	2783	2783	2783
Log Likelihood	-10336.917			

Table 3.8: Input Regression Results (Cotton)

Pigeonpea Inputs **Explanatory Variables**	**(1)** $nitro_{it}$	**(2)** $phos_{it}$	**(3)** irr_{it}	**(4)** $area\_mix_{j,ht}$
$area_{ht}$				0.284^{***} (0.010)
$P_{F,dt}$	-0.037^{***} (0.009)	-0.028^{***} (0.009)	-0.071^{***} (0.021)	-0.047^{**} (0.020)
$P_{L,dt}$	0.110^{***} (0.016)	0.069^{***} (0.017)	0.002 (0.040)	0.170^{***} (0.047)
$P_{I,dt}$	-0.0003 (0.0004)	0.0004 (0.0003)	-0.002^{***} (0.0002)	-0.003^{***} (0.001)
irr_km_{i}	-0.0001 (0.003)	0.001 (0.004)	-1.060^{***} (0.183)	0.001 (0.007)
$P_{1,d,t-1}$	0.027^{***} (0.007)	0.030^{***} (0.007)	-0.022 (0.014)	-0.004 (0.024)
$P_{2,d,t-1}$	-0.002 (0.007)	-0.003 (0.008)	-0.016 (0.061)	0.003 (0.021)
own_{ht}	0.077 (0.073)	0.071 (0.078)	0.244 (0.190)	-0.168 (0.214)
age_{h}	-0.002 (0.002)	-0.003 (0.002)	0.001 (0.005)	-0.000 (0.005)
edu_{h}	-0.004 (0.006)	-0.013^{**} (0.006)	0.014 (0.015)	-0.053^{***} (0.017)
fwd_{h}	-0.092 (0.064)	0.059 (0.069)	0.259^{*} (0.156)	-0.055 (0.163)
$problem_{i}$	-0.025 (0.043)	-0.046 (0.046)	-0.425^{***} (0.113)	0.438^{***} (0.121)
$deep_{i}$	0.236^{***} (0.072)	0.227^{***} (0.077)	-0.462^{*} (0.255)	-0.690^{***} (0.183)
$water_{i,t-1}$	-0.047 (0.051)	-0.065 (0.055)	1.126^{***} (0.133)	-0.413^{**} (0.166)
$nfix_{t-1}$	-0.112^{*} (0.063)	-0.082 (0.067)	0.099 (0.146)	4.010^{***} (0.385)
$temp_{vt}$	-0.308^{***} (0.068)	-0.327^{***} (0.073)	-0.327^{**} (0.164)	-0.037 (0.174)
rf_{vt}	-0.0001^{***} (0.00001)	-0.0006^{*} (0.0003)	-0.0007 (0.0009)	-0.001^{**} (0.0005)
Trend	-0.133^{***} (0.017)	-0.089^{***} (0.019)	0.349^{***} (0.066)	-0.234^{***} (0.045)
Constant	10.455^{***} (1.887)	10.693^{***} (2.036)	5.957 (4.501)	-1.028 (4.764)
District Fixed Effects		Yes		
Observations	2783	2783	2783	2783
Log Likelihood		-10336.917		

Table 3.9: Input Regression Results (Pigeonpea)

Predictors	μ (Paddy (Kharif) Yield)	ϕ (Paddy (Kharif) Yield)	μ (Paddy (Rabi) Yield)	ϕ (Paddy (Rabi) Yield)
$nitro_{it}$	0.26*** (0.09)	0.39* (0.20)	-0.31 (0.22)	-0.32 (0.57)
$nitro_{it}^2$	-0.06*** (0.07)	-0.07*** (0.02)	-0.01*** (0.003)	0.04* (0.02)
$phos_{it}$	-0.08 (0.05)	-0.28** (0.13)	0.36 (0.45)	-0.70 (1.20)
irr_{it}	-0.01 (0.01)	-0.04** (0.02)	-0.00 (0.01)	-0.03* (0.02)
$problem_i$	-0.09*** (0.03)	-0.01 (0.07)	0.08 (0.07)	0.51** (0.21)
$deep_i$	0.01 (0.05)	-0.32*** (0.12)	0.25 (0.18)	-0.37 (0.39)
$bund_i$	0.20* (0.10)	0.78 (0.22)	-0.153* (0.08)	0.06* (0.27)
age_h	0.00 (0.001)	-0.005** (0.002)	-0.002 (0.003)	-0.001 (0.008)
edu_h	-0.00 (0.003)	0.002 (0.006)	0.013 (0.009)	-0.05* (0.03)
t	0.03*** (0.009)	0.004 (0.02)	0.011 (0.01)	0.06* (0.03)
$temp_{vt}$	-0.07 (0.009)	-0.80*** (0.10)	-0.30*** (0.11)	-0.72* (0.37)
rf_{vt}	-0.00 (0.00)	-0.001*** (0.00)	0.00* (0.00)	0.001** (0.001)
$nitro\_residuals_{it}$	0.10 (0.08)	-0.01 (0.19)	0.38* (0.22)	0.19 (0.57)
$irr\_residuals_{it}$	0.009*** (0.002)	0.01** (0.006)	-0.005 (0.01)	0.08** (0.03)
$phos\_residuals_{it}$	-0.06 (0.04)	0.02 (0.12)	-0.39 (0.44)	0.85 (1.21)
Constant	-0.83*** (0.16)	0.27 (0.34)	0.66*** (0.08)	2.11*** (0.34)
Observations	7824		692	
District Fixed Effects	Yes			
Log-Likelihood	6150.324		231.86	

Table 3.10: Yield Regression Results (Paddy (Kharif) and Paddy (Rabi))

Predictors	μ (Wheat Yield)	ϕ (Wheat Yield)
$nitro_{it}$	0.27*** (0.06)	0.10 (0.24)
$nitro_{it}^2$	-0.00 (0.006)	-0.02* (0.01)
$phos_{it}$	-0.55*** (0.09)	0.65** (0.30)
irr_{it}	0.19*** (0.03)	-0.30*** (0.10)
$problem_i$	-0.06** (0.02)	-0.12 (0.13)
$deep_i$	-0.08 (0.05)	0.19 (0.19)
age_h	-0.004*** (0.001)	0.007 (0.005)
edu_h	0.005 (0.003)	0.003 (0.015)
t	0.01 (0.01)	0.15*** (0.04)
$temp_{vt}$	0.05 (0.03)	0.13 (0.15)
rf_{vt}	-0.00 (0.00)	0.001** (0.000)
$nitro\_residuals_{it}$	-0.16*** (0.06)	-0.05 (0.25)
$irr\_residuals_{it}$	-0.04* (0.02)	0.09 (0.09)
$phos\_residuals_{it}$	0.58*** (0.09)	-0.57* (0.31)
Constant	-0.27 (0.19)	0.70 (0.66)
Observations	4108	
District Fixed Effects	Yes	
Log-Likelihood	1673.3	

Table 3.11: Yield Regression Results (Wheat (Rabi))

Predictors	μ (Cotton Yield)	ϕ (Cotton Yield)	μ (Pigeonpea Yield)	ϕ (Pigeonpea Yield)
$nitro_{it}$	0.142 (0.309)	0.064 (0.827)	-0.310 (0.486)	1.289 (0.936)
$nitro_{it}^2$	-0.012 (0.207)	-1.343*** (0.478)	-0.521* (0.297)	0.386 (0.487)
$phos_{it}$	1.425*** (0.447)	-0.916 (1.244)	1.596** (0.627)	-1.885* (1.106)
irr_{it}	0.081 (0.054)	0.535 (0.712)	0.404 (1.250)	-0.245 (2.066)
$mix_{j,ht}$	0.697*** (0.180)	0.431 (0.444)	0.816*** (0.227)	-0.828** (0.295)
$problem_i$	-0.113* (0.068)	0.136 (0.159)	0.081 (0.085)	-0.093 (0.179)
$deep_i$	-0.044 (0.112)	-0.023 (0.285)	-0.412*** (0.124)	-0.150 (0.213)
$bund_i$	0.287** (0.093)	-0.025 (0.033)	-0.13** (0.06)	0.202** (0.8)
$water_{i,t-1}$	0.752*** (0.108)	-0.687*** (0.192)	0.451*** (0.109)	-0.184 (0.204)
$nfix_{t-1}$	0.104* (0.063)	0.162 (0.194)	0.223*** (0.056)	-0.168* (0.03)
age_h	-0.008*** (0.003)	0.000 (0.007)	-0.007** (0.003)	-0.002 (0.007)
edu_h	0.015 (0.009)	-0.042* (0.025)	0.030* (0.015)	-0.084** (0.033)
t	0.085*** (0.013)	-0.093*** (0.029)	-0.000 (0.021)	-0.105*** (0.033)
$temp_{vt}$	0.232 (0.142)	-0.589** (0.297)	0.597*** (0.201)	0.327 (0.334)
rf_{vt}	-0.001*** (0.000)	-0.001 (0.001)	0.001*** (0.000)	0.000 (0.000)
$nitro\_residuals_{it}$	0.407 (0.312)	0.328 (0.853)	0.930 (0.570)	-2.225** (1.113)
$irr\_residuals_{it}$	-0.007** (0.003)	0.009*** (0.003)	-0.016*** (0.004)	0.015*** (0.003)
$mix\_residuals_{it}$	-0.048*** (0.013)	0.161*** (0.040)	0.018 (0.012)	0.075* (0.039)
$phos\_residuals_{it}$	-0.996** (0.439)	0.297 (1.264)	-1.323** (0.641)	1.908* (1.126)
Constant	-1.945*** (0.059)	2.760*** (0.144)	-3.084*** (0.095)	3.261*** (0.166)
District Fixed Effects	Yes			
Log-Likelihood	1270.54		2492.37	

Table 3.12: Yield Regression Results (Cotton and Pigeonpea)

Predictors	μ (Intercropped Cotton Yield)	μ (Intercropped Pigeonpea Yield)
$nitro_{it}$	-0.126 (0.203)	0.301 (0.462)
$nitro_{it}^2$	0.005** (0.002)	-0.205** (0.096)
$phos_{it}$	0.894*** (0.212)	-0.854 (0.625)
irr_{it}	0.050 (0.032)	0.306*** (0.043)
$mix_{j,ht}$	0.91*** (0.05)	0.81*** (0.01)
$problem_i$	0.048 (0.034)	-0.126 (0.077)
$deep_i$	0.058 (0.065)	0.306*** (0.078)
$bund_i$	0.3** (0.093)	-0.11** (0.06)
$water_{i,t-1}$	0.219*** (0.043)	0.175** (0.072)
$nfix_{t-1}$	0.148*** (0.039)	0.175** (0.072)
age_h	0.001 (0.002)	-0.003 (0.003)
edu_h	0.006 (0.005)	0.009 (0.007)
t	0.101*** (0.009)	-0.021 (0.022)
$temp_{vt}$	-0.191*** (0.067)	-0.277*** (0.105)
rf_{vt}	0.0001* (0.000)	-0.000 (0.000)
$nitro\_residuals_{it}$	0.301 (0.209)	0.311 (0.415)
$irr\_residuals_{it}$	-0.014** (0.006)	-0.144*** (0.036)
$mix\_residuals_{it}$	-0.822*** (0.213)	1.035* (0.625)
$phos\_residuals_{it}$	-1.896*** (0.107)	-3.278*** (0.388)
Constant	1.744*** (0.313)	2.594*** (0.467)
District Fixed Effects	Yes	
Log-Likelihood	3650.7065	
Covariance	0.146***	

Table 3.13: Yield Regression Results (Cotton and Pigeonpea (Simultaneous))

3.10 References

Ahlstrom, A., M. Raupach, G. Schurgers, B. B.N. Smith, A. Arneth, M. Jung, M. Reichstein, et al. 2015. "The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink." Science 348: 6237 895-899.

Altieri, M. A. & Nicholls, C. I., 2017. The adaptation and mitigation potential of traditional agriculture in a changing climate. Climatic Change, 40, pp. 33-45.

Anderson, J.R., J.L. Dillon and B. Hardaker (1977), Agricultural Decision Analysis (Iowa State University Press, Ames, IA).

Antle, John M. 1983. "Testing the Stochastic Structure of Production: A Flexible Moment-Based Approach." *Journal of Business and Economic Statistics* 1 (3): 192. https://doi.org/10.2307/1391337.

Antle, John M. 1987. "Econometric Estimation of Producers' Risk Attitudes." American Journal of Agricultural Economics 69 (3): 509–22. https://doi.org/10.2307/1241687.

Antle, J. M., and W. J. Goodger. 1984. "Measuring Stochastic Technology: The Case of Tulare Milk Production." *American Journal of Agricultural Economics* 66 (3): 342–50. https://doi.org/10.2307/1240801.

Antón, J., Kimura, S., Lankoski, J. & Cattaneo, A., 2012. A Comparative Study of Risk Management in Agriculture under Climate Change, Paris: OECD Publishing.

Aggarwal, P. K. (2008). Global climate change and Indian agriculture: Impacts, adaptation and mitigation. Indian Journal of Agricultural Sciences,78(11), 0019–5022.

Aggarwal, P. K., & Mall, R. K. (2002). Climate change and rice yields in diverse agro-environments of India. II. Effect of uncertainties in scenarios and crop models on impact assessment. Climatic Change,52(3), 331–343.

Arora, G., Feng, H., Anderson, C. & Hennessy, D. A., 2020. Evidence of climate change impacts on crop comparative advantage and land use. Agricultural Economics, 51(2), pp. 221-236.

Arora, G., Agarwal, S. K., Shukla, S. 2021. Agricultural input use and index insurance adoption: Concept and evidence. Working Paper. https://doi.org/10.22004/AG.ECON.304508

Arrow, K. J., 1958. Essays in the Theory of Risk-Bearing. Chicago: Markham

Atwood, Joseph, Saleem Shaik, and Myles Watts. 2003. "Are Crop Yields Normally Distributed? A Reexamination." American Journal of Agricultural Economics, 85(4) 888-901.

Babcock, B. A. & Hennessy, D. A., 1996. Input Demand under Yield and Revenue Insurance. American Journal of Agricultural Economics, 78(2), pp. 416-427.

Barrett, Christopher B., and Leah E. M. Bevis . 2015. "The Self-Reinforcing Feedback between Low Soil Fertility and Chronic Poverty." Nature Geoscience (8) 907-912.

Binswanger, Hans P. 1980. "Attitudes Toward Risk: Experimental Measurement in Rural India." American Journal of Agricultural Economics 62 (3): 395–407. https://doi.org/10.2307/1240194.

Binswanger, Hans P. 1981. "Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India." The Economic Journal, 91(364) 867-890.

Blaise, D., G. Majumdar, and K.U. Tekale. 2004. "On-farm Evaluation of Fertilizer Application and Conservation Tillage on Productivity of Cotton+Pigeonpea Strip Intercropping on Rainfed Vertisols of Central India." Soil and Tillage Research 84 (1): 108–17. https://doi.org/10.1016/j.still.2004.09.016.

Buccola, S.T. 1986. Testing for Nonnormality in Farm Net Returns. American Journal of Agricultural Economics 68:334-343.

Buschena, David E., and David Zilberman. 1994. "What Do We Know About Decision Making Under Risk and Where Do We Go from Here?" Journal of Agricultural and Resource Economics, 19(2) 425-445.
Byerlee, D., Janvry, A., & Sadoulet, E. 2009. Agriculture for Development: Toward a New Paradigm. Annual Review of Resource Economics, 1, 15-31.
Calzadilla, A. et al., 2013. Climate change impacts on global agriculture. Climatic Change, Volume 120, pp. 357-374.
Campbell, B.M., S. Jeffrey, W. Kozanayi, M. Luckert, M. Mutamba, and C. Zindi. 2002. Household Livelihoods in Semiarid Regions: Options and Constraints. Research Publication, Indonesia: Center for International Forestry Research (CIFOR). Accessed here.
Cervantes-Godoy, D., S. Kimura and J. Antón (2013-06-10), "Smallholder Risk Management in Developing Countries", OECD Food, Agriculture and Fisheries Papers, No. 61, OECD Publishing, Paris. Accessed here.
Chambers, RG & Quiggin, J. 2000, Uncertainty, Production, Choice and Agency: The State-Contingent Approach. vol. 1, First edn, Cambridge University Press, New York.
Chavas, Jean-Paul, and Matthew T. Holt. 1990. "Acreage Decisions under Risk: The Case of Corn and Soybeans." American Journal of Agricultural Economics 72, no. 3: 529–38. https://doi.org/10.2307/1243021.
Chavas, Jean-Paul, and Matthew T. Holt. 1996. "Economic Behavior Under Uncertainty: A Joint Analysis of Risk Preferences and Technology." The Review of Economics and Statistics 78 (2): 329. https://doi.org/10.2307/2109935.
Coe, R., Njoloma, J. & Sinclair, F. To control or not to control: How do we learn more about how agronomic innovations perform on farms? Experimental Agriculture 55(S1), 303–309 (2017)
Collender, R. N., and J. A. Chalfant. 1986. "An Alternative Approach to Decisions under Uncertainty Using the Empirical Moment-Generating Function." Amer. J. Agr. Econ. 68:727-31
Cunningham, Scott. 2021. "Chapter 1, Section 1.3: Optimization Makes Everything Endogenous." In Causal Inference: The Mixtape., by Scott Cunningham, 19-21. New Haven and London: Yale University Press.
Day, Richard. 1965. Probability Distributions of Field Crop Yields. Journal of Farm Economics, 47, 713-41.
Dercon, Stefan. 1996. "Risk, Crop Choice, and Savings: Evidence From Tanzania." Economic Development and Cultural Change 44 (3): 485–513. https://www.jstor.org/stable/1154463.
Dewbre, J., D. Cervantes-Godoy and S. Sorescu (2011), Agricultural Progress and Poverty Reduction: Synthesis Report", OECD Food, Agriculture and Fisheries Papers, No. 49, OECD Publishing. Accessed here.
Di Falco, Salvatore, and Charles Perrings. 2003. "Crop Genetic Diversity, Productivity and Stability of Agroecosystems. A Theoretical and Empirical Investigation." Scottish Journal of Political Economy 50 (2): 207–16. https://doi.org/10.1111/1467-9485.5002006.
Di Falco, Salvatore, and Charles Perrings. 2005. "Crop Biodiversity, Risk Management and the Implications of Agricultural Assistance." Ecological Economics 55 (4): 459–66. https://doi.org/10.1016/j.ecolecon.2004.12.005.
Di Falco, S. 2006. "Crop Genetic Diversity, Farm Productivity and the Management of Environmental Risk in Rainfed Agriculture." European Review of Agricultural Economics 33 (3): 289–314. https://doi.org/10.1093/eurrag/jbl016.
Di Falco, Salvatore, Jean-Paul Chavas, and Melinda Smale. 2007. "Farmer Management of Production Risk on Degraded Lands: The Role of Wheat Variety Diversity in the Tigray

Region, Ethiopia." Agricultural Economics 36 (2): 147–56. https://doi.org/10.1111/j.1574-0862.2007.00194.x.
Di Falco, Salvatore, and Jean-Paul Chavas. 2009. "On Crop Biodiversity, Risk Exposure, and Food Security in the Highlands of Ethiopia." American Journal of Agricultural Economics 91 (3): 599–611. https://doi.org/10.1111/j.1467-8276.2009.01265.x.
Di Falco, Salvatore, Mintewab Bezabih, and Mahmud Yesuf. 2010. "Seeds for Livelihood: Crop Biodiversity and Food Production in Ethiopia." Ecological Economics 69 (8): 1695–1702. https://doi.org/10.1016/j.ecolecon.2010.03.024.
Di Falco, Salvatore, and Elisavet Zoupanidou. 2016. "Soil Fertility, Crop Biodiversity, and Farmers' Revenues: Evidence From Italy." AMBIO 46 (2): 162–72. https://doi.org/10.1007/s13280-016-0812-7.
Donfouet, Hermann Pythagore Pierre, Aleksandra Barczak, Cécile Détang-Dessendre, and Elise Maigné. 2017. "Crop Production and Crop Diversity in France: A Spatial Analysis." Ecological Economics 134 (January): 29–39. https://doi.org/10.1016/j.ecolecon.2016.11.016.
Dillon, J. The Analysis of Response in Crop and Livestock Production. Oxford: Pergamon Press, 1979
Du, Xiaodong, David A. Hennessy, and Cindy L. Yu. 2012. "Testing Day's Conjecture That More Nitrogen Decreases Crop Yield Skewness." American Journal of Agricultural Economics 94 (1): 225–37. https://www.jstor.org/stable/41331238.
Evans, L. T. 1998. *Feeding the Ten Billion: Plants and Population Growth.* Cambridge: Cambridge University Press.
Fafchamps, Marcel. 1992. "Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World." American Journal of Agricultural Economics 74 (1): 90–99. https://doi.org/10.2307/1242993.
FAO. 2015**.** *Social Protection and Agriculture: Breaking the Cycle of Rural Poverty.* Rome: FAO. https://www.fao.org/3/i4742e/i4742e.pdf
FAO. 2016**.** *The State of Food and Agriculture: Climate Change, Agriculture, and Food Security.* Rome, Italy: FAO. https://www.fao.org/3/a-i6030e.pdf
FAO. 2022. The State of the World's Forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies. Rome, FAO. Accessed here.
FAO and DWFI. 2015. *Yield Gap Analysis of Field Crops: Methods and Case Studies.* By V. O. Sadras, K. G. G. Cassman, P. Grassini, A. J. Hall, W. G. M. Bastiaanssen, A. G. Laborte, A. E. Milne, G. Sileshi, and P. Steduto. FAO Water Reports, No. 41. Rome, Italy: FAO. https://www.fao.org/3/i4953e/i4953e.pdf
Ferrari, S., and F. Cribari-Neto. 2004**.** "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31 (7): 799–815.
Fischer, Tony, Derek Byerlee, and Greg Edmeades. 2014**.** *Crop Yields and Global Food Security: Will Yield Increase Continue to Feed the World?* ACIAR Monograph No. 158. Canberra, Australia: Australian Centre for International Agricultural Research (ACIAR). https://www.aciar.gov
Fisher, E., J. Hellin, H. Greatrex, and N. Jensen. 2019**.** "Index Insurance and Climate Risk Management: Addressing Social Equity." *Development Policy Review* 37 (5): 581–602.
Franzel, Steven, and Richard Coe. 2002. "Participatory On-Farm Technology Testing: The Suitability of Different Types of Trials for Different Objectives." In *Quantitative Analysis of Data from Participatory Methods in Plant Breeding*, 1–8. Mexico: International Maize and Wheat Improvement Center (CIMMYT).
Fujimoto, Takefumi, and Aya Suzuki. 2024. "Different Strategies of Crop Diversification Between Poor and Non-poor Farmers: Concepts and Evidence From Tanzania."

Ecological Economics 227 (September): 108369. https://doi.org/10.1016/j.ecolecon.2024.108369.
Gallagher, P. 1986. "U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Nonsymmetric Disturbances." *Journal of Agricultural Economics* 8: 109–122.
Gallagher, P. 1987**.** "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Distributions." *American Journal of Agricultural Economics* 69: 796–803.
Garcia-Franco, Noelia, Eleanor Hobley, Rico Hübner, and Martin Wiesmeier. 2018. "Chapter 23: Climate-Smart Soil Management in Semiarid Regions." In *Soil Management and Climate Change*, 349–368. Academic Press: Elsevier
George, Thomas. 2014. "Why Crop Yields in Developing Countries Have Not Kept Pace with Advances in Agronomy." *Global Food Security* 3 (1): 49–58.
Grilli, Gianluca, Francesco Pagliacci, and Paola Gatto. 2024. "Determinants of Agricultural Diversification: What Really Matters? A Review." Journal of Rural Studies 110 (July): 103365. https://doi.org/10.1016/j.jrurstud.2024.103365.
Goodwin, Barry K., and Alan P. Ker. 1998**.** "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80 (1): 139–153. https://doi.org/10.2307/3180276.
Harri, Ardian, Cumhur Erdem, Keith H. Coble, and Thomas O. Knight. 2009**.** "Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality." *Review of Agricultural Economics* 31 (1): 163–182. https://www.jstor.org/stable/30224852.
Himmelstein, J., A. Ares, D. Gallagher, and J. Myers. 2016. "A Meta-analysis of Intercropping in Africa: Impacts on Crop Yield, Farmer Income, and Integrated Pest Management Effects." International Journal of Agricultural Sustainability 15 (1): 1–10. https://doi.org/10.1080/14735903.2016.1242332.
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.
Holling, C. S. 1973. "Resilience and Stability of Ecological Systems." Annual Review of Ecology and Systematics 4: 1–23. https://www.jstor.org/stable/2096802.
Just, R., and R. Pope. 1979**.** "Production Function Estimation and Related Risk Considerations." *American Journal of Agricultural Economics* 61: 276–284.
Just, R. E., and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *American Journal of Agricultural Economics* 81 (2): 287–304.
Just, R. E., and W. Candler. 1985. "Production Functions and Rationality of Mixed Cropping." European Review of Agricultural Economics 12 (3): 207–31. https://doi.org/10.1093/erae/12.3.207.
Ker, Alan P., and Keith Coble. 2003. "Modeling Conditional Yield Densities." *American Journal of Agricultural Economics* 85 (2): 291–304. https://doi.org/10.1111/1467-8276.00120.
Ker, Alan P, and Tor N Tolhurst. 2019. "On The Treatment of Heteroscedasticity in Crop Yield Data." *American Journal of Agricultural Economics* 101 (4): 1247–61. https://doi.org/10.1093/ajae/aaz004.
Khanal, Aditya R., and Ashok K. Mishra. 2017. "Enhancing Food Security: Food Crop Portfolio Choice in Response to Climatic Risk in India." Global Food Security 12 (December): 22–30. https://doi.org/10.1016/j.gfs.2016.12.003.
Kravchenko, Alexandra N., Sieglinde S. Snapp, and G. Philip Robertson. 2017. "Field-scale Experiments Reveal Persistent Yield Gaps in Low-input and Organic Cropping Systems."

Proceedings of the National Academy of Sciences 114 (5): 926–31. https://doi.org/10.1073/pnas.1612311114.

Krishnamurthy, L., M. Zaman-Allah, R. Purushothaman, M. Irshad, and V. Vadez. 2011. “Plant Biomass Productivity Under Abiotic Stresses in SAT Agriculture.” In *InTech eBooks*. https://doi.org/10.5772/17279.

Laajaj, Rachid, Karen Macours, Cargele Masso, Moses Thuita, and Bernard Vanlauwe. 2020. “Reconciling Yield Gains in Agronomic Trials With Returns Under African Smallholder Conditions.” *Scientific Reports* 10 (1): 14286. https://doi.org/10.1038/s41598-020-71155-y.

Leff, Billie, Navin Ramankutty, and Jonathan A. Foley. 2004. “Geographic Distribution of Major Crops Across the World.” *Global Biogeochemical Cycles* 18 (1). https://doi.org/10.1029/2003gb002108.

Lehman, Clarence L., and David Tilman. 2000. “Biodiversity, Stability, and Productivity in Competitive Communities.” The American Naturalist 156 (5): 534–52. https://doi.org/10.1086/303402.

Lemken, Dominic, Achim Spiller, and Marie Von Meyer-Höfer. 2017. “The Case of Legume-Cereal Crop Mixtures in Modern Agriculture and the Transtheoretical Model of Gradual Adoption.” Ecological Economics 137 (March): 20–28. https://doi.org/10.1016/j.ecolecon.2017.02.021.

Lin, Brenda B. 2011. “Resilience in Agriculture Through Crop Diversification: Adaptive Management for Environmental Change.” BioScience 61 (3): 183–93. https://doi.org/10.1525/bio.2011.61.3.4

Maggio, Giuseppe, and Nicholas J. Sitko. 2020. “Diversification Is in the Detail: Accounting for Crop System Heterogeneity to Inform Diversification Policies in Malawi and Zambia.” The Journal of Development Studies 57 (2): 264–88. https://doi.org/10.1080/00220388.2020.1769072.

Malthus T., 1798. An Essay on the Principle of Population. London (Electronic Scholarly Publishing Project), Accessed here.

Mani A, Mullainathan S, Shafir E and Zhao J, 2013. Poverty impedes cognitive function. Science 341: 976–80

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. Microeconomic Theory. New York: Oxford University Press.

Matiu, Michael, Donna P. Ankerst, and Annette Menzel. 2017. “Interactions Between Temperature and Drought in Global and Regional Crop Yield Variability During 1961-2014.” *PLoS ONE* 12 (5): e0178339. https://doi.org/10.1371/journal.pone.0178339.

McArthur, John W., and Gordon C. McCord. 2017. “Fertilizing Growth: Agricultural Inputs and Their Effects in Economic Development.” *Journal of Development Economics* 127 (March): 133–52. https://doi.org/10.1016/j.jdeveco.2017.02.007.

McDonald, James B., and Yexiao J. Xu. 1995. “A Generalization of the Beta Distribution With Applications.” *Journal of Econometrics* 66 (1–2): 133–52. https://doi.org/10.1016/0304-4076(94)01612-4.

Michler, Jeffrey D., and Anna L. Josephson. 2017. “To Specialize or Diversify: Agricultural Diversity and Poverty Dynamics in Ethiopia.” World Development 89 (September): 214–26. https://doi.org/10.1016/j.worlddev.2016.08.011.

Moschini, Giancarlo, and David A. Hennessy. 2001. "Chapter 2: Uncertainty, Risk Aversion, and Risk Management for Agricultural Producers." In Handbook of Agricultural Economics, Volume 1, by Bruce L. Gardner and Gordon C. Rausser, 88-144. Netherlands: North-Holland Publishing.

Moss, Charles B., and J. S. Shonkwiler. 1993. "Estimating Yield Distributions With a Stochastic Trend and Nonnormal Errors." *American Journal of Agricultural Economics* 75 (4): 1056–62. https://doi.org/10.2307/1243993.

Norwood, Bailey, Matthew C. Roberts, and Jayson L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86 (4): 1032–43. https://www.jstor.org/stable/4492790.

Nelson, Carl H., and Paul V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71 (2): 370–78. https://doi.org/10.2307/1241595.

Nelson, Carl H. 1990. "The Influence of Distributional Assumptions on the Calculation of Crop Insurance Premia." *Applied Economic Perspectives and Policy* 12 (1): 71–78. https://doi.org/10.1093/aepp/12.1.71.

Paddock W. and Paddock P. 1968. Famine–1975! Weidenfeld and Nicolson: London

Racine, Jeff, and Alan Ker. 2006. "Rating Crop Insurance Policies With Efficient Nonparametric Estimators That Admit Mixed Data Types." *Journal of Agricultural and Resource Economics* 31 (1): 27–39. https://www.jstor.org/stable/40987304.

Ramirez, O.A. 1997. Estimation and use of a Multi- variate Parameteric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean and Wheat Yields. American Journal of Agricultural Economics. 79:191-205

Ramirez, Octavio A., Sukant Misra, and James Field. 2003. "Crop-Yield Distributions Revisited." American Journal of Agricultural Economics 85 (1): 108–20. https://www.jstor.org/stable/1244930.

Raseduzzaman, Md., and Erik Steen Jensen. 2017. "Does Intercropping Enhance Yield Stability in Arable Crop Production? A Meta-analysis." European Journal of Agronomy 91 (September): 25–33. https://doi.org/10.1016/j.eja.2017.09.009.

Ravisankar, Natesan, Meraj A. Ansari, Mohammad Shamim, Ashisa K. Prusty, Raghuveer Singh, Azad S. Pan, Debashis Dutta, et al. 2022. "Sustainable livelihood security of small farmers improved through a resilient farming system in the semiarid region of India." Land Degradation and Development 1-13.

Ray, Deepak K., James S. Gerber, Graham K. MacDonald, and Paul C. West. 2015. "Climate Variation Explains a Third of Global Crop Yield Variability." *Nature Communications* 6 (1): 5989. https://doi.org/10.1038/ncomms6989.

Raza, Muhammad Ali, Atta Mohi Ud Din, Hassan Shehryar Yasin, Hina Gul, Ghulam Abbas Shah, Wang Zhiqi, Muhammad Jawad Hassan, et al. 2025. "Maize/Legume Intercropping Increases Nutrient Uptake, Crop Yields, Land Productivity, and Economic Profits in Resource-intensive Arid-irrigated Areas." Journal of Soil Science and Plant Nutrition 25 (3): 8316–32. https://doi.org/10.1007/s42729-025-02675-7.

Schauberger, Bernhard, Tamara Ben-Ari, David Makowski, Tomomichi Kato, Hiromi Kato, and Philippe Ciais. 2018. "Yield Trends, Variability and Stagnation Analysis of Major Crops in France Over More Than a Century." *Scientific Reports* 8 (1): 16865. https://doi.org/10.1038/s41598-018-35351-1.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan. 2016. "The Psychological Lives of the Poor." *American Economic Review* 106 (5): 435–40. https://doi.org/10.1257/aer.p20161101.

Schmidhuber, Josef, and Francesco N. Tubiello. 2007. "Global Food Security Under Climate Change." *Proceedings of the National Academy of Sciences* 104 (50): 19703–8. https://doi.org/10.1073/pnas.0701976104.

Sherrick, Bruce J., Fabio C. Zanini, Gary D. Schnitkey, and Scott H. Irwin. 2004. "Crop Insurance Valuation Under Alternative Yield Distributions." American Journal of Agricultural Economics 86 (2): 406–19. https://doi.org/10.1111/j.0092-5853.2004.00587.x.

Singh, Hardeep, and Poonam Rani. 2025. "Adoption and Risk-reducing Effects of Mixed Cropping: Evidence From India." Applied Economic Perspectives and Policy 47 (3): 1119–38. https://doi.org/10.1002/aepp.13514.

Singh, R. P., and P. B. R. Hazell. 1993."Rural Poverty in the Semi-Arid Tropics of India- Identification, Determinants and Policy Interventions." 1993. Economic and Political Weekly. March 20, 1993. https://www.epw.in/journal/1993/12-13/review-agriculture-review-issues-specials/rural-poverty-semi-arid-tropics-india.

Smale, Melinda, Jason Hartell, Paul W. Heisey, and Ben Senauer. 1998. "The Contribution of Genetic Resources and Diversity to Wheat Production in the Punjab of Pakistan." American Journal of Agricultural Economics 80 (3): 482–93. https://doi.org/10.2307/1244551.

Smithson, Michael, and Jay Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-likelihood Regression With Beta-distributed Dependent Variables." *Psychological Methods* 11 (1): 54–71. https://doi.org/10.1037/1082-989x.11.1.54.

Stevanović, Miodrag, Alexander Popp, Hermann Lotze-Campen, Jan Philipp Dietrich, Christoph Müller, Markus Bonsch, Christoph Schmitz, Benjamin Leon Bodirsky, Florian Humpenöder, and Isabelle Weindl. 2016. "The Impact of High-end Climate Change on Agricultural Welfare." *Science Advances* 2 (8): e1501452. https://doi.org/10.1126/sciadv.1501452.

Taylor, C. Robert. 1990. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *American Journal of Agricultural Economics* 72 (1): 210–17. https://doi.org/10.2307/1243160.

Terza, Joseph V., Anirban Basu, and Paul J. Rathouz. 2007. "Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics* 27 (3): 531–43. https://doi.org/10.1016/j.jhealeco.2007.09.009.

Tesfaye, Wondimagegn, and Nyasha Tirivayi. 2019. "Crop Diversity, Household Welfare and Consumption Smoothing Under Risk: Evidence From Rural Uganda." World Development 125 (September): 104686. https://doi.org/10.1016/j.worlddev.2019.104686.

Tilman, David, Christian Balzer, Jason Hill, and Belinda L. Befort. 2011. "Global Food Demand and the Sustainable Intensification of Agriculture." *Proceedings of the National Academy of Sciences* 108 (50): 20260–64. https://doi.org/10.1073/pnas.1116437108.

Tilman, David, Stephen Polasky, and Clarence Lehman. 2005. "Diversity, Productivity and Temporal Stability in the Economies of Humans and Nature." Journal of Environmental Economics and Management 49 (3): 405–26. https://doi.org/10.1016/j.jeem.2004.03.008

Van Ittersum MK, Cassman K.; Grassini P.; Wolf J.; Tittonell P.; Hochman Z. 2013. Yield gap analysis with local to global relevance—A review. Field Crops Research 143:4–17.

Vogel, E. & Meyer, R., 2018. Chapter 3 - Climate Change, Climate Extremes, and Global Food Production—Adaptation in the Agricultural Sector. In: Resilience: The Science of Adaptation to Climate Change. United Kingdom: Elsevier, pp. 31-49.

Waha, Katharina, Jan Philipp Dietrich, Felix T. Portmann, Stefan Siebert, Philip K. Thornton, Alberte Bondeau, and Mario Herrero. 2020. "Multiple Cropping Systems of the World and the Potential for Increasing Cropping Intensity." Global Environmental Change 64 (August): 102131. https://doi.org/10.1016/j.gloenvcha.2020.102131.

Waldman, Kurt B, Peter M Todd, Shahera Omar, Jordan P Blekking, Stacey A Giroux, Shahzeen Z Attari, Kathy Baylis, and Tom P Evans. 2020. "Agricultural Decision Making and Climate Uncertainty in Developing Countries." *Environmental Research Letters* 15 (11): 113004. https://doi.org/10.1088/1748-9326/abb909.

World Bank, 2010. World Development Report 2010: Development and Climate Change, Washington, DC: The World Bank. Accessed here.

Yin, Xiaogang, Zhenwei Song, Shanheng Shi, Zhiyuan Bai, Yulin Jiang, Axiang Zheng, Wenhai Huang, Nana Chen, and Fu Chen. 2024. "Developments and Prospects of Multiple Cropping in China." Farming System 2 (2): 100083. https://doi.org/10.1016/j.farsys.2024.100083.

You, Liangzhi, Stanley Wood, and Ulrike Wood-Sichra. 2009. "Generating Plausible Crop Distribution Maps for Sub-Saharan Africa Using a Spatially Disaggregated Data Fusion and Optimization Approach." *Agricultural Systems* 99 (2–3): 126–40. https://doi.org/10.1016/j.agsy.2008.11.003.

Zheng, Qiujie, H. Holly Wang, and Qing Hua Shi. 2014. "Estimating Bivariate Yield Distributions and Crop Insurance Premiums Using Nonparametric Methods." *Applied Economics* 46 (18): 2108–18. https://doi.org/10.1080/00036846.2014.894630.

3.11 Supplementary Information

Reference	Year	Geography	Unit of Analysis; Data Type	Crops	Inputs	Methodological Approach	Conclusion about Shape of Distribution
(Day 1965)	1928-1957	Mississippi, USA	Plot-Level; Experimental	Cotton, Corn, Oats	Nitrogen	Method of Moments, Parametric Approach | Pearson Distribution Family	Crop yields are generally non-normal. Beta distribution emerges as the best fit. Nitrogen reduces skewness and kurtosis.
(Just and Pope 1979)	1928-1957	Mississippi, USA	Plot-Level; Experimental	Corn, Oats	Nitrogen	Parametric Approach | Model mean and variance as function of inputs. Normal/ Log-Normal Distribution	Mean and variance of yields are specified independently as functions of nitrogen. Nitrogen has a yield-variance increasing effect.
(Antle 1983)	1981 (30 months)	California, USA	Dairy-Level; Experimental	Milk	Feed Input	Non-Parametric, Flexible-Input-Conditioned Moment Approach (count based on Stieltjes Condition).	Moments significantly driven by inputs. Cross-moment restrictions implied by standard models (e.g., Just and Pope (1978)) rejected.
(Antle and Goodger 1984)	1981 (30 months)	California, USA	Dairy-Level; Experimental	Milk	Feed Input	As in Antle (1983).	Suggests moments (including higher order i.e., >2) are function of inputs. Relevant properties of the distribution must be specified.
(Gallagher 1987)	1941-84	USA	USA, Observational	Soybean	N.A.	Parametric Approach | Gamma Distribution	Find evidence of positive skew.
(Nelson and Preckel 1989)	1961-1970	Iowa, USA	Plot-Level; Experimental	Corn	Nitrogen	Parametric Approach | Conditional Beta Distribution	Find negatively skewed yields. Nitrogen influences moments significantly.
(Nelson 1990)	1964-69	Iowa, USA	County-Level Averages based on Experimental Farms	Corn	N.A.	Parametric Approach | Normal and Beta Distributions	Argue that central limit theorem cannot justify normality for crop premia calculation since loss events are not independent. Use of normality overstates yield loss probability relative to Beta.

Reference	Year	Geography	Unit of Analysis; Data Type	Crops	Inputs	Methodological Approach	Conclusion about Shape of Distribution
(Taylor 1990)	1945-87	Illinois, USA	County-Level; Observational	Corn, Soybean, Wheat	N.A.	Parametric, Multivariate Non-Normal	Find evidence for skewness and kurtosis.
(Moss and Shonkweiler 1993)	1930-1990	USA	USA; Observational	Corn	N.A.	Parametric; Test for Normality	Find and correct skewness and kurtosis using an inverse hyperbolic sine distribution.
(Babcock and Hennessey 1996)	1986-1991	Iowa, USA	Farm-Level; Experimental	Corn	Nitrogen	Parametric, Beta Distribution	Evidence of yield skewness driven (in part) by nitrogen input.
(Ramirez 1997)	1950-89	Corn Belt, USA	Corn Belt; Observational	Corn, Soybean, Wheat	N.A.	Parametric, Hyperbolic Sine Distribution	Evidence of left skew for corn and soybean while wheat yields are found to be normally distributed.
(Goodwin and Ker 1998)	1962-1994	USA	County; Observational	Multiple Crops.	N.A.	Non-Parametric Approach	Negatively skewed yields across crops. Find evidence of bimodality which may be missed by parametric approaches.
(Just and Weninger 1999)	Subsets between 1926-94 across crops.	USA	Farm-Level, County-Level; Observational	Corn, Sorghum, Soybean, Wheat, Alfalfa	N.A.	Parametric Approach, Test for Normality	Do not reject normality in farm or county level yield data. Propose a central limit theorem-based explanation.
(Ramirez et al. 2003)	Subsets between 1970-1999 by spatial granularity.	USA	County-Level; Observational | Farm-Level; Experimental	Corn, Soybean, Cotton, Wheat	N.A.	Parametric Approach, Normal Distribution	Find that some aggregate and farm level distributions are non-normal, kurtotic and left/right skewed depending on circumstances.
(Atwood et al. 2003)	1990-1999	Kansas, USA	Farm-Level; Observational	Multiple Crops.	N.A.	Parametric Approach; Test for Normality	Normality of yields is rejected.

Reference	Year	Geography	Unit of Analysis; Data Type	Crops	Inputs	Methodological Approach	Conclusion about Shape of Distribution
(Norwood et al. 2004)	Subsets between 1970-1999 by spatial granularity.	USA	County-Level; Observational | Farm-Level; Experimental	Corn, Soybean, Wheat	N.A.	Parametric and Non-Parametric Approaches compared on out-of-sample fit. | Multiple Distributions	Semiparametric performs better. Find evidence for skewness and kurtosis.
(Sherrick et al. 2004)	1972-1999	Illinoi, USA	Farm; Experimental	Corn, Soybean	N.A.	Parametric Approach; Test of Multiple Distributions	Reject normality. Weibull distribution performs best for corn. Beta distribution provides best fit across crops.
(Racine and Ker 2006)	1962-19994	Illinoi, USA	County USA	Corn.	N.A.	Non-Parametric Approach to jointly model across-county yields.	Method performed better than standard non-parametric approaches.
(Atwood et al. 2007)	1988-97	USA	Farm-Level pooled to State and Region-Level	Multiple Crops.	N.A.	Parametric Approach; Test for Normality	Find consistent non-normality of crop yields.
(Ozaki et al. 2008)	1990-2002	Brazil	County level	Soybean, Corn, Wheat	N.A.	Parametric Approach: Normal and Beta Distributions; Non-Parametric Approach	Find evidence of positive skew and bimodality.
(Harri et al. 2009)	Subsets between 1965-2005 across crops.	USA	County level	Corn, Cotton, Soybean, Wheat	N.A.	Parametric models for trends | Test for Normality	Find negative skew in corn, soybean and wheat yields and positive skew in cotton yields with regional variation within counties cultivating the same crop.
(Antle 2010)	1980-92	Ecuador	Farm-Level; Observational	Potato	Nitrogen, Fungicide	Semi-Parametric (Partial Moments Based) Approach	Fertilizer use likely reduces skewness while fungicide and labour use likely increase skewness.

Reference	Year	Geography	Unit of Analysis; Data Type	Crops	Inputs	Methodological Approach	Conclusion about Shape of Distribution
(Ramirez et al. 2010)	1959-2003	Illinoi, USA	Farm-Level, Experimental	Corn	N.A.	Parametric Approach; Introduce a system of distributions that can span the *entire* mean-variance-skewness-kurtosis space.	Find evidence of yield skewness. Beta sub-family of distributions fit well.
(Lanou et al. 2010)	1972-2008	Illinoi, USA	Farm level; Observational	Corn	N.A.	Semi-Parametric (Weibull, Gamma, Normal, Beta), Non-Parametric	Find evidence for skewness. Compare parametric and non-parametric approaches on the bias-efficiency tradeoff.
(Claassen and Just 2011)	Subsets between 1983-2008 by spatial granularity.	USA	Farm level; County-Level Observational	Wheat and Corn	N.A.	Parametric – Reverse Log Normal	Find support for reverse lognormal, reverse gamma, beta and normal distributions across risk specifications.
(Du et al. 2012)	Subsets between 1973-2003 across crops.	USA	Farm-level Experimental Data	Corn, Cotton	Nitrogen	Parametric Approach (Beta Distribution); Quantile Regression	Find evidence of negative (positive) skewness for corn (cotton) yields. Weak evidence of Day's (1965) conjecture.
(Tack et al. 2012)	1972-2005	USA	County-Level; Observational	Cotton	Climate, Irrigation	Non-Parametric Approach	Climate and irrigation affect yield skewness significantly.
(Zheng et al. 2014)	1995-2006	China	Farm-level, Observational	Wheat, Corn	N.A.	Bi-variate Non-Parametric Approach	Find evidence of multimodal distributions across crops.
(Ker and Tolhurst 2019)	1951-2017	USA	County-Level; Observational	Corn, Soybean, Wheat	N.A.	Parametric Approach for modelling heteroskedasticity.	Find evidence of asymmetric heteroskedasticity (increasing volatility in the lower tail) in corn and soybean but not wheat.

Table 1(S1): Key Literature on Crop Yield Density Estimation; **Source:** Author's Compilation

Notes: N.A. refers to Not Applicable since the analysis is typically time-series without a focus on specific inputs.

References (SI, S1):

Antle, John M. 1983. “Testing the Stochastic Structure of Production: A Flexible Moment-Based Approach.” *Journal of Business and Economic Statistics* 1 (3): 192. https://doi.org/10.2307/1391337.

Antle, J. M., and W. J. Goodger. 1984. “Measuring Stochastic Technology: The Case of Tulare Milk Production.” *American Journal of Agricultural Economics* 66 (3): 342–50. https://doi.org/10.2307/1240801.

Antle, John M. 2010. “Asymmetry, Partial Moments, and Production Risk.” *American Journal of Agricultural Economics* 92 (5): 1294–1309. https://doi.org/10.1093/ajae/aaq077.

Atwood, Joseph, Saleem Shaik, and Myles Watts. 2003. “Are Crop Yields Normally Distributed? A Reexamination.” *American Journal of Agricultural Economics* 85 (4): 888–901. https://www.jstor.org/stable/1244771.

Atwood, Joseph, Saleem Shaik, and Myles Watts. 2002. “Can Normality of Yields Be Assumed for Crop Insurance?” *Canadian Journal of Agricultural Economics/Revue Canadienne D Agroeconomie* 50 (2): 171–84. https://doi.org/10.1111/j.1744-7976.2002.tb00426.x.

Babcock, B. A. & Hennessy, D. A., 1996. Input Demand under Yield and Revenue Insurance. American Journal of Agricultural Economics, 78(2), pp. 416-427.

Claassen, Roger, and Richard E. Just. 2011. “Heterogeneity and Distributional Form of Farm-Level Yields.” *American Journal of Agricultural Economics* 93 (1): 144–60. https://www.jstor.org/stable/41240265.

Day, Richard H. 1965. “Probability Distributions of Field Crop Yields.” *Journal of Farm Economics* 47 (3): 713. https://doi.org/10.2307/1236284.

Du, Xiaodong, David A. Hennessy, and Cindy L. Yu. 2012. “Testing Day’s Conjecture That More Nitrogen Decreases Crop Yield Skewness.” American Journal of Agricultural Economics 94 (1): 225–37. https://www.jstor.org/stable/41331238.

Gallagher, Paul. 1987. “U.S. Soybean Yields: Estimation And Forecasting With Nonsymmetric Disturbances.” *American Journal of Agricultural Economics* 69 (4): 796–803. https://doi.org/10.2307/1242190.

Goodwin, Barry K., and Alan P. Ker. 1998. “Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts.” *American Journal of Agricultural Economics* 80 (1): 139–53. https://doi.org/10.2307/3180276.

Harri, Ardian, Cumhur Erdem, Keith H. Coble, and Thomas O. Knight. 2009**.** “Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality.” *Review of Agricultural Economics* 31 (1): 163–182. https://www.jstor.org/stable/30224852.

Just, Richard E., and Rulon D. Pope. 1979. “Production Function Estimation and Related Risk Considerations.” *American Journal of Agricultural Economics* 61 (2): 276–84. https://doi.org/10.2307/1239732.

Just, Richard E., and Quinn Weninger. 1999. “Are Crop Yields Normally Distributed?” *American Journal of Agricultural Economics* 81 (2): 287–304. https://doi.org/10.2307/1244582.

Ker, Alan P, and Tor N Tolhurst. 2019. “On The Treatment of Heteroscedasticity in Crop Yield Data.” American Journal of Agricultural Economics 101 (4): 1247–61. https://doi.org/10.1093/ajae/aaz004.

Lanoue, Christopher & Sherrick Bruce J. & Woodard, Joshua D. & Paulson, Nicholas D. 2010. "Evaluating Yield Models for Crop Insurance Rating." *Ideas.Repec.Org*. https://ideas.repec.org/p/ags/aaea10/61761.html.

Moss, Charles B., and J. S. Shonkwiler. 1993. "Estimating Yield Distributions With a Stochastic Trend and Nonnormal Errors." *American Journal of Agricultural Economics* 75 (4): 1056–62. https://doi.org/10.2307/1243993.

Nelson, Carl H., and Paul V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71 (2): 370–78. https://doi.org/10.2307/1241595.

Nelson, Carl H. 1990. "The Influence of Distributional Assumptions on the Calculation of Crop Insurance Premia." *Applied Economic Perspectives and Policy* 12 (1): 71–78. https://doi.org/10.1093/aepp/12.1.71.

Norwood, Bailey, Matthew C. Roberts, and Jayson L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86 (4): 1032–43. https://www.jstor.org/stable/4492790.

Ozaki, Vitor A., Barry K. Goodwin, and Ricardo Shirota. 2008. "Parametric and Nonparametric Statistical Modelling of Crop Yield: Implications for Pricing Crop Insurance Contracts." *Applied Economics* 40 (9): 1151–64. https://doi.org/10.1080/00036840600749680.

Racine, Jeff, and Alan Ker. 2006. "Rating Crop Insurance Policies With Efficient Nonparametric Estimators That Admit Mixed Data Types." *Journal of Agricultural and Resource Economics* 31 (1): 27–39. https://www.jstor.org/stable/40987304.

Ramirez, O.A. 1997. Estimation and use of a Multi- variate Parameteric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean and Wheat Yields. American Journal of Agricultural Economics. 79:191-205

Ramirez, Octavio A., Tanya U. McDonald, and Carlos E. Carpio. 2010. "A Flexible Parametric Family for the Modeling and Simulation of Yield Distributions." *Journal of Agricultural and Applied Economics* 42 (2): 303–19. https://doi.org/10.1017/s1074070800003473.

Ramirez, Octavio A., Sukant Misra, and James Field. 2003. "Crop-Yield Distributions Revisited." American Journal of Agricultural Economics 85 (1): 108–20. https://www.jstor.org/stable/1244930.

Sherrick, Bruce J., Fabio C. Zanini, Gary D. Schnitkey, and Scott H. Irwin. 2004. "Crop Insurance Valuation Under Alternative Yield Distributions." *American Journal of Agricultural Economics* 86 (2): 406–19. https://www.jstor.org/stable/30139564.

Tack, Jesse, Ardian Harri, and Keith Coble. 2012. "More Than Mean Effects: Modeling the Effect of Climate on the Higher Order Moments of Crop Yields." *American Journal of Agricultural Economics* 94 (5): 1037–54. https://doi.org/10.1093/ajae/aas071.

Taylor, C. Robert. 1990. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *American Journal of Agricultural Economics* 72 (1): 210–17. https://doi.org/10.2307/1243160.

Zheng, Qiujie, H. Holly Wang, and Qing Hua Shi. 2014. "Estimating Bivariate Yield Distributions and Crop Insurance Premiums Using Nonparametric Methods." Applied Economics 46 (18): 2108–18. https://doi.org/10.1080/00036846.2014.894630.

Model	Endogenous Variable	F-Statistic	Inference
Paddy (Kharif) Yields	Nitrogen	48.72	Reject the null.
	Phosphorus	482.75	Reject the null.
	Electric Irrigation	44.41	Reject the null.
Paddy (Rabi) Yields	Nitrogen	101.8	Reject the null.
	Phosphorus	18.78	Fail to reject the null.
	Electric Irrigation	336	Reject the null.
Wheat (Rabi) Yields	Nitrogen	53.32	Reject the null.
	Phosphorus	60.16	Reject the null.
	Irrigation	92.25	Reject the null.
Cotton Yields	Nitrogen	78	Reject the null.
	Phosphorus	48	Reject the null.
	Electric Irrigation	51	Reject the null.
	Mixed-Cropping	61	Reject the null.
Pigeonpea Yields	Nitrogen	100	Reject the null.
	Phosphorus	82	Reject the null.
	Electric Irrigation	80	Reject the null.
	Mixed-Cropping	63	Reject the null.

Table 2(S1): Results for instrument relevance. Null Hypothesis: Instruments are weak.

Model	p-value	Inference
Paddy (Kharif)	0.15	Fail to reject the null.
Paddy (Rabi)	15.75	Reject the null.
Wheat (Rabi)	0.51	Fail to reject the null.
Cotton Yields	0.37	Fail to reject the null.
Pigeonpea Yields	0.84	Fail to reject the null.

Table 3(S1): Results for Sargan's Test. Null Hypothesis: Instruments are correctly excluded from the yield model.

Chapter 4: Plot-level land quality data pertaining to semi-arid tropics in India: A construction using the Village Dynamics Studies in South Asia Dataset

Sumedha Shukla

(IIIT-Delhi)

sumedhas@iiitd.ac.in

Gaurav Arora

(IIIT-Delhi)

gaurav@iiitd.ac.in

Anoushka Kumar

(Salesforce, AMTS)

anoushka.kumar@salesforce.com

Plot-level land quality data pertaining to semi-arid tropics in India: A construction using the Village Dynamics Studies in South Asia Dataset

Abstract

The semi-arid tropical (SAT) regions, home to a majority of agricultural poor, are vulnerable to increasing uncertainty in agricultural incomes due to climate change. Microdata aimed at understanding plot-level seasonal agricultural operations in these regions is critical for effective agricultural policy design. The Village Dynamics in South Asia (VDSA) program provides a unique, plot-level, panel dataset on farming operations for 1689 farming households located across thirty Indian SAT villages for the period 1975-2014. Structured across twelve modules, the VDSA records household endowments, landholding status, non-farm activities, consumption, transactions and cropping patterns, and is widely regarded as a global public resource. However, the landholding module which contains critical information on household's biophysical constraints, specifically plot-level soil descriptors such as depth, slope, texture, colour, and quality, suffers from inconsistencies in plot identifiers due to data compilation errors. An attempt to merge the plot-level information on cropping patterns and landholding quality leads to a loss of 52% of data across all major crops. We design and implement an algorithmic merging procedure followed by a manual merging criterion based on phonetic similarities and vernacular identifiers to address this issue, which reduces the data losses by 50 percentage points. Our algorithm entails a general contribution to the extensive literature using the VDSA microdata to study farm household decisions in the SAT regions.

Keywords: ICRISAT-VDSA micro-level dataset, data construction algorithm, semi-arid regions

JEL Codes: Q10, Q12, Y1

Plot-level land quality data pertaining to semi-arid tropics in India: A construction using the Village Dynamics Studies in South Asia Dataset

4.1 Introduction

"In the history of modern development economics, no single data set has yielded as many important microeconomic papers as ICRISAT Village Level Studies." (Morduch 2002, p.1)

The semi-arid regions accounting for 30 percent of the world's geographical area are home to a majority of the world's rural-poor and are considered "hot-spots" for climate change (Singh and Hazell 1993; Evans 1998; Sivakumar et al. 2005; Krishnamurthy, et al. 2011; Ahlstorm et al. 2015; Garcia-Franco, et al. 2018; Ravisankar et al. 2022). The regions are characterized by the natural prevalence of abiotic stressors (e.g., heat) resulting in diversified cropping patterns featuring multi-cropping and inter-cropping (Isom & Worker 1979). In India, they account for 37% of the cropping area and are characterized by small landholdings and heavily diversified cropping to cope with the high production risks (Krishnamurthy et al. 2011).

Developing effective agricultural risk management strategies for the semi-arid regions requires an in-depth understanding of the farmers' decision-making processes including the underlying production environment and economic constraints in these regions. A comprehensive picture of the farmers' decision making environment relies critically on longitudinal microdata, i.e., data at the farm or household-level collected over time (Mullen 2016; Carletto 2021; Carletto et al. 2021; Food and Agricultural Organization (FAO) 2023). However the costly nature of such data collection exercises has proven to be prohibitive, especially for a majority of the developing world (Leff et al. 2004; You et al. 2009; Van Ittersum et al. 2013; FAO 2015; Carletto 2021; Ravisankar, et al. 2022). The Village Dynamics Studies in South Asia dataset program initiated by the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) provides a unique, open-source, plot-level, panel dataset on farming operations for 1689 farming households located across thirty SAT villages spanning eight Indian states (see Figure 3.1).

The VDSA program spans multiple waves between 1975-2014 and provides a comprehensive account of farming households' endowment status, landholding information, cropping patterns, consumption expenditure, transactions (i.e., savings and borrowings), nutrition, and non-farm activities. The dataset has been widely recognized as a global public good for providing valuable insights into farm-level management across the world (Badiani et al. 2007; Mullen 2016). With over 300 studies using this dataset (Figure 3.2 (Panel A)) cited 41519 times (as of 15th May, 2024, Figure 3.2 (Panel B)), the VDSA data are responsible for several stylized facts in the development economics literature, and the development of panel data methodologies that extend well beyond SAT regions (Jodha 1986; Battese & Coelli 1988; Walker & Ryan 1990; Battese & Coelli 1995; Chambers & Quiggin 2000; Badiani et al. 2007; Balagtas et al. 2013; Binswanger-Mkhize 2013; Michler & Balagtas 2017; Narayanan et al. 2023).

The VDSA dataset is organized across twelve modules, each containing a set of identifiers, some of which are common across all modules. Among these, there are three plot-level agricultural-production-specific modules: (a) crop output schedule, (b) crop input schedule, and (c) landholding schedule. However information losses during data compilation are such that the identifiers in the landholding module are inconsistent. Specifically, an attempt to merge the plot-level information on cropping patterns and land quality leads to a loss of 52 percent of data across all major crops. In this study, we design and implement an algorithmic merging procedure followed by a manual merging criterion based on phonetic similarities and vernacular identifiers to address this issue which reduces the data losses by 50 percentage points. We then present a summary of the dataset obtained and document the information gained from our algorithm. Given the criticality of the landholding module in the study of agricultural production and other household operations, we believe that the information gains from our algorithm entail a general contribution to the extensive literature using the VDSA microdata to study farm households in the SAT regions. In the next section, we will detail our data merging algorithm, explaining the steps and rationale behind the merging process.

4.2 Methods

In this section, we describe the general procedure followed for merging the landholding (hereafter LH) file with the crop output (hereafter CO) file.

Setup

Each household has an identifier (i.e., VDSID_HHID) which contains information on the village the household is located in. The village is the finest geographical unit whereby villages are located in districts which are a part of states across the country. Each household can operate multiple plots in the *kharif* (July-October), *rabi* (October-April), and *zaid* (May-June) seasons in each year. In each season, the household may cultivate multiple crops on the same plot. Importantly, for each household, the information on agricultural output - made available in the CO file - is recorded crop-wise at the plot-level in each season in a year. On the other hand, since each plot can be used to cultivate multiple crops, the crop input information (contained in the CI file) is recorded at the plot-level for each season in a year. Finally, the landholding information on farmer-reported soil type, soil quality, soil fertility, soil depth, land slope, ease of farming operations, and irrigation status on the plot is recorded at the plot-level for each household in a year.

In order to merge the information from the LH file with the other agricultural modules, we choose to merge the LH file with the CO file to obtain a revised CO file which also contains the landholding details for each plot. This is done to ensure that all information gained in this merging process can be assessed in terms of the crops grown. The revised CO file can further be integrated with the CI file as well as all the other VDSA modules as required. The plot identifiers in the CO file are the plot name and plot code while those in the LH file are the plot name, plot code and the plot serial number. Due to inconsistencies in the plot identifiers (described in detail later), a crude merge based on the raw plot identifiers for a household in a year yields a match for less than 50% of the data in the crop output file. To correct for this, as part of the algorithm, we broke down the merging process into six steps as described below.

Basic Pre-Processing

We started with the CO file for the period 2001-14 which had 43406 rows of data. We revised the crop name to account for spelling inconsistencies and local crop names as described in Table 4.1. This leads to a list of 127 unique crop names listed in Table 4.2. If the crop name field were empty, then based on the remarks field, we treat the plot as fallow. We then constructed a plot-level, crop-wise, identifier for each household for in a year, i.e., the VDSPCYID. It is likely that the CO file has duplicate entries by this VDSPCYID because a plot may be used to grow the same crop in two different seasons in the same year, or a plot may be used to grow different varieties of the same crop in the same year (and even season).

We are choosing this ID as our base identifier for the CO file because, in general, we would like to translate any data losses from this merging procedure at the crop level. For instance, if we are unable to find matches for 500 plots in the CO file, we would like to understand which crops were being grown on these plots. If there are significant data losses for a specific crop, it may be more appropriate to use alternative approaches for incorporating landholding information for that crop. Finally, we generate a plot-level identifier for each household in each year denoted as the VDSPYID. Obviously, the CO file will have multiple entries for each VDSPYID but the landholding information to be incorporated from the LH file will be the same for all these entries.

Next, we processed the landholding file for all the years from 1975-2014 which had 39910 rows of 106 data. First, we remove any rows where both plot code and plot name are missing since these cannot be effectively utilized for a plot-level merge. We then constructed a plot identifier for each household in a year (i.e., a VDSPYID) which corresponded to the plot code when available, and was substituted with plot name in the cases where plot code was missing.

Query Details

Query 1:

Based on this revised plot identifier, i.e., VDSPYID, we performed a many-to-one merge between the crop output and landholding files respectively which furnished matches for 36 percent of the data. Importantly, at each step of the merging algorithm, before merging entries from the CO file with the LH file, we removed any rows in the LH file which were duplicated based on the VDSPYID to avoid erroneous merges and always relied on a many-to-one merge between the two files. The information from these deleted, duplicated rows - if useful - was incorporated at the sixth step of the merging algorithm – i.e., the manual merging stage.

Query 2:

In the second step, we revised the VDSPYID generated in the first step to include only the first letter of plot code (in the crop output file) to account for the plot naming nomenclature followed in the VDSA data collection whereby a plot was divided into sub-plots or sub-sub-plots based on the cultivated crop(s) or irrigation status (Binswanger & Jodha 1978; see Figure 4.3). Since the sub-plots or sub-sub-plots are derived from the same main plot, we extracted the soil quality information for these subplots based on the available information for the main plot. A merge on the basis of this revised identifier yielded a match for 40.6 percent of the original data.

Query 3:

In the third step, we revised the plot identifier generated in the first step to remove any information on plot serial number which was rendered redundant because plot serial number information was only available in the landholding file and not the crop output file (VDSA 2013). Merging based on this revised identifier provided matches for 3.5 percent of the data.

Query 4:

In the fourth step, we accounted for cases where arbitrary symbols were used while concatenating plot codes with sub-plot codes or serial numbers by retaining only the initial portion of the plot identifier string. Based on this revised identifier, we merged 2.9 percent of the original data.

Query 5:

In the fifth step, we constructed an alternate plot identifier which corresponded to the plot name when available, and was substituted with plot code in cases where plot name was missing. This alternate plot identifier yielded merges for 0.5 percent of the data.

Query 6 (Manual Merging):

In the sixth step, for the remaining 16 percent of data, we undertook a manual merging exercise which furnished matches for 84 percent of this leftover data (i.e., 14 percent of the original data). As part of the manual merging, we were able to incorporate landholding information in cases where: (a) the plot names were spelled inconsistently across the two files but were phonetically similar (for example *chalka, chhalka*, and *chalaka* all refer to the same plot); (b) plot names were recorded with vernacular identifiers as part of the plot name in one of the files but not the other (for example, a piece of farm-land can be referred to as shet/sheth/khet in languages spoken in Maharashtra); (c) plot codes were same but recorded in the serial number column in the landholding file; and (d) serial number (in the landholding file) was same as the plot code recorded in the crop output file. Often, a combination of these conditions was used to obtain the match hence a manual merging process was deemed most suitable. We repeat this merging process for the CO file for the years 1975-1989 and are able to incorporate 99% of the data using the first two queries. For the rest of the data, we rely on manual merging and are able to merge all the entries in the CO file.

4.3 Results

Table 4.3 presents the results of the algorithmic merge procedure, specifically, the count of CO entries incorporated from each stage of the merging algorithm. In the last stage, i.e., the manual merging stage, we merged 2927 of the 3425 entries based on the criterion described in Table 4.4. Table 4.5 presents the results of the algorithmic merging procedure for the 1975-

89 CO dataset. All the manual merges for this period were based on the plot code. Finally, Figures 4.4, 4.5, and 4.6 compare the data gains from incorporating the landholding information at the plot-level using our algorithm with the data generated from a ‘crude merge’ based on the pre-existing plot identifiers (i.e., without using our algorithm) for three land quality characteristics. Evidently, a household summary based on the crudely merged dataset would likely result in significant measurement errors, which can be minimized by using our algorithm.

4.4 Conclusion

This study develops a systematic approach for resolving long-standing identifier inconsistencies in the VDSA micro data, central to development economics. By leveraging plot nomenclature and phonetic similarity, the algorithm reduces data loss from 52% to <2%. The enhanced information on land quality allows for richer analysis using the VDSA dataset.

4.5 References

Ahlström, Anders, Michael R. Raupach, Guy Schurgers, Benjamin Smith, Almut Arneth, Martin Jung, Markus Reichstein, et al. 2015. "The Dominant Role of Semi-arid Ecosystems in the Trend and Variability of the Land CO 2 Sink." *Science* 348 (6237): 895–99. https://doi.org/10.1126/science.aaa1668.

Badiani, Reena, Stefan Dercon, and Pramila Krishnan. 2007. "Changes in Living Standards in Villages in India 1975-2004: Revisiting the ICRISAT Village Level Studies." *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.1653343.

Balagtas, Joseph V., Humnath Bhandari, Ellanie R. Cabrera, Samarendu Mohanty, and Mahabub Hossain. 2013. "Did The Commodity Price Spike Increase Rural Poverty? Evidence From a Long-run Panel in Bangladesh." *Agricultural Economics* 45 (3): 303–12. https://doi.org/10.1111/agec.12066.

Battese, George E., and Tim J. Coelli. 1988. "Prediction of Firm-level Technical Efficiencies With a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38 (3): 387–99. https://doi.org/10.1016/0304-4076(88)90053-x.

Binswanger, M., Jodha, N.S. 1978. Manual of Instructions for Economic Investigators in ICRISAT's Village Level Studies. Accessed here.

Binswanger-Mkhize, H. Dealing with risks and uncertainty in rainfed agriculture. In Deb, U. (ed.) Raising hopes and nurturing options for agricultural development: Essays in honor of Cynthia Bantilan (ICRISAT, Patancheru 502 324, Andhra Pradesh, India, 2013).

Carletto, Calogero. 2021. "Better Data, Higher Impact: Improving Agricultural Data Systems for Societal Change." European Review of Agricultural Economics 48 (4): 719–40. https://doi.org/10.1093/erae/jbab030.

Carletto, Calogero, Andrew Dillon, and Alberto Zezza. 2021. "Agricultural Data Collection to Minimize Measurement Error and Maximize Coverage." In *Handbook of Agricultural Economics*, 4407–80. https://doi.org/10.1016/bs.hesagr.2021.10.008.

Chambers, RG & Quiggin, J 2000, *Uncertainty, Production, Choice and Agency: The State-Contingent Approach*. vol. 1, First edn, Cambridge University Press, New York

Evans, LT., 1998, *Feeding the Ten Billion Plants and Population Growth*, Cambridge University Press, Cambridge.

Food and Agricultural Organization. 2015. *Social Protection and Agriculture: Breaking the Cycle of Rural Poverty*. Rome: FAO.

FAO. 2023. "The State of Food and Agriculture 2023." 2023. https://openknowledge.fao.org/items/1516eb79-8b43-400e-b3cb-130fd70853b0.

Garcia-Franco, Noelia, Eleanor Hobley, Rico Hübner, and Martin Wiesmeier. 2018. "Climate-Smart Soil Management in Semi-arid Regions." In *Elsevier eBooks*, 349–68. https://doi.org/10.1016/b978-0-12-812128-3.00023-9.

Isom, W.H., Worker, G.F. (1979). Crop Management in Semi-Arid Environments. In: Hall, A.E., Cannell, G.H., Lawton, H.W. (eds) Agriculture in Semi-Arid Environments. Ecological Studies, vol 34. Springer, Berlin, Heidelberg.

Jat, R. A., S. P. Wani, and K. L. Sahrawat. "Conservation Agriculture in the Semi-Arid Tropics: Prospects and Problems." In *Advances in Agronomy*, edited by D. L. Sparks, Vol. 117, 191–273. Academic Press. https://doi.org/10.1016/B978-0-12-394278-4.00004-0

Jodha, N. S. 1986. "Common Property Resources and Rural Poor in Dry Regions in India." *Economic and Political Weekly* 21: 224.
Krishnamurthy, L., M. Zaman-Allah, R. Purushothaman, M. I. Ahmed, and V. Vadez. 2011. "Plant Biomass Productivity under Abiotic Stresses in SAT Agriculture." In *Biomass – Detection, Production and Usage*, edited by M. D. Matovic, 247–264. IntechOpen.
Leff, B., N. Ramankutty, and J. A. Foley. 2004. "Geographic Distribution of Major Crops across the World." *Global Biogeochemical Cycles* 18: 1–27.
Matiu, M., Ankerst, D. P. & Menzel, A. Interactions between temperature and drought in global and regional crop yield variability during 1961-2014. PLoS ONE 12 (2017).
Michler, J. D., and J. V. Balagtas. 2017. "The Importance of the Savings Device in Precautionary Savings: Empirical Evidence from Rural Bangladesh." *Agricultural Economics* 48: 129–141. https://doi.org/10.1111/agec.12320.
Mullen, J. D. 2016. *Impact Assessment of ICRISAT Village Level Studies: 1975 to 2013*. Technical report, independent external review commissioned by ICRISAT. Accessed here.
Morduch, J. 2002**.** *Consumption Smoothing across Space: Testing Theories of Risk-Sharing in the ICRISAT Study Region of South India*. WIDER Working Paper DP2002-55. World Institute for Development Economics Research (UNU-WIDER).
Narayanan, S., D. S. Negi, and T. Gupta. 2023. "Separability, Spillovers, and Segmented Markets: Evidence from Dairy in India." *Agricultural Economics* 54: 884–899. https://doi.org/10.1111/agec.12786
Ravisankar, Natesan, Meraj A. Ansari, Mohammad Shamim, Ashisa K. Prusty, Raghuveer Singh, Azad S. Panwar, Debashis Dutta, et al. 2022. "Sustainable Livelihood Security of Small Farmers Improved Through a Resilient Farming System in The semi arid region of India." *Land Degradation and Development* 33 (15): 2830–42. https://doi.org/10.1002/ldr.4358.
Singh, R. P., and P. B. R. Hazell. 1993. "Rural Poverty in the Semi-Arid Tropics of India: Identification, Determinants and Policy Interventions." *Economic and Political Weekly* 28: A9–A15
Sivakumar, M., H. Das, and O. Brunini. 2005. "Impacts of Present and Future Climate Variability and Change on Agriculture and Forestry in the Arid and Semi-Arid Tropics." *Climatic Change* 70: 31–72. https://doi.org/10.1007/s10584-005-5937-9
Van Ittersum, Martin K., Kenneth G. Cassman, Patricio Grassini, Joost Wolf, Pablo Tittonell, and Zvi Hochman. 2012. "Yield Gap Analysis With Local to Global relevance—A Review." *Field Crops Research* 143 (December): 4–17. https://doi.org/10.1016/j.fcr.2012.09.009.

You, Liangzhi, Stanley Wood, and Ulrike Wood-Sichra. 2009. "Generating Plausible Crop Distribution Maps for Sub-Saharan Africa Using a Spatially Disaggregated Data Fusion and Optimization Approach." *Agricultural Systems* 99 (2–3): 126–40. https://doi.org/10.1016/j.agsy.2008.11.003.

4.6 Figures

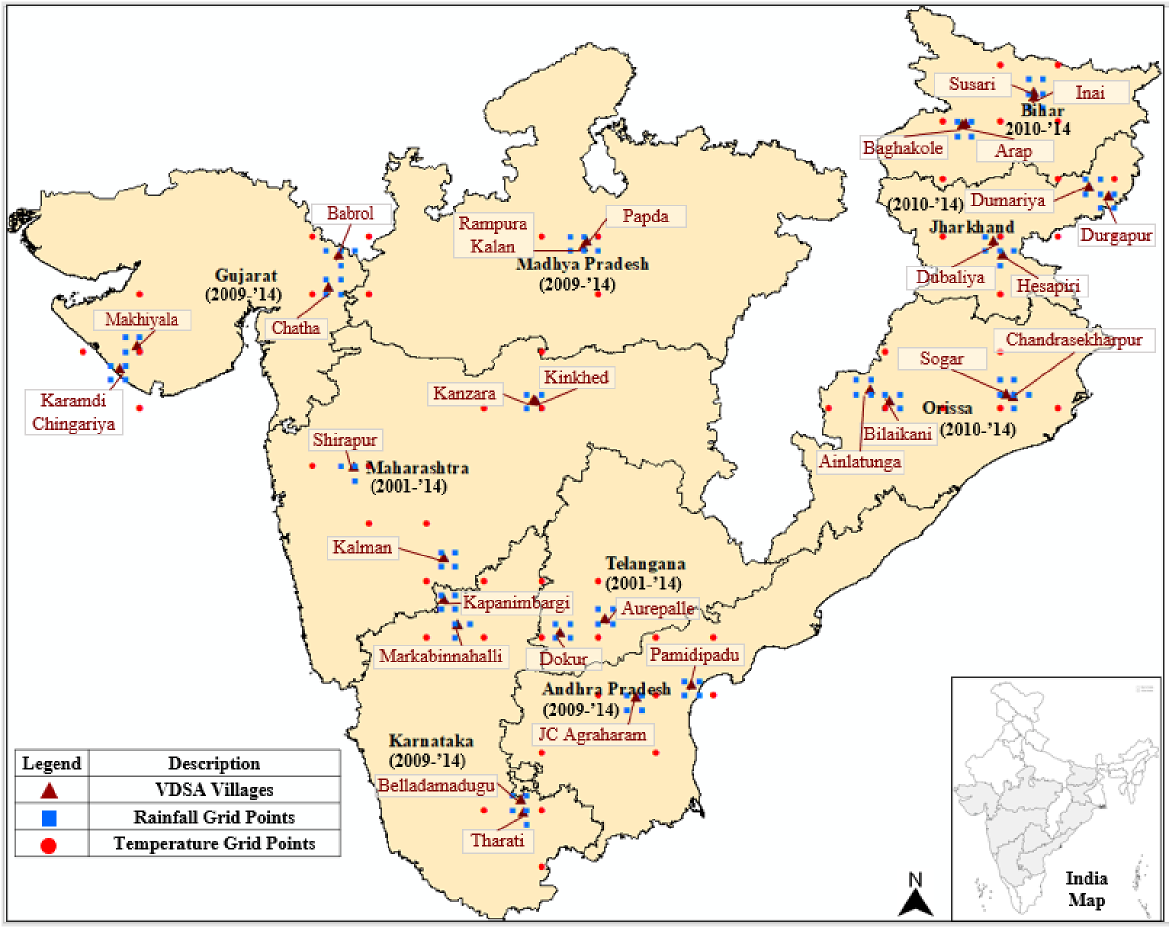


Figure 4.1: Spatiotemporal span of the VDSA Microdata (2001-14). The full list for 1975-2014 is provided in Appendix Table A1.

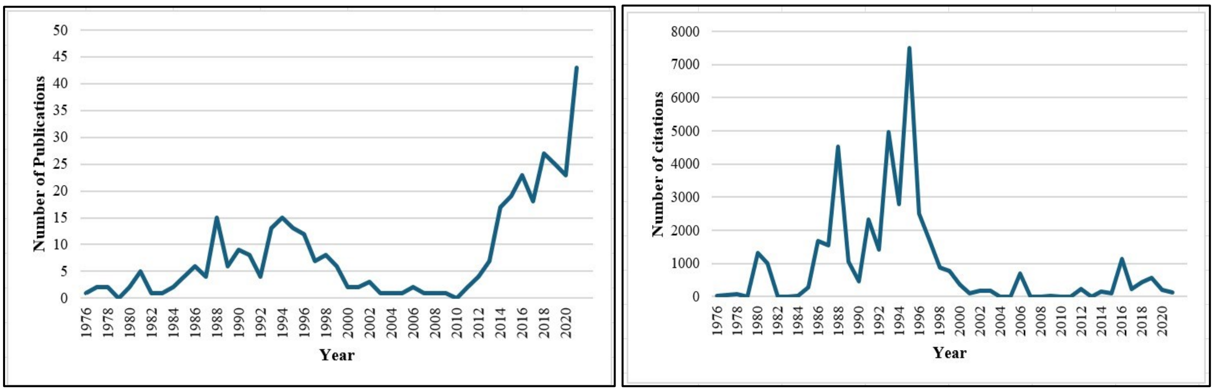


Figure 4.2: Year-Wise Count of Publications and Year-Wise Count of Citations using VDSA Microdata, Full details are provided in Appendix Table A2.

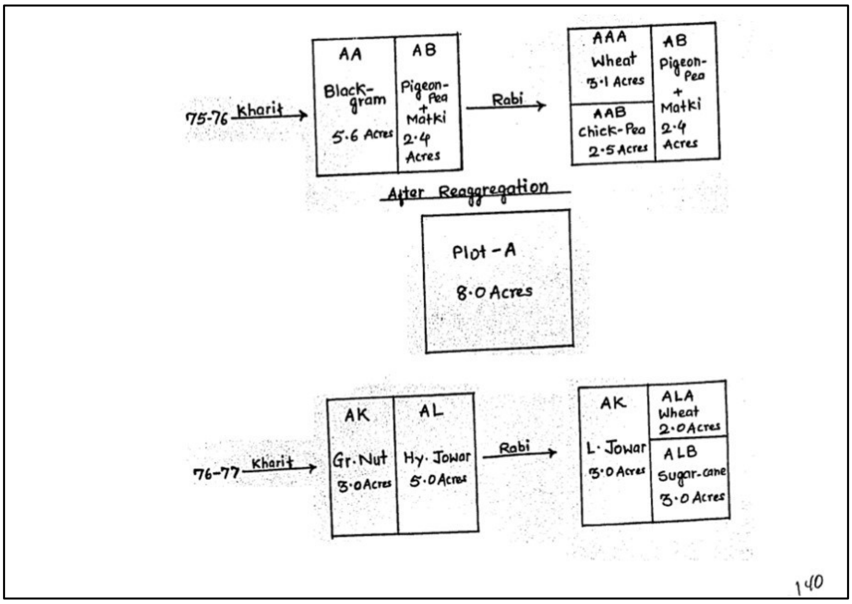


Figure 4.3: Accounting for Plot Naming Nomenclature followed in the VDSA Studies (Screenshot taken from the VDSA Data Manual (2001))

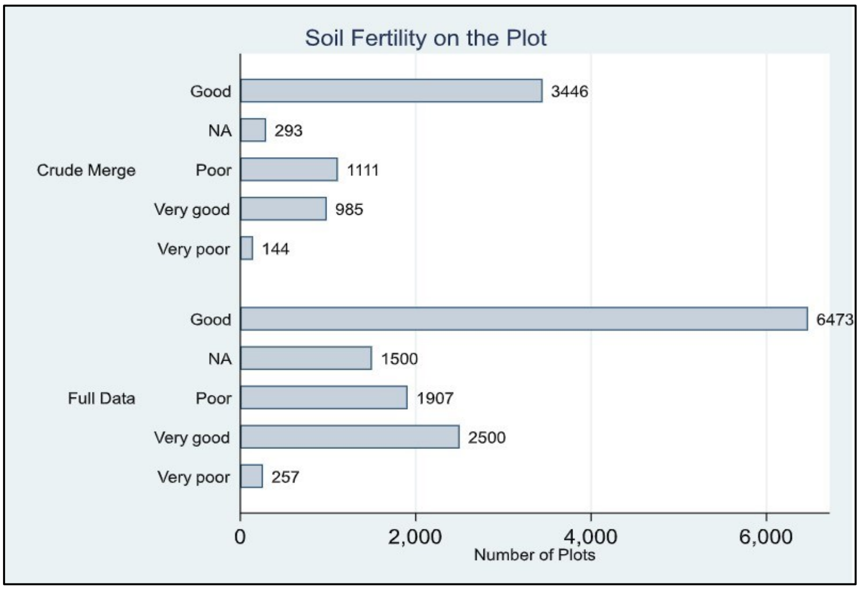


Figure 4.4: Data Gains for Soil Fertility

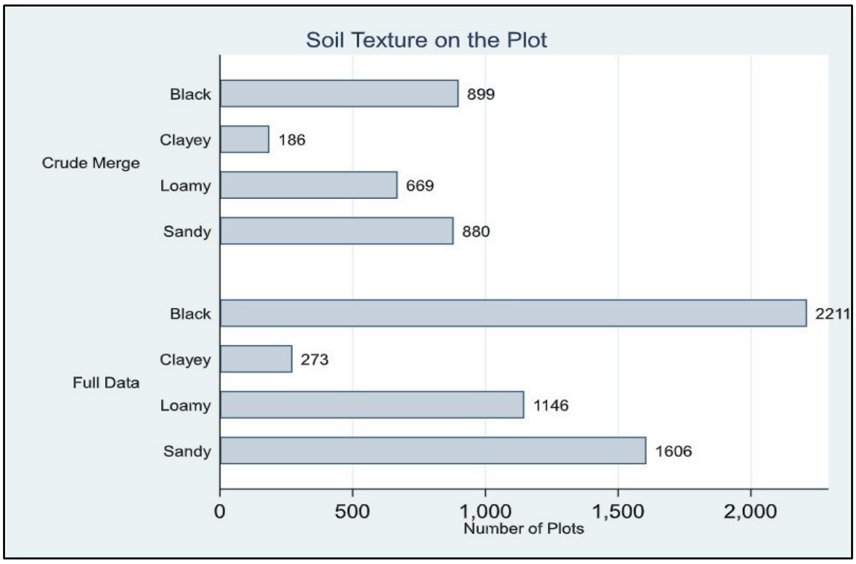


Figure 4.5: Data Gains for Soil Texture

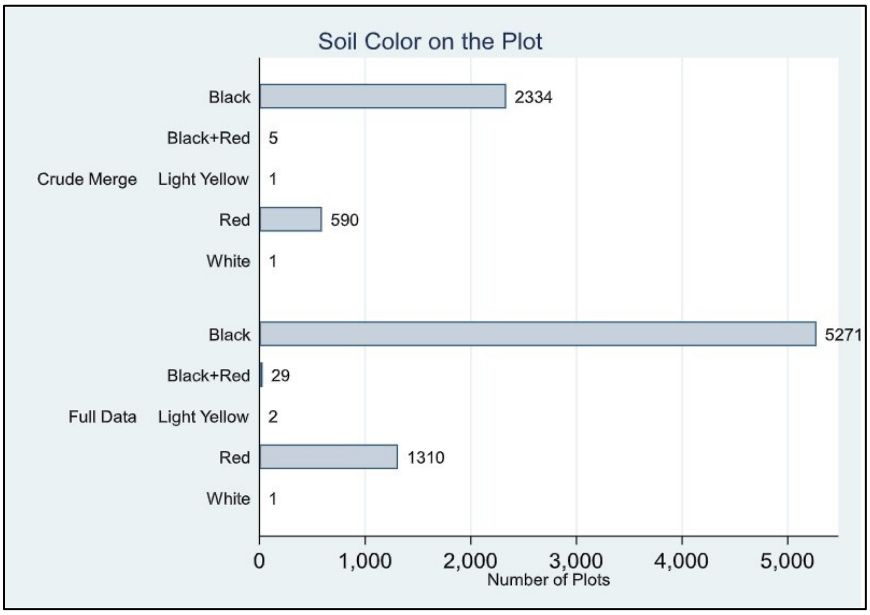


Figure 4.6: Data Gains for Soil Texture

4.7 Tables

Table 4.1: Crop Name changes in the CO File

Original Crop Name	Replaced With
Avare	Cluster Bean
Bajra	Pearl Millet
Beans (Bokala/Fawa Beans)	Beans(Bokala/Fawa/Runner Beans)
Berseem	Fodder
Berseem (Egyptian Clover)	Fodder
Betel Vine	Others
Black Gram	Blackgram
Bokala	Beans(Bokala/Fawa/Runner Beans)
Chillies	Chillies
Chillies Green	Chillies
China Aster	Aster
Coriender	Coriander
Cucurbit	Gourd
Dhemsa	Tinda
Egyptian Clover	Fodder
Fallow	Current Fallow
Fodder (Egyptian Clover)	Fodder
Fodder Oats (Jai)	Fodder
Freanch Bean	French Bean
Freanch Beans	French Bean
French Bean	French Bean
Gajaraj Grass	Fodder
Gajraj Grass	Fodder
Gardenpea	Green Pea
Gherkin	Cucumber
Grape	Grapes
Grass	Fodder
Green Chilly	Green Chillies
Green Fodder	Fodder
Green Fodder (Janer)	Fodder
Green Gram	Greengram
Greengram Fodder	Fodder
Ground Nut	Groundnut
Gumguvar	Guvar Gum
Horse Gram	Horsegram
Hulga	Horsegram
Ivy Gourd	Gourd

Kazilime	Kagzi Lime
Kolla Ganjeru Fodder	Fodder
Lady's Finger	Ladys Finger
Leafy	Leafy Vegetable
Lucern Grass	Fodder
Maize Fodder	Fodder
Mix Vegetable	Vegetable
Moong	Greengram
Napier Grass	Fodder
Oats (Jai)	Fodder
Orchard(Bagicha)	Flower
Ots (Green Fodder/Ots/Jai)	Fodder
Paddy Local	Paddy
Paddy Seed	Paddy
Paddy Seedling	Paddy
Pady	Paddy
Para Grass	Fodder
Pea	Green Pea
Pearl Millet Fodder	Fodder
Pigeon Pea	Pigeonpea
Pillipesara	Fodder
Runner Beans	Beans(Bokala/Fawa/Runner Beans)
Sarguja	Sajgura
Sesbania Aculeata (Dhaicha)	Dhaincha
Small Millet (Sarguja)	Sajgura
Sorghum Fodder	Fodder
Soyabean	Soybean
Subabul	Fodder
Sweet Orange	Orange
Tewara	Tewada
Turmuric	Turmeric

Table 4.2: List of Unique Crop Names

S. No.	Crop Name
1	Acarus Calamus
2	Agashi
3	Amaranthus
4	Amla
5	Arecanut
6	Aster
7	Avestor
8	Banana
9	Barley
10	Beans
11	Beans(Bokala/Fawa/Runner Beans)
12	Ber
13	Billion Dollar Grass (Sawan)
14	Bitter Gourd
15	Black Mustard
16	Blackgram
17	Bottle Gourd
18	Brinjal
19	Cabbage
20	Capsicum
21	Carrot
22	Castor
23	Cauliflower
24	Chickpea
25	Chillies
26	Chrysanthemum
27	Citrus
28	Cluster Bean
29	Coconut
30	Colocasia
31	Coriander
32	Cotton
33	Cowpea
34	Cucumber
35	Cumin
36	Current Fallow
37	Custard Apple
38	D Lab Lab
39	Dhaincha

40	Drumstick
41	Fenugreek
42	Finger Millet
43	Flower
44	Fodder
45	French Bean
46	Garlic
47	Gelardia
48	Ginger
49	Gourd
50	Grapes
51	Green Chillies
52	Green Grass
53	Green Leaves
54	Green Pea
55	Greengram
56	Groundnut
57	Guava
58	Guvar Gum
59	Horsegram
60	Jasmine
61	Jungle Tree
62	Jute
63	Kagzi Lime
64	Ladys Finger
65	Lathyrus
66	Leafy Vegetable
67	Lemon
68	Lentil
69	Lettuce
70	Linseed
71	Lotus Seed
72	Mahua
73	Maize
74	Mango
75	Marigold
76	Matki
77	Minor Millet
78	Minor Pulse
79	Mulberry
80	Mustard

81	Niger Seed
82	Oats
83	Onion
84	Orange
85	Others
86	Paddy
87	Papaya
88	Pearl Millet
89	Peas
90	Phaseolus Trilobus
91	Pigeonpea
92	Pomegranate
93	Potato
94	Pumpkin
95	Rabda
96	Radish
97	Rapeseed
98	Red Green Leaf
99	Red Leaf
100	Ridge Gourd
101	Rose
102	Safflower
103	Sajgura
104	Sapota
105	Sesamum
106	Shevaga
107	Small Millet (Marua)
108	Smooth Gourd
109	Soft Gourd
110	Sorghum
111	Soybean
112	Spinach
113	Sugarcane
114	Sunflower
115	Sweet Potato
116	Tamarind
117	Teak
118	Tewada
119	Tinda
120	Tobacco
121	Tomato

122	Trees
123	Turmeric
124	Variga
125	Vegetable
126	Watermelon
127	Wheat

Table 4.3: Results of algorithmic merging procedure (2001-14)

Query	Count of CO entries
1	20527
2	18000
3	320
4	663
5	471
6	3425
Total	43406

Table 4.4: Results of manual merging procedure (2001-14)

Merge Condition	Count of CO entries
Phonetics	1873
Phonetics + Serial Number	2
Plot Code + Serial Number	694
Serial Number	227
Year	131
Total	2927

Table 4.5: Results of algorithmic merging procedure (1975-89)

Query	Count of CO entries
1	20845
2	577
3	Not Applicable
4	Not Applicable
5	Not Applicable
6	28

4.9 Appendix

Table A1: Village Level Data Availability

State	District	Taluk	Village	Year/Period
Andhra Pradesh (AP)	Mahbubnagar	Devarakadra	Dokur	1975-79, 1983-84, 1989, 2001-14
		Madgul	Aurepalle	1975-79, 1983-84, 1989, 2001-14
	Prakasam	Bestavaripeta	JC Agraharam	2009-14
		Korisapadu	Pamidipadu	
Bihar (BH)	Darbhanga	Baheri	Inai Susari	2010-14
	Patna	Bikram	Arap Bagakole	2010-14
Gujarat (GJ)	Junagadh	Junagadh	Makhiyala	2009-14
		Mangrol	Karamdi Chingariya	
	Panchmahal	Goghamba	Chatha	2009-14
		Santrampur	Babrol	
	Sabarkantha	Prantij	Boriya Rampura	1980-83, 1989
Jharkhand (JK)	Dumka	Jarmundi	Dumariya	2010-14
		Sikaripara	Durgapur	
	Ranchi	Kanke	Dubaliya	2010-14
		Namkum	Hesapiri	
Karnataka (KN)	Bijapur	Basavana Bagewadi	Markabbinahalli	2009-14
		Indi	Kapanimbargi	
	Tumkur	Koratagere	Tharati	2009-14
		Madhugiri	Belladamadugu	
Madhya Pradesh (MP)	Raisen	Gairatganj	Papda Rampura Kalan	1982-83, 2009-14
Maharashtra (MH)	Akola	Murtizapur	Kanzara Kinkhed	1975-84, 1989, 2001-14
	Solapur	Mohol	Shirapur	1975-84, 1989, 2001-14
		North Solapur	Kalman	1975-84, 1989, 2001-14

Odisha (OD)	Bolangir	Patnagarh	Ainlatunga	2010-14
		Punital	Bilaikani	
	Dhenkanal	Gondiya	Chandrasekharpur	2010-14
		Kamakhya Nagar	Sogar	

Table A2: List of studies utilizing the VDSA dataset is linked here. | **Source:** Author's Compilation

Chapter 5: Assessing systematic biases in farmers' soil quality perceptions and their linkage with farming decisions: Evidence from rice-growers in India

Sumedha Shukla

(IIIT-Delhi)

sumedhas@iiitd.ac.in

Gaurav Arora

(IIIT-Delhi)

gaurav@iiitd.ac.in

Working Paper Link: https://ageconsearch.umn.edu/record/336014?ln=en&v=pdf

DOI: https://ageconsearch.umn.edu/record/336014

Assessing systematic biases in farmers' soil quality perceptions and their linkage with farming decisions: Evidence from rice-growers in India

Abstract

Maintaining soil quality in the face of increasing human pressure on land is a major challenge. Policy efforts to improve soil quality aim to influence farm-level decisions, but the adoption of such interventions depends on farmers' perceptions of soil quality, which shape their subjective cost–benefit calculations. These perceptions may diverge from data-based evidence, as expectations about unobserved soil quality are typically formed using observable but imperfect proxies such as crop yields. In this study, we characterize soil quality misperceptions for a large-scale, representative sample of rice-growing farmers in India by matching geo-referenced plot locations with high-resolution gridded soil maps. We construct a plot-level measure of soil quality misperception as the gap between farmer-reported soil quality and corresponding data-based soil measures. We then document the farm-level distribution of biases in soil quality perceptions and identify their underlying drivers, including growers' socio-economic characteristics and demographics. Finally, we examine a previously unexplored two-way causal relationship between soil quality misperceptions and farm-level fertilizer application using an instrumental variable (IV) framework. We find that farmers perceive soil texture with greater consistency than soil fertility. Fertility perceptions exhibit systematic bias with over-perception exceeding under-perception. The IV results indicate a bidirectional relationship: misperceptions significantly influence nitrogen, phosphorus, and potassium application, while higher fertilizer use reduces the likelihood of over-perceiving soil fertility, consistent with belief updating driven by self-serving attribution. Our analysis highlights the joint determination of soil misperceptions and input decisions, with important implications for land management policy design.

JEL Codes: Q120; Q150; C210

Keywords: Soil Quality Perceptions; Farming Decisions; Ordered Logit Model

Assessing systematic biases in farmers' soil quality perceptions and their linkage with farming decisions: Evidence from rice-growers in India

5.1 Introduction:

Human dependence on soil resources spans centuries with soil quality shaping the rise and decline of civilizations through productivity and sustainability of cultivation on land (Montgomery 2007; Parikh and James 2012). Beyond food production, soils provide critical ecosystem services including carbon storage, water regulation and supply of raw materials such as gravel (Blum et al. 1988; McNeill and Winiwarter 2004; Baveye et al. 2016). However, soil health is under threat with approximately one-third of the world's soils classified as moderately to highly degraded (Smith et al. 2024), which poses serious risk to global food security and rural livelihoods (Perrings 1989; Barrett and Bevis 2015; Barbier and Hochard 2016). Policy efforts to maintain soil quality typically target shifts in farm-level practices including crop selection, input use and land management[47].

Adoption of diverse policy mandates depends on farmer perceptions about soil quality that inform *subjective* assessments of costs, returns and perceived need for such innovations and practices (Rogers 1995). That perceptions matter in farmer decision-making processes is well-established (Niemijer and Mazzucato 2003; Marenya et al. 2008; Essougong et al. 2020; Delgado and Stoorvogel 2022). For example, Ervin and Ervin (1982) posit that farmer perception about soil erosion and its impact on farm returns and land value are potentially correlated with farmers' willingness to pay for soil conservation practices. Similarly, Zhou et al. (2020) show that soil pollution perception is positively associated with farm-level adoption of pro-environmental behaviour in China (e.g., judicious chemical fertilizer use and

[47]Examples include conservation agriculture programs that encourage zero tillage, crop residue retention and crop diversification/rotation in Sub-Sahran Africa (Bolliger et al. 2006; Hobbs 2008; Barrett et al. 2002; Giller et al. 2009; Krah et al. 2019; Ofstehage and Nehring 2021); land retirement in the United States (Collins et al. 2012; Hoffman 2025); fertilizer reduction in China (Chen et al. 2022) and nutrient information-based intervention such as Soil Health Card scheme in India (Morton et al. 2023).

avoidance of straw burning). Tittonell et al. (2005) find differences in the timing and intensity of crop management by farmers' perceived soil quality in Kenya.

Importantly, however, farmer perceptions about soil quality may *not* match with data-based evidence (Barrios and Trejo 2003; Gray and Morant 2003; Berazneva et al. 2018). Soil dynamics are complex and spatially heterogeneous (Antle et al. 2006; Barrett 2008) where key indicators like organic carbon, nitrogen and phosphorus differ non-linearly with soil texture, mineralogy, and land use (Moebius-Clune et al. 2011). Moreover, farmers do not directly observe soil quality but infer it from noisy and lagged signals such as crop yield (Moges and Holden 2007; Karltun et al. 2013), environmental stressors and soil input responsiveness (Martey and Kuwornu 2021; Essougong et al. 2020; Marenya et al. 2008). Institutional extension services and peer networks can mitigate the relevant informational gap, however access to such resources is not uniform across farmers (Green and Heffernan 1987; Elwell et al. 2018; Wang et al. 2020; Kolady et al. 2021; Alkaabi et al. 2025). Besides informational constraints, farmers face cognitive limitations in processing complex and dynamic soil–yield relationships, leading them to rely on simplified mental models and form expectations from noisy signals and production outcomes (Johnson-Laird 1983; Morecroft 1983; Marenya et al. 2008). In addition, psychological mechanisms such as self-serving attribution and motivated beliefs may influence soil quality perceptions (Miller and Ross 1975; Mezulis et al. 2004; Bénabou and Tirole 2002; Castagnetti and Schmacker 2022). Specifically, farmers may attribute unfavourable outcomes to soil conditions rather than prior decisions, whereas favourable outcomes get attributed to their 'smart' decisions. Alternatively, farmers receiving adverse signals about soil quality may strategically preserve optimistic soil fertility belief in order to sustain the motivation to cultivate and invest in farm operations. Together, these mechanisms imply a bidirectional relationship between soil quality perceptions and farming decisions, where perceptions shape farming decisions while past decisions (and outcomes) feed back into the formation of perceptions.

In this study we provide a geo-coded measure of potential biases in farmer perceptions about soil fertility and soil texture for 8,281 rice-growing plots across eight major rice-producing states in India (see Figure 1 for plot locations). These data were collected during 2018 *Kharif* (Monsoon) season by the International Maize and Wheat Improvement Centre (CIMMYT) as part of their landscape diagnostic surveys on rice production practices in India (Ajay et al. 2022). Farmers' soil quality mis-perceptions are defined as the *difference* between primary survey-based soil quality perceptions and corresponding data-based evidence recorded in the form of gridded soil maps (made available by the International Soil Reference and Information Centre). We seek to identify socioeconomic and geographic drivers that underlie how and why farmer perceptions differ from scientific measurements of soil quality, including the linkage between soil quality misperception and farming decisions. Ethnopedology (or study of indigenous knowledge systems related to soil and land management) provides several original accounts of divergence between farmers' soil quality perceptions and corresponding scientific or data-based measurement. Early studies sought to overlay locally informed soil classification used by farmers onto a map of scientifically measured soil properties with the aim to identify (chemical and physical) scientific indicators that best correspond to farmers' soil assessments (Liebig and Doran 1999; Murage et al. 2000; Barrios and Trejo 2003; Andrews et al. 2003; Gruver and Weil 2007). Observable soil attributes such as soil texture (measured by clay content) are perceived with reasonable accuracy, whereas less visible and more dynamic fertility indicators including nitrogen, organic carbon, phosphorus, and pH present mixed evidence (Krogh and Paarup-Laursen 1997; Liebig and Doran 1999; Murage et al. 2000; Barrios and Trejo 2003; Andrews et al. 2003; Desbiez et al. 2004; Nord and Snapp 2020). For instance, Murage et al. (2000) find that soils perceived as more productive exhibit significantly higher phosphorus, nitrogen, and pH in Kenya. whereas Gruver and Weil (2007) report no significant relationship between farmer-assessed soil quality and pH, phosphorus, or potassium levels among Mid-Atlantic farmers.

Much of this evidence is based on relatively small samples (typically fewer than 100 farmers), limited geographic scope (often within a single village), and concentrated in Sub-Saharan Africa with a handful of studies in Nepal and United States. This limited coverage is partly driven by the high cost of collecting plot-level soil measurements through field sampling and laboratory analysis, which constrains the scale of such studies. Moreover, existing analyses largely rely on unconditional mean comparisons or ANOVA-based approaches to contrast perceived and measured soil quality. While informative about the relative importance of different soil indicators in specific contexts, these approaches provide limited insight into the distribution of perception gaps at the farm level and potential socioeconomic and behavioural mechanisms that may underlie these. More recent work uses larger and more spatially diverse samples (Berazneva et al. 2018; Delgado and Stoorvogel 2022) but the linkages between soil perceptions and farming decisions remain underexplored.

We leverage geo-coded gridded soil data (similar to Berazneva et al. 2018) as a scalable, low-cost alternative to plot-level soil samples, enabling granular measurement of gaps in soil quality perception over large and diverse geographic regions. We acknowledge that gridded soil maps are not ground-truth alternative to sensor-based field observations and may suffer from measurement error due to spatial aggregation/interpolation (Cressie 1992). While such data do not capture exhaustive biophysical details, they are considered to meaningfully capture economic relationships pertaining to farm production and input decisions (Berazneva et al. 2018). We address the concern of potential measurement error through robustness checks, ranging from use of multiple soil quality indicators to varying “neighbourhood” size when matching plot-level survey locations to gridded soil categories.

We construct plot-level measure of soil fertility misperceptions as the gap between farmer’s soil quality perceptions and corresponding data-based records in gridded soil maps. In the process, we first provide a regression-based, sample-wide summary of misperceptions by modelling farmers' soil quality perceptions as a function of soil quality data conditional on a variety of socioeconomic and geographical factors. We further document farm-level unconditional distribution of soil fertility misperceptions derived from cross-tabulations between soil quality perceptions and data-based evidence.

Finally, we examine the linkage between soil fertility misperceptions and farm-level fertilization decision through which we envision informing effective nutrient management policies for rice production in India (Mondal and Basu 2009; Pingali 2012). Despite extensive subsidization of urea-based nitrogenous fertilizers, fertilizer costs (excluding application costs) accounted for more than 10% of total operational cost for rice cultivation in India in 2018. We recognize that while misperceptions can influence nutrient application (e.g., Tittonell et al. 2005; Zhou et al. 2020), input use may also shape perceptions (Miller and Ross 1975; Mezulis et al. 2004; Bénabou and Tirole 2002). Theoretically, these relationships are directionally ambiguous.

Overestimating soil quality may reduce nutrient application if farmers perceive limited need, or increase it in anticipation of high returns from productive plots. Conversely, underestimating soil quality may suppress fertilization due to pessimistic expectations or induce nutrition substitution for perceived deficiencies. In the reverse direction, farmers may attribute favourable yields to their own input use decisions rather than soil quality, where poor outcomes may be (mis)attributed to soil conditions rather than own choices (Miller and Ross 1975; Mezulis et al. 2004).

We assess the linkage between soil quality misperceptions and fertilizer application in two steps. First, we employ two-way contingency tables to test whether farmers' soil perceptions are on average independent of their input application decisions. Second, we employ an instrumental variable approach through multivariate regression to identify the causal link in both directions — the effect of soil misperceptions on farming decisions and the effect of farming decisions on soil misperceptions. While some literature has examined the former linkage i.e., how soil quality perceptions shape farmer decisions consistent with the standard decision-theoretic framework in which beliefs about the production environment enter the farmer's optimization problem (Delgado and Stoorvogel 2020), the latter linkage is largely overlooked — decisions may themselves generate and sustain the very beliefs that motivate them. Establishing the direction and magnitude of both causal channels is essential not only for understanding how soil quality misperceptions arise, but also for designing interventions that can break the self-reinforcing cycle between distorted beliefs and suboptimal input decisions. To the best of our knowledge, no existing study jointly analyses soil quality misperceptions and input decisions in a unified manner and at scale.

Our study makes three main contributions to the literature. First, we construct a plot-level measure of soil fertility misperceptions as the gap between farmer's soil quality perceptions and corresponding high-resolution gridded soil quality data. Second, we articulate farm-level distribution of biases in soil quality perceptions and identify the underlying drivers, including growers' socio-economic status and demographics. Third, we examine previously unexplored two-way causal linkage between soil quality misperceptions and farm-level input decision. Our analysis highlights the joint determination of soil misperceptions and input decisions with implications for effective policy designs.

Section 2 outlines the survey-based and auxiliary data used to construct and explain soil quality misperceptions along with providing modelling details. Section 3 provides the main results of our study, followed by discussions and conclusion in Section 4.

5.2 Materials and Methods

5.2.1 Data and Data Processing

Survey Data: soil quality perceptions and input application decisions

We utilize geo-referenced data on 8,281 Indian farmers' soil quality perceptions for their "primary" (or largest) rice-cultivated plot during the 2018 *Kharif* (Monsoon) season. Figure 1 shows plot locations across eight Indian states of Punjab, Haryana, Bihar, Uttar Pradesh, West Bengal, Orissa, Chhattisgarh and Andhra Pradesh, which represent the major rice-producing regions of India. In sampling terms, observations are distributed in proportion to the eligible population of rice growers across districts within each state (Ajay et al. 2022; Coggins et al. 2025). Data summaries are presented in Table 1 and further discussed in the next section.

It is noteworthy that the primary purpose of these surveys was to collect information on the economics of rice production, which includes data related to land preparation, crop varieties, fertilizer use, irrigation, crop protection, and harvesting. We argue that eliciting farmer perceptions within their production environment can allows us to systematically relate on farm decisions and belief formation. Our setting also contrasts with existing work where attention is directed towards soil characteristics as a primary survey objective, which in turn can interfere with the perception reports (Diersch and Walther 2016).

In our survey, farmers are asked to report their soil perceptions along two dimensions: soil "quality" and soil "texture", both of which relate to the broader concept of soil quality in the ethnopedology literature (Andrews et al. 2003; Gruver and Weil 2007). To differentiate functional and physical attributes of soil, we interpret reported perceptions on soil "quality" as capturing perceptions about soil "fertility" (i.e., chemical and biological nutrient availability), while soil "texture" is treated as a distinct physical attribute related to soil structure (Gruver and Weil 2007; Berazneva et al. 2018). Denote farmer perception about soil quality of plot *i* by by $\tilde{S}_i$ which is recorded as a set of two ordered attributes: "Fertility ($\tilde{F}_i$)"

and "Texture ($\tilde{T}_i$)". Specifically, survey data comprises information on $\tilde{S}_i \in \{\tilde{F}_i, \tilde{T}_i\}$ such that $\tilde{S}_i$=1 if plot i is perceived to have {"Low Fertility", "Light Texture"} soils; =2 if i is perceived to have {"Medium Fertility", "Medium Texture"} soil; and =3 if is perceived to have {"High Fertility", "Heavy Texture"} soil.

With regards to soil fertility perceptions, the largest sample proportion (~88%) of farmer perceptions were recoreded to be "Medium" fertility, whereas only 4% (333 farmers) perceived "Low" fertility and 8% (658 farmers) perceived "High" quality soils on their primary plots. Similarly, with regards to texture, only 7% and 14% farmers perceived their primary plots to have "Light" and "Heavy" textured soils, respectively, whereas the remaining 78% sample farmers perceived their soil to be of "Medium" texture. , i.e, neither Heavy nor Light. The concentration of responses in the "Medium" category is consistent with the well-documented anchoring bias associated with the presentation of a coarse Likert-scale (Tversky and Kanheman 1974). Given that soil quality is complex and partially observable, it makes sense for farmers to not map beliefs onto extreme categories (Douven 2018).

As discussed earlier, we evaluate potential causal link between soil fertility misperceptions and fertilizer use on farm plots. To do this we construct nutrient-specific data for Nitrogen (N_i), Phosphorus (P_i) and Potassium (K_i) levels. These data are based on the fertilizer type (i.e., urea, Diammonium Phosphate, Single Superphosphate, Sulphate of Potash, Muriate of Potash, and compound NPK and NPKS fertilizers) and quantity (kilograms per-acre) applied across four rounds (basal, first, second and third top-dressing) within the growing season. We use standard nutrient content coefficients associated with each fertilizer type (Fertilizer Association of India 2022) to convert quantities applied into nutrient-specific amounts of N, P and K. About 98% sample farmers apply mineral fertilizer indicating near-universal adoption of chemical inputs among rice-growing farmers. Below we provide details on physical data used, including those on soil quality.

Gridded Soil Data and Climate Data:

Data-based soil quality measurements denoted as $S_i \in \{F_i, T_i\}$, were obtained by matching plot *i*'s location to spatially-delineated (250m x 250m resolution) global digital soil maps provided by the International Soil Reference and Information Centre. Plot-level soil quality values are constructed by averaging soil quality indicator values within a circular buffer of 250 metres radius centred on the plot location[48]. F_i is measured as <u>organic carbon stock</u> (O_i^S) (tonnes/hectare i.e., t/ha) and <u>nitrogen content</u> (N_i^S) (centigram/kilogram i.e., cg/kg) while T_i is measured as <u>clay content</u> (C_i^S) (gram/kilogram i.e., g/kg) in top soil layer which is 0-30 centimetres depth from the surface. To obtain the soil quality values for top soil, i.e., soil depth ranging from 0-30 cm, all soil quality indicators values were averaged for the soil depths ranges from 0-5 cm., 5-15 cm., and 15-30 cm.

We find that the average soil nitrogen level across our sample locations is ~137 centigrams/kilogram of soil and displays significant heterogeneity with over-depletion in states like Punjab and Haryana. The average organic carbon stock level across our sample locations is ~40 tonnes/hectare and displays significant heterogeneity with states like Bihar and West Bengal having above average organic carbon stock likely due to the clayey texture of the soil which helps in retention of organic matter. The average soil clay content is ~316 grams per kilogram of soil with heavier soils in coastal states like West Bengal & Odisha.

We note that the range of soil nitrogen and organic carbon in our data often lies above the 'high' agronomic thresholds for these measures, reducing their informativeness for assessing

[48] Given that the native resolution of the soil data is 250 metres, a 250-metre buffer ensures that the extracted value captures at least one full pixel centred on the plot location while averaging over any partial pixel overlap at the buffer boundary. As a robustness check, we repeated the extraction using buffers of 125 metres and 300 metres radius and found the distribution of soil quality indicators to be substantially identical with correlations exceeding 0.99 across all three buffer specifications, suggesting that our results are not sensitive to the choice of buffer radius. Please see Supplementary Information (SM), Section 1.

agreement between farmer perceptions and data-based measures of soil fertility. This pattern is consistent with evidence of widespread fertilizer over-application in these regions (see Coggins et al. 2025).

Climate data are drawn from the AReNA DHS-GIS dataset (International Food Policy and Research Institute 2018), which provides monthly temperature and rainfall records for DHS survey locations across India spanning the period 1980-2018. For each farm household in our sample, climate indicators are constructed by aggregating across the nearest five DHS locations[49]. We construct the coefficient of variation of temperature and rainfall separately for Kharif season (June through October) across 1980-2017. Seasonal coefficients of variation capture agronomically meaningful climatic variability and are used as instruments for soil fertility misperceptions in our empirical analysis, on the grounds that long-run climatic variability shapes how farmers form beliefs about their production environment.

5.2.2 Analysis of Farmers' Soil Quality Perceptions

Multivariate link between farmers' soil perceptions and data-based evidence:

We model each element of farmers' soil quality perception as an ordered logistic regression function (McCullagh 1980) of the respective data-based soil quality measurement alongside control vector $\vec{x}_i$. The ordinal logistic model is appropriate because farmer perceptions represent a clear ordering of soil quality (e.g., from low to high fertility or from light to heavy texture). Specifically, under the proportional odds assumption we have:

$$\begin{aligned}\text{logit}(\Pr(\tilde{S}_i \le 1)) &= \log\left(\theta_{i1}/(\theta_{i2}+\theta_{i3})\right) = \kappa_1 + \beta_s S_i + \vec{x}_i \vec{\gamma} + \varepsilon_{1i} \\ \text{logit}(\Pr(\tilde{S}_i \le 2)) &= \log\left((\theta_{i1}+\theta_{i2})/\theta_{i3}\right) = \kappa_2 + \beta_s S_i + \vec{x}_i \vec{\gamma} + \varepsilon_{2i}\end{aligned} \tag{1}$$

Here, $\theta_{ij} = \Pr(\tilde{S}_i = j), j = 1,2,3$. We estimate model (1) separately for separately for soil

[49] We repeated the construction using the nearest three and nearest ten DHS cluster locations respectively. Our results are robust. See Supplementary Materials, Section 2.

fertility perception and soil texture perception i.e., $\tilde{S}_i \in \{\tilde{F}_i, \tilde{T}_i\}$. Model (1) specifies the proportional odds of perception responses in a manner that the probabilities accumulate in an ascending order (i.e., they capture perception odds of being in a lower soil quality category relative to the higher ones): $\tilde{S}_i = 1$ relative to $\tilde{S}_i > 1$; and $\tilde{S}_i \le 2$ relative to $\tilde{S}_i = 3$; as a function of S_i conditional on $\vec{x}_i$. Model parameter β_s measures the degree of consistency between plot-level soil quality perceptions and corresponding data-based measurements. Given that soil quality is higher when increases, a negative β_s value indicates agreement between soil perceptions and data-based evidence for an average sample farmer. On the other hand, a positive value for β_s will reflect on-average biased soil quality perceptions.

Model parameters in $\vec{\gamma}$ represent the ceteris paribus shift the log odds of being in a lower perception category relative to a higher one, upon a marginal increase in the respective explanatory variable. Both β_s and $\vec{\gamma}$ are assumed to be the same regardless of the perception response interval, which is known as the proportional odds assumption. Parameters κ_1 and κ_2 characterize the average log odds of perception level $\tilde{S} = 1$ (e.g., low soil fertility relative to medium or high fertility levels) and $\tilde{S} \le 2$ (e.g., low or medium soil fertility levels relative to high fertility) respectively, when all other explanatory variables are zero.

Control variables in $\vec{x}_i$ include: (a) farmer demography: education, caste and household size; (b) economic situation: income share from agriculture, property rights and land size (Berazneva et al. 2018); (c) irrigation availability (Desbiez et al. 2004; Moges and Holden 2007), (d) climatic controls for the Kharif season: long-term (1980-2017) temperature and rainfall mean as well as coefficient of variation (i.e., climate variability) (Martey and Kuwornu 2021; Tesfahugnen et al. 2021); and (e) geographic controls: latitude, longitude (Murage et al. 2000; Arora and Feng 2024). Overall, Model (1) assesses whether farmers' soil quality perceptions are, on average, consistent with corresponding data-based measures, which we formalize in the following hypothesis.

Hypothesis I: Farmers' soil quality perceptions are on average consistent with corresponding data-based evidence conditional on control vector $\vec{x}$, i.e., $H_0 : \beta_s \le 0$ against the alternative that $\beta_s > 0$.

We next focus our misperceptions in soil fertility, where discrepancies between perceived and data-based measures are quite pronounced (please see details in the 'Results' section).

Articulating biases in farmers' soil fertility perceptions:

As a first step, we cross tabulate $\tilde{F}_i$ and F_i to characterize the joint distribution of biased and consistent soil fertility perceptions. We categorize each soil fertility measure into equal-frequency tertiles (Shanmugam and Chattamvelli 2015) to align with the ordered response categories used to elicit farmer perceptions i.e, low, medium and high. This approach partitions the distribution of measured soil fertility F_i into equal-sized bins based on empirical thresholds that are independent of the perception data. The discretization of measured soil fertiltiy is meant to facilitate quantitative comparisons between discrete perceptions categories and continuous soil fertility measures (Arora and Feng 2024). The tertile-based classification is comparabile across distinct soil metrics (i.e., N_i^S and O_i^S discussed in Section 2.1.2) given that there is no single preferred soil fertility indicator in the literature (Andrews et al. 2003; Gruver and Weil 2007; Lal et al. 2024).

Specifically, we classify soil nitrogen data (N_i^S) as: $N_i^S \le 120.5$ (Low); $N_i^S \in (120.5, 142.8]$ (Medium); and $N_i^S > 142.8$ (High) to match perceived soil fertility categories $\tilde{F}_i \in \{\text{Low Fertility, Medium Fertility, High Fertility}\}$. Similarly soil organic carbon stock categories $O_i^S \in \{\le 38.3 \text{ (Low)}, (38.3, 42.6] \text{ (Medium)}, > 42.6 \text{ (High)}\}$ are matched to the

$\tilde{F}_i \in \{\text{Low Fertility, Medium Fertility, High Fertility}\}$. We present these cross-tabulations in Table 2 which provides counts of sample farmers across joint categories of perceived and data-based soil fertility. The diagonal elements in Table 2 provide the count of farmers whose soil fertility beliefs are consistent with data-based soil fertility measurement while the off-diagonal elements refer to those farmers whose perceived soil fertility was inconsistent with data-based measurement. We define the off-diagonal elements in Table 2 as a measure of soil fertility *mis*perceptions, which we will further analyze in relation to their fertilizer use decisions.

5.2.3 Linking Farmers' Soil Fertility Misperceptions to Fertilizer Application Decisions: We examine whether soil fertility misperceptions generate systematic patterns with farm-level fertilization decisions. In particular, we assess whether farmers who over-perceive soil fertility adjust fertilizer application downward, believing their soil to be sufficiently fertile, or instead increase its use in an attempt to maximize expected returns. Similarly, we investigate whether farmers who under-perceive soil fertility reduce fertilization due to pessimistic expectations, or increase fertilizer use to compensate for perceived soil deficiencies.

We also consider the reverse relationship whereby past input use decisions may shape farmers' soil fertility perceptions (Miller and Ross 1975; Bénabou and Tirole 2002; Mezulis et al. 2004; Castagnetti and Schmacker 2022). In this framework, farmers may adjust their beliefs to justify prior decisions or protect self-image. For example, a farmer who applies high levels of fertilizer and observes favourable yields may attribute these outcomes to their own input choices rather than to underlying soil fertility, thereby attenuating over-perceptions of soil fertility. Conversely, if such a farmer observes poor yields despite high input use, they may attribute this to low soil fertility, reinforcing pessimistic beliefs. Similarly, a farmer who applies low levels of fertilizer and observes poor yields may attribute outcomes to inherently low soil fertility, reinforcing under-perceptions. In contrast, favourable yields under low input use may be attributed to high soil fertility, thereby moderating such pessimistic beliefs. Together, these mechanisms suggest a bidirectional relationship between soil fertility perceptions and input use decisions. We formalize these hypotheses below.

Hypothesis IIA (Soil fertility misperceptions impact fertilization decisions): Farmers' soil fertility misperceptions affect their input use decisions. In other words, farmers who over- or under-perceive soil fertility may systematically differ in amount of fertilizer applied *relative* to farmers who accurately perceive soil fertility of their farms.

Hypothesis IIB (Fertilizer use impacts farmer soil fertility beliefs): On farm fertilizer use decisions systematically impact soil fertility misperceptions. In other words, farmers who apply higher or lower fertilizer systematically overperceived or underperceived their soil fertility *relative* to farmers who accurately perceive their soil fertility.

To test hypotheses IIA and IIB, we first conduct a Chi-squared (χ^2) test of independence between farmers' soil fertility misperceptions and application of fertilizers: nitrogen (N_i), phosphorus (P_i) and potassium (K_i). Second, we employ an instrumental variable (IV) approach to evaluate the potential for a causal link between farmers' soil fertility perceptions and fertilizer application decisions. We focus our subsequent analysis exclusively on farmers who perceive soil fertility as either "Low" or "High" categories where the direction of misperception is clear. This analytical focus on directional misperceptions is analogous to the approach of Arora and Feng (2024), who concentrate on farmers who misperceive drier conditions as wetter or wetter conditions as drier, setting aside the share of farmers who report 'about the same' weather change.

Within this subsample, we identify two types of misperceptions that are the basis of our farming decision analysis: over-perceptions and under-perceptions. We denote soil fertility over-perceptions by $B_i^L \in \{L \to H, L \to L\}$ where $L \to H$ denotes a scenario where soils with low measured soil fertility (N_i^S or O_i^S) are perceived as high fertility while $L \to L$ denotes cases of consistent perceptions of low fertility soils. Similarly, soil fertility under-perceptions are denoted by $B_i^H \in \{H \to L, H \to H\}$ where $H \to L$ captures cases where soils with high measured fertility (N_i^S or O_i^S) are perceived as low fertility while $H \to H$ denotes cases of consistent perceptions of high fertility soils.

As noted earlier, we focus on fertilizer utilization decisions, which we categorize into "Low" or "High" based on a median cut-off (Shanmugam and Chattamvelli 2015). Specifically, define $\tilde{N}_i = 1\{N_i \geq \text{median}(N_i)\}$ for farm-level nitrogen application, $\tilde{P}_i = 1\{P_i \geq \text{median}(P_i)\}$ for farm-level phosphorus application and $\tilde{K}_i = 1\{K_i \geq \text{median}(K_i)\}$. We define a χ^2 statistic based on a cross tabulation of over-perceptions i.e., B_i^L and each of $\tilde{N}_i$, $\tilde{P}_i$ and $\tilde{K}_i$, and under-perceptions i.e., B_i^H and each of $\tilde{N}_i$, $\tilde{P}_i$ and $\tilde{K}_i$ in a 2 × 2 contingency table where each cell entry represents the count of farmers.

Specifically, let $M_{\text{Row}(r)\,;\,\text{Column}(c)}$ denote the number of farmers in cell (r, c) where r indexes fertilizer use categories and c indexes soil fertility misperception categories. The total number of observations is $M = \sum_r \sum_c M_{r,c}$ with row and column marginals given by $M_{r,.}$ and $M_{.,c}$ respectively. We define the corresponding joint probability as $\rho_{r,c} = M_{r,c} / M$ with marginals $\rho_{r,.}$ and $\rho_{.,c}$. Under the null hypothesis of independence between fertilizer use and soil fertility misperceptions, the joint probabilities factorize as $\rho_{r,c} = \rho_{r,.} \times \rho_{.,c} \forall r, c$. This implies that expected cell frequencies $\hat{M}_{r,c} = \left(M_{r,.} \times M_{.,c}\right) / M$. The test statistic is defined as:

$X = \sum_r \sum_c \dfrac{(M_{rc} - \hat{M}_{rc})^2}{\hat{M}_{rc}} \sim \chi_d^{\,2}$ where $d = (R-1)(C-1)$ are the degrees of freedom. If $X > \chi^2_{d,5\%}$

(or $X \le \chi^2_{d,5\%}$) then we will reject (or fail to reject) the null hypothesis that input decisions and soil fertility misperceptions are independent of each other.

The contingency tables do not provide sufficient evidence that soil fertility misperceptions causally lead to systematic biases in farmers' input use decisions, or that input use decisions causally lead to systematic biases in soil fertility misperceptions. Soil fertility misperceptions and input use decisions may be endogenous to each other. We propose to address endogeneity by specifying instruments that directly impact the variable of interest in each direction but influence the outcome only through the hypothesized causal channel.

5.2.4 Assessing how soil fertility misperceptions impact of fertilizer use decisions:

We specify a two-stage least squares (2SLS) regression framework to identify the impact of soil fertility misperceptions on farmers' fertilizer use decisions. We instrument soil fertility perceptions with lagged weather variability and pest incidence. To address endogeneity between farmer belief formation and on farm decisions (Arora and Feng 2024), we instrument soil fertility perceptions using long-term temperature variability (σ_i^{temp}) and insect incidence from the past (insect_i). The identifying assumption is that weather variability and biotic stressors influence how farmers form beliefs about their production environment including soil fertility (Martey and Kuwornu 2021; Tesfahugnen et al. 2021). For example, a period of high temperature may lead farmers to revise their assessment of soil quality downward, attributing poor yields partly to soil conditions. At the same time, the exposure to past weather shocks is plausibly exogenous to current input decisions: weather and pest realizations in prior seasons are not determined by present fertilizer application choices.

The identification strategy relies on the assumption that cropping risk incidence in the past will influence present-day fertilization only through their impact on farmers' soil fertility perceptions. This is plausible as both factors are salient drivers of past yields and used to infer underlying soil conditions, while being exogenous to current input decisions. A potential concern is that these variables may directly affect input use through persistent effects on true soil fertility or household resources. We mitigate this by focusing on temperature variability, which is less likely to induce lasting changes in soil characteristics than other climatic factors, and by controlling for agricultural income shares to account for persistent economic effects. To the extent that any direct channels remain, our estimates should be interpreted as conservative, capturing both agronomic and belief-driven responses.

Specifically, we model outcomes B_i^L and B_i^H as a function of instruments (σ_i^{temp} and insect_i) and socioeconomic and demographic controls contained in vector $\vec{Z}_i$. $\vec{Z}_i$ consists of (a) farmer demography: education, caste and household size; (b) economic situation: income share from agriculture, property rights and land size (Berazneva et al. 2018) and (c) irrigation availability (Desbiez et al. 2004; Moges and Holden 2007). In the first stage:

$$B_i^L = \alpha_0 + \alpha_1 \sigma_i^{temp} + \alpha_2 \text{insect}_i + \vec{Z}_i' \vec{\lambda} + v_i \qquad (2)$$

We use the predicted over-perception probability $\hat{B}_i^L$ from model (2) to then estimate a model for nitrogen application decision N_i as a function of $\hat{B}_i^L$, $price_{i,t-5}$ and $\vec{Z}_i$.

$$N_i = \delta_0 + \delta_1 \hat{B}_i^L + \delta_2 price_{i,t-5} + \vec{Z}_i' \vec{\omega} + \upsilon_i \qquad (3)$$

$\hat{\delta}_1$ is the change in expected nitrogen application associated with a marginal increase in the predicted probability of over-perceiving soil fertility, conditional on price and other controls. δ_2 and $\vec{\omega}$ are the coefficients associated with the relevant covariates. We estimate analogous specifications for under-perception of soil fertility and phosphorus and potassium application (please see Results).

5.2.5 Assessing how past fertilizer application impacts soil fertility misperceptions:

To examine the reverse relationship i.e., the impact of nutrient application decisions on soil fertility misperceptions, we instrument farmers' nutrient application decisions using lagged, farmer-reported rice output prices (at the farmgate), averaged over the previous five years i.e., $price_{i,t-5}$. The identifying variation arises from the fact that higher output prices increase expected returns to cultivation and thereby incentivize greater input use, independently of prior beliefs about soil fertility (Mundlak 2001). The exclusion restriction requires that lagged prices influence soil fertility perceptions only through this input–yield channel, and not directly. This is plausible as output prices may not convey agronomic information about underlying soil conditions. For example, Woodard (2016) demonstrates that soil type significantly affects crop yield risk, yet this variation is not captured in the United States Department of Agriculture crop insurance premium ratings which rely on aggregated data for output prices and crop yields.

We model fertilizer application N_i as a function of the instrument $price_{i,t-5}$ which captures the expected returns to cultivation and thereby incentivizes greater input use independently of soil fertility beliefs, while controlling for farmer socioeconomic information in control vector $\vec{Z}_i$:

$$N_i = \psi_0 + \psi_1 price_{i,t-5} + \vec{Z}_i'\vec{\eta} + \zeta_i \qquad (4)$$

We use the predicted fertilizer application $\hat{N}_i$ to further estimate a model for soil fertility over-perceptions that we specify as follows:

$$\ln\left(\frac{\Pr(B_i^L = \{L \to H\}}{\Pr(B_i^L = \{L \to L\}}\right) = \mu_0 + \mu_1 \hat{N}_i + \vec{Z}_i'\vec{\phi} + \iota_i \qquad (5)$$

$\hat{\mu}_1$ measures the change in the log-odds of over-perceiving soil fertility relative to perceiving it consistently, associated with a one-unit increase in predicted nitrogen application, holding other covariates constant. μ_1 and $\vec{\phi}$ are the coefficients associated with the relevant covariates. We estimate analogous specifications for phosphorus and potassium application.

5.3 Results

This section reports results on the determinants of soil quality perceptions (fertility and texture), documents the prevalence of misperceptions (Hypothesis 1), and examines the linkage between soil fertility misperceptions and fertilizer (i.e., nitrogen, phosphorus and potassium) application decisions (Hypothesis 2). We assess Hypothesis 1 separately for soil texture and fertility perceptions; however, given these results, the subsequent analysis focuses on soil fertility. For Hypothesis 2, we present results for both IIA and IIB in the case of over-perceptions, while for under-perceptions we are limited to IIA due to instrument relevance concerns.

5.3.1 Drivers of farmers' soil quality perceptions

Multivariate link between farmers' soil perceptions and data-based evidence

Table 2 reports estimation results from the ordered logistic models for farmers' soil fertility and soil texture perceptions. We find relatively greater consistently in how farmers perceive soil texture than they perceive soil fertility when compared with the corresponding data-based measurements. Specifically, a marginal increase in soil clay content is associated with significantly lower odds that farmers perceiving 'light' texture soils relative to 'heavy' soils. Similarly, marginally higher soil nitrogen levels is associated with weaker odds of perceiving 'low' fertility relative to 'high' fertility. In contrast, organic carbon acts as a mixed signal of soil fertility, in that results vary substantially across model specifications. Our findings align with the existing agronomic evidence that farmers more accurately perceive nutrients they directly manage (e.g., nitrogen fertilizers) whereas less salient properties like organic carbon are inferred indirectly through yield (Liebig and Doran 1999; Barrios and Trejo 2003; Berazneva et al. 2018; Delgado and Stoorvogel 2020; Lal et al. 2024).

Geographic and socioeconomic regression covarites provide some interesting insights. Households having irrigation access and greater income share from agriculture tend to positively perceive soil fertility on their farms, which is logical given that farming conditions and economic returns are often found to influence farmers' belief systems (Moges and Holden 2007; Berazneva et al., 2018). Education and caste identities also shape soil perceptions. More educated farmers and those from forward castes tend to report higher soil fertility, even after controlling for objective data. Landownership is negatively associated with perceived soil fertility, although this result is not robust across model specifications. Local climate exhibits contrasting effects on farmers' soil perceptions. Absent geographic controls, long-term temperature mean and variability terms are found to negatively influence fertility perceptions, which makes sense in light of the evidence that links heat stress to soil degradation and crop yield risk (Hansen et al., 2018; Arora et al., 2020). However, these effects attenuate with the inclusion of geographic controls, indicating that spatial heterogeneity absorbs a substantial share of climate-related variation in our sample.

In light of the results that farmer perceptions of soil texture are on average consistent with soil clay measurements but soil fertility perceptions may differ across indicators of soil quality, in the subsequent analysis we lay focus on soil fertility (mis)perceptions.

5.3.2 Soil Fertility *Mis*perceptions

Articulating misperceptions: soil fertility perceptions versus soil nitrogen levels

Tables 3A-B present cross-tabulations between farmers' survey-based soil fertility perceptions ($\tilde{F}_i$) and data-driven nitrogen (N_i^S) and organic carbon (O_i^S) measurements, respectively. These cross-tabulations, representing joint distribution of perceptions and data-based evidence, are made possible by dividing the continuous data-based measurements N_i^S

and O_i^S into three equal sized groups (or tertiles) of 'Low', 'Medium' and 'High' levels (i.e., in tune with the perceptions categories in survey data). In particular, soil nitrogen data are categorized as: {Low, Medium, High}, and soil organic carbon perceptions are categorized as: {Low, Medium, High}. The diagonal elements of this joint distribution provide frequency of consistent farmer perceptions whereas off-diagonal elements provide frequency of inconsistent or biased perceptions when compared with scientific evidence. . We find 'Medium' category perceptions to occur most frequently in sample records, i.e., 7,290 out of 8,281 farmers (~88%), which we classify as anchoring bias that arises due to a coarse Likert scale based inquiry of farmer perceptions.

Overall, Tables 3A-B reveal that majority farmers misperceive soil fertility conditions relative to data-based evidence. Only 34.3% fertility perception categories match with the corresponding soil nitrogen categories, and only 33.4% of these perception categories match with soil organic carbon data. Moreover, we find 97% misperceived soil quality among 2,667 'Low' soil nitrogen farms, out of which 2,513 (94%) farms fall in 'Medium' fertility category. Similarly, about 89% misperception instances are recorded on 2,767 farms with 'High' nitrogen levels, out of which we 2,513 (84%) farmer reported 'Medium' fertility soils. On 2,565 farms with 'Medium' soil nitrogen levels, 2,455 farmers consistently perceive fertility category while 198 and 112 farmers identify their lands as 'high' and 'low' fertility, respectively. We find similar distributional patterns between soil organic carbon categories and farmers' fertility perception reports.

We focus on the directionally meaningful misperceptions: (1) over-perceptions, i.e., farmers whose perception categories outmatch both data-driven soil nitrogen as well as organic carbon levels; and (2) under-perceptions, i.e., perceptions underestimate data-driven soil nitrogen and organic carbon levels. We observe that over-perceptions are more prevalent in our sample than are under-perceptions, even though both groups are small in absolute size relative to the full sample. Notably, large divergence between farmer perceptions and objective soil quality measures is not unexpected given that fertility is a complex, multidimensional and dynamic attribute that is not directly observed by farmers (Lal 1993; Antle et al. 2006; Barrett 2008; Moebius-Clune et al. 2011). Farmers often infer fertility from noisy and lagged signals such as crop yields (Moges and Holden 2007; Karltun et al. 2013), which are themselves shaped by multitude factors (alongside soils) such as inputs, weather variability, and management practices. In addition, socioeconomic and information constraints can influence how farmers interpret such signals (Marenya et al. 2008; Essougong et al. 2020; Martey and Kuwornu 2021; Tesfahugnen et al. 2021; Alkaabi et al. 2025).

5.3.3 Linking soil fertility misperceptions and fertilizer application decisions

We recognise a two-way relationship between farm-level soil fertility misperceptions and farmers' input use decisions. Beliefs about soil fertility likely play a central role in shaping input application and production choices (Ervin and Ervin 1982; Tittonell et al. 2005; Marenya et al. 2008; Zhou et al. 2020; Delgado and Stoorvogel 2022). On the other hand, these beliefs may themselves be influenced by past production outcomes such as farmers mis-attributing yield outcomes to soil conditions even though other factors may be at play (Miller and Ross 1975; Mezulis et al. 2004; Castagnetti and Schmacker 2022). Since productivity is a function of prior fertilization, misperceptions about soil quality may be attributed to behavioral aspects of farmer decision-making (Benabou and Tirole 2002; Arora and Feng, 2024). We assess the linkage between soil fertility misperceptions and input decisions in two steps: (1) two-way contingency tables allow us to test whether soil perceptions are on average independent of input application decisions; and (2) 2SLS regression based on instrumental variable approach to identify two-sided causal relationships.

Table 4 reports the relationship between fertilizer application decisions and soil fertility misperceptions. We distinguish between consistent perceptions and biased perceptions, where the latter capture directionally meaningful misperceptions, i.e., over- or under-perceptions depending on underlying soil fertility category. Based on statistically significant chi-squared statistics across all nutrient-types and categories, we reject the hypothesis that farmers' input application decisions are independent of their soil fertility perceptions. When data-based measures indicate 'low' soil fertility and farmers apply 'high' fertilizer (NPK) levels, then these farmers tend to overperceive soil fertility relative to those who apply 'low' fertilizer levels on less fertile lands. Conversely, 'low' fertilizer application on nitrogen rich soils is frequently associated with under perception of soil fertility. Therefore, farm-level input application seems to be systematically correlated with the direction of soil fertility misperceptions. Below we move beyond descriptive results to investigate two-way causal relationship: whether soil fertility beliefs shape input use, and/or input use decisions influence soil fertility beliefs.

Do soil fertility misperceptions influence fertilizer use decisions?

Tables 5A-5C present the results from 2SLS framework to establish that farmers' soil fertility misperceptions systematically drive their fertilizer application decisions. In Table 5A (first-stage regression), we find that long-term temperature variability is associated with a higher likelihood of over-perceiving soil fertility and a lower likelihood of under-perceiving it. This is in line with our findings from the perceptions models in Table 2. Further, past pest incidence is related to higher (lower) odds of over-perception (under-perception) relative to consistent soil fertility beliefs, which also aligns with evidence that frequent infestation indicates biologically active (fertile) soils (Lavelle et al. 2006).

Tables 5B-C provides evidence on causal impact of soil fertility misperceptions on nutrient application decisions of farmers. While OLS results in positive (negative), statistically-significant association between over- (under-)perception of soil fertility and fertilizer application, the IV regression estimates are reverse the sign across all nutrient types found in fertilizer inputs. The instrument relevance condition is satisfied everywhere and overidentification test fails to reject instrument validity for all nutrients but potassium. Therefore, we indicate caution in interpreting causal linkage with potassium applciaiton, even though the impact sign is consistent with nitrogen and phosphorous nutrients.

Our IV estimates reveal that farmers who over-perceive their soil as more fertile than it is objectively, apply significantly less fertilizer than their consistently-perceiving counterparts. This is consistent with understanding that an optimistic belief about soil fertility suppresses the perceived need for soil amendment leading to systematic under-investment in inputs. On the other hand, soil fertility under-perceptions too negatively influence nutrient application decisions. Farmers who underperceive soil quality likely exhibit pessimistic marginal returns on degraded land and suppress nutrient application (Tittonell et al. 2005; Berazneva et al. 2018). This generates an inefficiency trap in which pessimistic soil beliefs and low input use are mutually reinforcing (Barrett 2002). Such directional associations have been documented in the past (Tittonell et al. 2005; Zhou et al. 2020), however we contribute to this literature through first instance of causal linkage between perceptions and farm management decisions.

The magnitude of these effects is also economically meaningful. Over-perception of soil fertility leads nitrogen application reduction by 180 kg/acre and phosphorus by 320 kg/acre relative to those who accurately perceive soil quality. For the case of under-perceived soil fertility potassium application is reduced by 600 kg/acre. These behavioural impacts on farm management decisions represent substantial departures from average levels in Table 1.

Can fertilizer use decisions lead to systematically misperceived soil fertility?

Tables 6A-6B present the results from 2SLS framework to establish that farmers' soil fertility misperceptions systematically drive their fertilizer application decisions. Table 6A presents the first stage regression of fertilizer application on lagged output price for the over-perceiving and under-perceiving subsamples separately. For the over-perceiving subsample, lagged price is a negative and statistically significant predictor of fertilizer application for all three nutrients satisfying the instrument relevance condition. The negative sign however is counter-intuitive and requires further examination since it indicates that higher output prices in the past are associated with lower current fertilizer application. For the under-perceiving subsample, the instrument fails the relevance condition for all three nutrients, with coefficients close to zero and statistically insignificant.

Table 6B examines the relationship between nutrient application decisions and soil fertility over-perceptions. We report probit estimates first, without adjusting for the endogeneity between fertilizer application and misperceptions, and then IV estimates for each nutrient. The probit estimates indicate a positive and statistically significant association between fertilizer application and over-perception across all three nutrients, implying that farmers who apply more fertilizer are also more likely to overestimate soil fertility. However, this correlation is not causal, since fertilizer use and soil fertility beliefs are jointly determined. By contrast, the IV estimates are negative for all three nutrients. This reversal is consistent with a motivated beliefs mechanism. Farmers who invest heavily in fertilizer and subsequently obtain good yields may credit these outcomes to their own management decisions rather than to favourable soil conditions, thereby reducing over-perceptions of soil fertility. On the other hand, when high input use is not matched by strong yield performance, farmers may interpret the outcome as evidence of poor soil quality, which can strengthen under-perceptions. Overall, our results suggest that nutrient application decisions can shape soil fertility perceptions through both actual production experience and the way farmers interpret observed signals of soil fertility.

5.4 Discussion and Conclusions

Farmers' soil fertility perceptions are recognized as a critical input for designing effective land management policy (Niemijer and Mazzucato 2003; Essougong et al. 2020; Delgado and

Stoorvogel 2022). However, such perceptions are likely based on locally observed and noisy signals such as crop yields and within-season input responsiveness (Marenya et al. 2008; Karltun et al. 2013)). There is recognition that farmers draw an understanding of complex soil process through local context and indigenous knowledge systems (Liebig and Doran 1999; Gruver and Weil 2007; Berazneva et al. 2018); climate variability (Martey and Kuwornu 2021; Tesfahugnen et al. 2021); peer effects and extension services (Green and Heffernan 1987; Elwell et al. 2018; Wang et al. 2020; Kolady et al. 2021; Alkaabi et al. 2025).

Overall, it is not surprising that soil perceptions may differ from scientific evidence (Liebig and Doran 1999; Murage et al. 2000; Barrios and Trejo 2003; Andrews et al. 2003; Gruver and Weil 2007). However, some important questions arise, such as how large is the *divergence*s? How are soil perceptions formed and whether they are systematic in the manner that they diverge from data? And, how might misperceptions influence farm management decisions? We addressed these questions in this paper by employing large-scale, geocoded farm surveys in conjunction with finely gridded soil maps; amounting to first-of-a-kind contribution to a fairly comprehensive literature on soil perceptions.

We find that soil texture, which is a physical attribute of soil quality, is perceived consistently by Indian rice growers in our sample whereas majority of farmers misperceive soil fertility relative to data-based measures such as soil nitrogen and organic carbon levels. Over-perceptions of soil fertility are found to be more prevalent than under-perceptions. We further recognize a two-way causal relationship: soil quality misperception impact farm fertilization decision; and these decisions influence soil quality misperceptions, and employ a IV based regression framework to produce evidence for these linkages. Over-perceptions as well as under-perception of soil fertility leads to systematic reduction in fertilizer use, consistent with correlational evidence; corroborated by past evidence in Tittonell et al. (2005) and Marenya et al. (2008). Higher fertilization reduces soil fertility over-perceptions consistent with self-serving misattribution in belief formation: when farmers achieve better yields through greater fertilizer use, they may attribute the outcome to their own input choices rather than to underlying soil quality, making them less likely to overestimate soil fertility (Mezulis et al. 2004).

Several limitations merit consideration. First, our classification of soil fertility into ‘low’, ‘medium’ and ‘high’ categories is derived from country-wide equi-sized thresholds that mask regional heterogeneity in soil quality, possibly over-attributing the role of ‘medium’ category soil data in analysis. As a consequence, the same country-wide percentile may correspond to diverse agronomic conditions depending on soil types, cropping systems, and nutrient use (Srivastava et al. 2015; Preez et al. 2025) and lead to misestimation of the extent of farmer misperceptions. We recommend developing spatially adaptive soil quality benchmarks as part of future research agenda. Second, the use of gridded soil data can introduce measurement errors due to spatial interpolation (Cressie 1992). Incorporating data based on denser, country-level ground truth, for example, in the Indian setting using sensor-based soil quality reported under the Soil Health Card scheme (Morton et al. 2023), can complement our work.

Third, our analysis abstracts from the full complexity of belief formation which is shaped by cognitive, social, and institutional factors that lie beyond the scope of this study. As economists, we focus on shedding light on potential economic channels for how farming decisions may interact with soil fertility perceptions. In addition, our causal analysis is based on cross-sectional data, which limits the extent to which dynamic belief formation (Benabou and Tirole 2002) can be explicitly modelled.

While our instrumental variable strategies provide evidence of plausible causal channels in both directions, they do not fully resolve the temporal sequencing between input decisions and belief updating. Nevertheless, our results suggest that land management policies based solely on scientifically measured soil quality may miss the full picture. Misperceptions are not merely passive informational gaps, but appear to be linked with farmers' economic decisions. Recognizing this can help reorient extension services which form a central pillar of agricultural policy in developing countries and bridge the gap between farmers' soil quality perceptions and corresponding data-based measurement.

5.5 References

Ajay, Anurag, Peter Craufurd, Virender Kumar, Arindam Samaddar, Rk Malik, Sachin Sharma, Harshit Ranjan, et al. 2022. "Large Survey Dataset of Rice Production Practices Applied by Farmers on Their Largest Farm Plot During 2018 in India." Data in Brief 45 (September): 108625. https://doi.org/10.1016/j.dib.2022.108625.

Alkaabi, Khaula, Muhammad Touseef Anjum, Asadik Hamed, and Omar Younes. 2025. "Modeling Farmers' Perception of Agricultural Practices and Soil Quality in Arid Regions of UAE: A PLS-SEM Approach to Sustainable Land Use." Environmental Challenges 20 (August): 101283. https://doi.org/10.1016/j.envc.2025.101283.

Antle, John M., Jetse J. Stoorvogel, and Roberto O. Valdivia. 2006. "Multiple Equilibria, Soil Conservation Investments, and the Resilience of Agricultural Systems." Environment and Development Economics 11 (4): 477–92. https://www.jstor.org/stable/44379116.

Arora, Gaurav, and Hongli Feng. "Assessing Systematic Biases in Farmers' Local Weather Change Perceptions." *Scientific Reports 14*, no. 1 (2024): 26641. Accessed April 15, 2026. https://doi.org/10.1038/s41598-024-76327-8.

Arora, Gaurav, Hongli Feng, Christopher J. Anderson, and David A. Hennessy. "Evidence of Climate Change Impacts on Crop Comparative Advantage and Land Use." *Agricultural Economics 51*, no. 2 (2020): 221-236. Accessed April 14, 2026. https://doi.org/10.1111/agec.12551.

Asafu-Adjaye, John. 2008. "Factors Affecting the Adoption of Soil Conservation Measures: A Case Study of Fijian Cane Farmers." *Journal of Agricultural and Resource Economics* 33 (1): 99–117. https://www.jstor.org/stable/41220616.

Barbier, Edward B., and Jacob P. Hochard. 2016. "Does Land Degradation Increase Poverty in Developing Countries?" *PLoS ONE* 11 (5): e0152973. https://doi.org/10.1371/journal.pone.0152973.

Barrett, Christopher B., and Leah E. M. Bevis. 2015. "The Self-reinforcing Feedback Between Low Soil Fertility and Chronic Poverty." *Nature Geoscience* 8 (12): 907–12. https://doi.org/10.1038/ngeo2591.

Barrios, E., and M.T. Trejo. 2002. "Implications of Local Soil Knowledge for Integrated Soil Management in Latin America." *Geoderma* 111 (3–4): 217–31. https://doi.org/10.1016/s0016-7061(02)00265-3.

Baveye, Philippe C., Jacques Baveye, and John Gowdy. 2016. "Soil 'Ecosystem' Services and Natural Capital: Critical Appraisal of Research on Uncertain Ground." *Frontiers in Environmental Science* 4 (June). https://doi.org/10.3389/fenvs.2016.00041.

Benabou, R., and J. Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics* 117 (3): 871–915. https://doi.org/10.1162/003355302760193913.

Berazneva, Julia, Linden McBride, Megan Sheahan, and David Güereña. 2018. "Empirical Assessment of Subjective and Objective Soil Fertility Metrics in East Africa: Implications for Researchers and Policy Makers." *World Development* 105 (January): 367–82. https://doi.org/10.1016/j.worlddev.2017.12.009.

Blum, W. E. H. 1988. "Problems of Soil Conservation." In *Nature and Environment Series*, vol. 39. Strasbourg: Council of Europe, Steering Committee for the Conservation and Management of the Environment and Natural Habitats (CDPE)

Bolliger, Adrian, Jakob Magid, Jorge Carneiro Telmo Amado, Francisco Skóra Neto, Maria De Fatima Dos Santos Ribeiro, Ademir Calegari, Ricardo Ralisch, and Andreas De Neergaard. 2006. "Taking Stock of the Brazilian 'Zero-Till Revolution': A Review of Landmark Research and Farmers' Practice." *Advances in Agronomy*, January, 47–110. https://doi.org/10.1016/s0065-2113(06)91002-5.

Castagnetti, Alessandro, and Renke Schmacker. 2022. "Protecting the Ego: Motivated Information Selection and Updating." *European Economic Review* 142 (January): 104007. https://doi.org/10.1016/j.euroecorev.2021.104007.
Chen, Xi, Mingzhe Pu, and Yu Zhong. 2022. "Evaluating China Food's Fertilizer Reduction and Efficiency Initiative Using a Double Stochastic Meta-Frontier Method." *International Journal of Environmental Research and Public Health* 19 (12): 7342. https://doi.org/10.3390/ijerph19127342.
Coggins, Sam, Andrew J. McDonald, João Vasco Silva, Anton Urfels, Hari Sankar Nayak, Sonam Rinchen Sherpa, Mangi Lal Jat, et al. 2025. "Data-driven Strategies to Improve Nitrogen Use Efficiency of Rice Farming in South Asia." *Nature Sustainability* 8 (1): 22–33. https://doi.org/10.1038/s41893-024-01496-3.
Collins, Harold P., Maysoon M. Mikha, Tabitha T. Brown, Jeffrey L. Smith, David Huggins, and Upendra M. Sainju. 2012. "Agricultural Management and Soil Carbon Dynamics." In *Elsevier eBooks*, 59–77. https://doi.org/10.1016/b978-0-12-386897-8.00005-x.
Cressie, N. A. C. (1992). *Statistics for Spatial Data* (Revised ed.). New York, NY: John Wiley & Sons.
Delgado, Luciana, and Jetse J. Stoorvogel. 2022. "Role of Soil Perception and Soil Variability by Smallholder Farmers in the Low Adoption Rates of Extension Packages in Central America." *Journal of Rural Studies* 93 (May): 92–103. https://doi.org/10.1016/j.jrurstud.2022.05.009.
Diersch, Nadine, and Eva Walther. 2016. "The Impact of Question Format, Context, and Content on Survey Answers in Early and Late Adolescence." *Journal of Official Statistics* 32 (2): 307–28. https://doi.org/10.1515/jos-2016-0018.
Desbiez, Arnaud, Robin Matthews, Bhaba Tripathi, and Jim Ellis-Jones. 2004. "Perceptions and Assessment of Soil Fertility by Farmers in the Mid-hills of Nepal." *Agriculture Ecosystems & Environment* 103 (1): 191–206. https://doi.org/10.1016/j.agee.2003.10.003.
Douven, Igor. 2017. "A Bayesian Perspective on Likert Scales and Central Tendency." *Psychonomic Bulletin & Review* 25 (3): 1203–11. https://doi.org/10.3758/s13423-017-1344-2.
Essougong, Urcil P. Kenfack, Maja Slingerland, Syndhia Mathé, Wouter Vanhove, Precillia I. Tata Ngome, Philippe Boudes, Ken E. Giller, Lotte S. Woittiez, and Cees Leeuwis. 2020. "Farmers' Perceptions as a Driver of Agricultural Practices: Understanding Soil Fertility Management Practices in Cocoa Agroforestry Systems in Cameroon." *Human Ecology* 48 (6): 709–20. https://doi.org/10.1007/s10745-020-00190-0.
Ervin, Christine A., and David E. Ervin. 1982. "Factors Affecting the Use of Soil Conservation Practices: Hypotheses, Evidence, and Policy Implications." *Land Economics* 58 (3): 277. https://doi.org/10.2307/3145937.
Gray, Leslie C., and Philippe Morant. 2002. "Reconciling Indigenous Knowledge With Scientific Assessment of Soil Fertility Changes in Southwestern Burkina Faso." *Geoderma* 111 (3–4): 425–37. https://doi.org/10.1016/s0016-7061(02)00275-6.
Foster, Andrew D., and Mark R. Rosenzweig. 1995. "Learning by Doing and Learning From Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103 (6): 1176–1209. https://www.jstor.org/stable/2138708.
Giller, Ken E., Ernst Witter, Marc Corbeels, and Pablo Tittonell. 2009. "Conservation Agriculture and Smallholder Farming in Africa: The Heretics' View." *Field Crops Research* 114 (1): 23–34. https://doi.org/10.1016/j.fcr.2009.06.017.
Green, Gary P., and William D. Heffernan. 1987. "Soil Erosion and Perception of the Problem." *Journal of Rural Studies* 3 (2): 151–57. https://doi.org/10.1016/0743-0167(87)90030-1.

Gruver, J.B., and R.R. Weil. 2007. “Farmer Perceptions of Soil Quality and Their Relationship to Management-sensitive Soil Parameters.” *Renewable Agriculture and Food Systems* 22 (4): 271–81. https://doi.org/10.1017/s1742170507001834.

Hoffman, Oliver, Christopher E. Chorpenning, Tad Trimarco, Nathan Carr, Cassidy Buchanan, and James A. Ippolito. 2025. “Conservation Reserve Program Soils Show Potential as a Soil Health Benchmark—A Southern Minnesota Case Study.” *Soil Systems* 9 (2): 46. https://doi.org/10.3390/soilsystems9020046.

Hobbs, Peter R, Ken Sayre, and Raj Gupta. 2007. “The Role of Conservation Agriculture in Sustainable Agriculture.” *Philosophical Transactions of the Royal Society B Biological Sciences* 363 (1491): 543–55. https://doi.org/10.1098/rstb.2007.2169.

Johnson-Laird, P. N. (1983). Mental Models. Towards a Cognitive Science of Language, Inference and Consciousness. Cambridge, UK: Cambridge University Press.

Karltun, E., M. Lemenih, and M. Tolera. 2011. “COMPARING FARMERS’ PERCEPTION OF SOIL FERTILITY CHANGE WITH SOIL PROPERTIES AND CROP PERFORMANCE IN BESEKU, ETHIOPIA.” *Land Degradation and Development* 24 (3): 228–35. https://doi.org/10.1002/ldr.1118.

Kerr, John, and N. K. Sanghi. "Indigenous soil and water conservation in India's semi-arid tropics." (1992): 28-pp.

Kolady, Deepthi, Weiwei Zhang, Tong Wang, and Jessica Ulrich-Schad. 2020. “Spatially Mediated Peer Effects in the Adoption of Conservation Agriculture Practices.” *Journal of Agricultural and Applied Economics* 53 (1): 1–20. https://doi.org/10.1017/aae.2020.24.

Krah, Kwabena, Hope Michelson, Emilie Perge, and Rohit Jindal. 2019. “Constraints to Adopting Soil Fertility Management Practices in Malawi: A Choice Experiment Approach.” *World Development* 124 (August): 104651. https://doi.org/10.1016/j.worlddev.2019.104651.

Krogh, Lars, and Bjarke Paarup-Laursen. 1997. “Indigenous Soil Knowledge Among the Fulani of Northern Burkina Faso: Linking Soil Science and Anthropology in Analysis of Natural Resource Management.” *GeoJournal* 43 (2): 189–97. https://doi.org/10.1023/a:1006894818609.

Lal, Bharat, Ashutosh Sharma, Ashish Kumar Gupta, Sati Sankar Singh, Susheel Kumar Chaturvedi, and Manoj Kumar Jhariya. 2024. “Soil Fertility Evaluation Through Farmer Knowledge and Scientific Approaches of Bundelkhand Region, Central India.” *Environment Development and Sustainability*, November. https://doi.org/10.1007/s10668-024-05729-5.

Lele, Uma. 1989. “Sources of Growth in East African Agriculture.” *The World Bank Economic Review* 3 (1): 119–44. https://doi.org/10.1093/wber/3.1.119.

Liebig, Mark A., and John W. Doran. 1999. “Evaluation of Farmers’ Perceptions of Soil Quality Indicators.” *American Journal of Alternative Agriculture* 14 (1): 11–21. https://doi.org/10.1017/s0889189300007967.

Marenya, Paswel, Christopher B. Barrett, and Trudi Gulick. 2008. “Farmers’ Perceptions of Soil Fertility and Fertilizer Yield Response in Kenya.” *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.1845546.

Martey, Edward, and John K.M. Kuwornu. 2020. “Perceptions of Climate Variability and Soil Fertility Management Choices Among Smallholder Farmers in Northern Ghana.” *Ecological Economics* 180 (October): 106870. https://doi.org/10.1016/j.ecolecon.2020.106870.

Matson, Amanda, Maria Fantappiè, Grant A. Campbell, Jorge F. Miranda-Vélez, Jack H. Faber, Lucas Carvalho Gomes, Rudi Hessel, et al. 2024. “Four Approaches to Setting Soil

Health Targets and Thresholds in Agricultural Soils." *Journal of Environmental Management* 371 (November): 123141. https://doi.org/10.1016/j.jenvman.2024.123141.
McCullagh, Peter. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society: Series B (Methodological) 42*, no. 2 (1979): 109-127. Accessed April 15, 2026. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x.
McNeill, J. R., and Verena Winiwarter. 2004. "Breaking the Sod: Humankind, History, and Soil." *Science* 304 (5677): 1627–29. https://doi.org/10.1126/science.1099893.

Meijer, Seline S., Delia Catacutan, Oluyede C. Ajayi, Gudeta W. Sileshi, and Maarten Nieuwenhuis. 2014. "The Role of Knowledge, Attitudes and Perceptions in the Uptake of Agricultural and Agroforestry Innovations Among Smallholder Farmers in sub-Saharan Africa." *International Journal of Agricultural Sustainability* 13 (1): 40–54. https://doi.org/10.1080/14735903.2014.912493.
Meissner, Luise, and Oliver Musshoff. "Revisiting the relationship between farmland prices and soil quality." *Q Open* 2, no. 2 (2022): qoac017.
Mezulis, Amy H., Lyn Y. Abramson, Janet S. Hyde, and Benjamin L. Hankin. "Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias." *Psychological bulletin* 130, no. 5 (2004): 711.
Miller, Dale T., and Michael Ross. 1975. "Self-serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin* 82 (2): 213–25. https://doi.org/10.1037/h0076486.
Moebius-Clune, B. N., H. M. Van Es, O. J. Idowu, R. R. Schindelbeck, J. M. Kimetu, S. Ngoze, Johannes Lehmann, and J. M. Kinyangi. "Long-term soil quality degradation along a cultivation chronosequence in western Kenya." *Agriculture, Ecosystems & Environment* 141, no. 1-2 (2011): 86-99.
Moges, Awdenegest, and N. Mf Holden. "Farmers' perceptions of soil erosion and soil fertility loss in Southern Ethiopia." *Land Degradation & Development* 18, no. 5 (2007): 543-554.
Mondal, Pinaki, and Manisha Basu. 2009. "Adoption of Precision Agriculture Technologies in India and in Some Developing Countries: Scope, Present Status and Strategies." *Progress in Natural Science Materials International* 19 (6): 659–66. https://doi.org/10.1016/j.pnsc.2008.07.020.
Montgomery, David R. *Dirt: The Erosion of Civilizations*. 2nd ed. University of California Press, 2012. http://www.jstor.org/stable/10.1525/j.ctt1pnhp3.
Morecroft, John Dw. 1983. "System Dynamics: Portraying Bounded Rationality." *Omega* 11 (2): 131–42. https://doi.org/10.1016/0305-0483(83)90002-6.
Morton, Claire M., Hemant Pullabhotla, Leah Bevis, and David B. Lobell. 2023. "Soil Micronutrients Linked to Human Health in India." *Scientific Reports* 13 (1): 13591. https://doi.org/10.1038/s41598-023-39084-8.
Mundlak, Y. (2001). Chapter 1 Production and supply. In *Handbook of agricultural economics* (pp. 3–85). https://doi.org/10.1016/s1574-0072(01)10004-6
Murage, Evah W, Nancy K Karanja, Paul C Smithson, and Paul L Woomer. 2000. "Diagnostic Indicators of Soil Quality in Productive and Non-productive Smallholders' Fields of Kenya's Central Highlands." *Agriculture Ecosystems & Environment* 79 (1): 1–8. https://doi.org/10.1016/s0167-8809(99)00142-5.
Niemeijer, David, and Valentina Mazzucato. 2002. "Moving Beyond Indigenous Soil Taxonomies: Local Theories of Soils for Sustainable Development." *Geoderma* 111 (3–4): 403–24. https://doi.org/10.1016/s0016-7061(02)00274-4.

Nord, Alison, and Sieglinde Snapp. 2020. “Documentation of Farmer Perceptions and Site-specific Properties to Improve Soil Management on Smallholder Farms in Tanzania.” *Land Degradation and Development* 31 (15): 2074–86. https://doi.org/10.1002/ldr.3582.
Ofstehage, Andrew, and Ryan Nehring. 2021. “No-till Agriculture and the Deception of Sustainability in Brazil.” *International Journal of Agricultural Sustainability* 19 (3–4): 335–48. https://doi.org/10.1080/14735903.2021.1910419.
Parikh, S. J. & James, B. R. (2012) Soil: The Foundation of Agriculture. *Nature Education Knowledge* 3(10):2. https://www.nature.com/scitable/knowledge/library/soil-the-foundation-of-agriculture-84224268/#:~:text=Throughout%20human%20history%2C%20our%20relationship,as%20the%20foundation%20of%20agriculture.&text=Human%20society%20has%20developed%20through,in%20our%20agriculturally%2Ddependent%20society.
Perrings, Charles. "An optimal path to extinction?: Poverty and resource degradation in the open agrarian economy." *Journal of Development Economics* 30, no. 1 (1989): 1-24.
Pingali, Prabhu L. "Green revolution: impacts, limits, and the path ahead." *Proceedings of the national academy of sciences* 109, no. 31 (2012): 12302-12308.
Preez, Gerhard du, Ané Loggenberg, Driekie Fourie, João Marcelo-Silva, Tvisha Martin, Dimakatso Ramphisa-Nghondzweni, Hendrik Smith, and Christine Sprunger. "Context matters: soil ecosystem status varies across diverse conservation agriculture systems." *Journal of Soil Science and Plant Nutrition* 25, no. 2 (2025): 2576-2589.
Rogers, E.M. (1995) Diffusion of Innovations. 4th Edition, the Free Press, New York.
Thanuja, Adama, Bhavana Dori, Chandan Kumar Panigrahi, Saransh Kumar Gautam, Jyoti Gawaria, Satyabrata Sarangi, Ganesh G. Chaware, and Ishanu Mandal. "Ecology of soil-dwelling insects and their influence on crop health: A review." *Int. J. Environ. Clim. Change* 15 (2025): 285-296.
Shanmugam, Ramalingam, and Rajan Chattamvelli. 2015. *Statistics for Scientists and Engineers*. https://doi.org/10.1002/9781119047063.
Nielsen, Uffe N., Diana H. Wall, and Johan Six. 2015. “Soil Biodiversity and the Environment.” *Annual Review of Environment and Resources* 40 (1): 63–90. https://doi.org/10.1146/annurev-environ-102014-021257.
Smith, Pete, Rosa M. Poch, David A. Lobb, Ranjan Bhattacharyya, Ghiath Alloush, Gaius D. Eudoxie, H. C. Anjos et al. "Status of the World's Soils." *Annual Review of Environment and Resources 49*, no. Volume 49, 2024 (2024): 73-104. Accessed April 7, 2026. https://doi.org/10.1146/annurev-environ-030323-075629.
Středová, Hana, Petra Fukalová, Filip Chuchma, Jan Haberle, and Tomáš Středa. "Nitrates directive restriction: To change or not to change in terms of climate change, that is the question." *Science of The Total Environment* 917 (2024): 170381.
Tesfahunegn, Gebreyesus Brhane, Elias T. Ayuk, and S. G. K. Adiku. "Farmers' perception on soil erosion in Ghana: Implication for developing sustainable soil management strategy." *Plos one* 16, no. 3 (2021): e0242444.
Thanuja, Adama, Bhavana Dori, Chandan Kumar Panigrahi, Saransh Kumar Gautam, Jyoti Gawaria, Satyabrata Sarangi, Ganesh G. Chaware, and Ishanu Mandal. "Ecology of soil-dwelling insects and their influence on crop health: A review." *Int. J. Environ. Clim. Change* 15 (2025): 285-296.
Tittonell, Pablo, Bernard Vanlauwe, P. A. Leffelaar, Edwin C. Rowe, and Ken E. Giller. "Exploring diversity in soil fertility management of smallholder farms in western Kenya: I. Heterogeneity at region and farm scale." *Agriculture, ecosystems & environment* 110, no. 3-4 (2005): 149-165.

Tversky, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." *Science 185*, no. 4157 (4157): 1124-1131. Accessed April 15, 2026. https://doi.org/1738360.

VERMA, SANTOSH. "Mapping the cost of cultivation of paddy and wheat in India: An analysis of CACP data." *International Journal of Social Sciences and Management Review* 6, no. 3 (2023): 101-108.

Wang, Geling, Qian Lu, and Sergio C. Capareda. "Social network and extension service in farmers' agricultural technology adoption efficiency." *Plos one* 15, no. 7 (2020): e0235927.

Zhou, Zhifang, Jinhao Liu, Huixiang Zeng, Tao Zhang, and Xiaohong Chen. "How does soil pollution risk perception affect farmers' pro-environmental behavior? The role of income level." Journal of environmental management 270 (2020): 110806.

Srivastava, Pankaj, Dilip Kumar Pal, K. Manini Aruche, Suhas P. Wani, and Kanwar L. Sahrawat. 2014. "Soils of the Indo-Gangetic Plains: A Pedogenic Response to Landscape Stability, Climatic Variability and Anthropogenic Activity During the Holocene." *Earth-Science Reviews* 140 (November): 54–71. https://doi.org/10.1016/j.earscirev.2014.10.010.

5.6 Tables

Variable Definitions	Acronym	N	Mean	SD	Median
Soil Fertility Perception $\{1: \text{Low Quality}, 2: \text{Medium Quality}, 3: \text{High Quality}\}$	$\tilde{F}_i$	8,281	2.04	0.34	2
Soil Texture Perception $\{1: \text{Light Texture}, 2: \text{Medium Texture}, 3: \text{Heavy Texture}\}$	$\tilde{T}_i$	8,281	2.06	0.46	2
Objective Soil Quality Measures					
Nitrogen (cg/kg)	N_i^S	8,281	137.7	35.22	131.5
Organic Carbon Stock (t/ha)	O_i^S	8,281	40.4	5.34	40.6
Clay Content (g/kg)	C_i^S	8,281	316.1	48.16	304.3
Socioeconomics and Demographic Characteristics					
Land Ownership (=1 if largest cultivated rice plot is owned, = 0 otherwise)	own_i	8,281	0.82	0.38	1
Lower Caste (=1 if surveyee identifies as Scheduled Caste or Scheduled Tribe, = 0 otherwise)	$caste_i$	8,281	0.19	0.4	1
Years of Education	edu_i	8,281	6.7	5.21	5
Agricultural Income Share (percentage)	inc_i	8,281	52.53	30.36	50
Irrigation (=1 if irrigation is available, = 0 otherwise)	irr_i	8,281	0.88	0.32	1
Size of Agricultural Household	$size_i$	8,281	0.38	0.22	0.33
Climate Variables					
Average Kharif temperature (1980-2017)	$temp_i$	8,281	28.98	0.72	29.07
Coefficient of variation of Kharif temperature (1980-2017)	σ_i^{temp}	8,281	1.17	0.16	1.18
Other Controls					
Lagged Rice Output Price (rupees/quintal)	$price_{i,t-5}$	8,281	1365.4	308.4	1300
Pest (= 1 if rice farmer experienced insect incidence, 0 otherwise)	insect_i	8,281	0.73	0.44	1
Input Application					
Nitrogen Applied (kg/acre)	N_i	8,281	70.7	149.9	16.56
Phosphorus Applied (kg/acre)	P_i	8,281	113.34	7.36	289.9
Potassium Applied (kg/acre)	K_i	8,281	80.98	1.54	239.94

Table 5.1: Summary Statistics

Table 5.2: Ordered Logistic Regression Results for Soil Fertility and Texture perceptions.

Explanatory Variables	$\tilde{F}_i$		$\tilde{T}_i$	
Models	**Without Latitude and Longitude**	**With Latitude and Longitude**	**Without Latitude and Longitude**	**With Latitude and Longitude**
Objective Soil Variables				
N_i^S	-0.004*** (0.001)	-0.004*** (0.001)	-	-
O_i^S	0.02** (0.01)	0.04*** (0.01)	-	-
C_i^S	-	-	-0.007*** (0.001)	-0.007*** (0.001)
Farmer & Household Characteristics				
own_i	0.20* (0.11)	0.18 (0.12)	0.42*** (0.1)	0.41*** (0.09)
inc_i	-0.02*** (0.001)	-0.02*** (0.001)	-0.006*** (0.001)	-0.007*** (0.001)
irr_i	-0.75*** (0.17)	-0.77*** (0.17)	-0.35** (0.16)	-0.34** (0.16)
$caste_i$	0.26** (0.11)	0.29** (0.11)	0.002 (0.11)	0.02 (0.11)
edu_i	-0.02** (0.007)	-0.02** (0.007)	-0.004 (0.006)	-0.003 (0.006)
$size_i$	0.27 (0.21)	0.36* (0.21)	0.64*** (0.16)	0.53*** (0.17)
Climate Variables				
$temp_i$	0.14** (0.06)	0.07 (0.08)	-0.11* (0.06)	-0.15** (0.07)
σ_i^{temp}	1.6*** (0.34)	0.75* (0.43)	-0.97** (0.38)	-1.25** (0.49)
Geographic Controls				
Latitude	-	0.05*** (0.02)	-	-0.01 (0.01)
Longitude	-	-0.04** (0.02)	-	-0.02 (0.02)
N | Log-Likelihood	8281 | -3523.08	8281 | -3511.75	8281 | -5276.4	8281 | -5270.87
Pseudo R^2	0.038	0.04	0.024	0.03

Notes: Standard errors (in parentheses) are clustered at village level. | *** p<0.01, ** p<0.05, * p<0.10.

Panel A: Soil Fertility Perceptions versus Soil Nitrogen				
Data-driven Soil Nitrogen Categories	**Soil Fertility Perceptions ($\tilde{F}_i$)**			**Total**
	Low	**Medium**	**High**	
Low ($N_i^S \leq 120.5$)	**82 (3%)**	2,513 (91%)	154 (6%)	2,749
Medium ($N_i^S \in (120.5, 142.8]$)	112 (4%)	**2,455 (89%)**	198 (7%)	2,765
High ($N_i^S > 142.8$)	139 (5%)	2,322 (84%)	**306 (11%)**	2,767
Total	**333**	**7,290**	**658**	**8,281**
Panel B: Soil Fertility Perceptions versus Soil Organic Carbon Stock				
Data-driven Soil Organic Carbon Categories	**Soil Fertility Perceptions ($\tilde{F}_i$)**			**Total**
	Low	**Medium**	**High**	
Low ($O_i^S \leq 38.3$)	**74 (3%)**	2,510 (92%)	155 (5%)	2,739
Medium ($O_i^S \in (38.3, 42.6]$)	117 (4%)	**2,426 (87%)**	235 (9%)	2,778
High ($O_i^S > 42.6$)	142 (5%)	2,354 (85%)	**268 (10%)**	2,764
Total	**333**	**7,290**	**658**	**8,281**

Notes: Rows represent objective soil quality categories (based on tertile cutoffs). Columns represent farmers' soil fertility perceptions. Green shading indicates cells where perceptions are consistent with objective data (diagonal). Red shading indicates directional misperceptions. Figures in parentheses represent row percentages.

Table 5.3: Cross-Tabulation between $\tilde{F}_i$ and $F_i \in \{N_i^S, O_i^S\}$.

Fertilizer Application Decisions	Measured Soil Fertility is Low		
	Consistent Perceptions $B_i^L \in \{L \rightarrow L\}$	Biased Perceptions $B_i^L \in \{L \rightarrow H\}$	Row Totals
Nitrogen Application Decision			
$\tilde{N}_i = 1$ (High)	188	147	335
$\tilde{N}_i = 0$ (Low)	266	74	340
Column Totals	454	221	675
Chi-Squared Statistic	37.47^{***}		
Phosphorus Application Decision			
$\tilde{P}_i = 1$ (High)	205	132	337
$\tilde{P}_i = 0$ (Low)	249	89	338
Column Totals	454	221	675
Chi-Squared Statistic	12.63^{***}		
Potassium Application Decision			
$\tilde{K}_i = 1$ (High)	209	125	334
$\tilde{K}_i = 0$ (Low)	245	96	341
Column Totals	454	221	675
Chi-Squared Statistic	6.59^{***}		
Fertilizer Application Decisions	**Measured Soil Fertility is High**		
	Consistent Perceptions $B_i^H \in \{H \rightarrow H\}$	**Biased Perceptions** $B_i^H \in \{H \rightarrow L\}$	**Row Totals**
Nitrogen Application Decision			
$\tilde{N}_i = 1$ (High)	272	63	335
$\tilde{N}_i = 0$ (Low)	220	128	348
Column Totals	492	191	683
Chi-Squared Statistic	27.38^{***}		
Phosphorus Application Decision			
$\tilde{P}_i = 1$ (High)	269	72	341
$\tilde{P}_i = 0$ (Low)	223	119	342
Column Totals	492	191	683
Chi-Squared Statistic	15.87^{***}		
Potassium Application Decision			
$\tilde{K}_i = 1$ (High)	261	78	339
$\tilde{K}_i = 0$ (Low)	231	113	344
Column Totals	492	191	683
Chi-Squared Statistic	8.21^{***}		

Table 5.4: Contingency Tables between Fertilizer Application Decisions and Farmers' Soil Fertility Misperceptions

Table 5.5: First-Stage binomial logit regression for farmers' soil fertility misperceptions (Over-Perceptions: $B_i^L \in \{L \to H, L \to L\}$, and Under-Perceptions: $B_i^H \in \{H \to L, H \to H\}$)

Explanatory Variables	Soil Fertility Overperceived $B_i^L \in \{L \to H, L \to L\}$	Soil Fertility Underperceived $B_i^H \in \{H \to L, H \to H\}$
Instruments		
σ_i^{temp}	0.43*** (0.04)	-0.37*** (0.09)
insect_i	0.12** (0.05)	-0.15** (0.04)
Controls	$own_i, caste_i, edu_i, ag_i, irr_i, size_i, price_{i,t-5}, temp_i$	

Notes: *** p<0.01 ** p<0.05 * p<0.10 \| Standard errors are robust to heteroskedasticity.

Table 5.6: Parent linear regression exploring influence of farmers' Over-Perceptions on fertilizer application.

Explanatory Variables	Nutrient Application					
	$nitro_i$		$phos_i$		$potash_i$	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
B_i^L	56.960*** (16.2)		103.89*** (31.87)		128.39*** (28.97)	
$\hat{B}_i^L$		-179.9*** (50.1)		-321.224*** (91.44)		-66.43 (74.55)
$price_{i,t-5}$	0.05** (0.02)	0.102*** (0.03)	0.005 (0.05)	0.09 (0.06)	-0.10** (0.05)	-0.06 (0.06)
own_i	-34.5* (20.3)	-9.9 (24.13)	-2.7 (34.5)	43 (41.36)	37.99 (28.99)	59.79* (32.29)
inc_i	0.71*** (0.21)	1.29*** (0.32)	1.88*** (0.43)	2.9*** (0.6)	1.13*** (0.38)	1.6*** (0.48)
irr_i	32.2 (20.07)	44.79* (23.92)	40.25 (40.63)	61.73 (46.51)	18.9 (38.16)	29.57 (38.79)
$caste_i$	31.06* (18.79)	62.28*** (22.97)	100.42*** (38.34)	157.01*** (46.44)	92.96** (37.68)	117.88*** (40.2)
edu_i	-3.69*** (1.35)	-4.98*** (1.57)	-9.71*** (2.6)	-11.85*** (2.9)	-6.77*** (2.18)	-7.92*** (2.3)
$size_i$	20.45 (32.36)	14.71 (38.99)	62.96 (63.22)	55.2 (74.94)	95.94* (57.03)	93.59 (59.94)
N	653	653	653	653	653	653
R^2	0.090	.	0.094	.	0.099	0.020
Instrument Relevance (F-Stat)	-	47.67***	-	47.72***	-	44.37***

Notes: *** p<0.01 ** p<0.05 * p<0.10 \| Standard errors are robust to heteroskedasticity.

Table 5.7: Parent linear regression exploring influence of farmers' Under-Perceptions on fertilizer application.

Explanatory Variables	Dependent Variables					
	$nitro_i$		$phos_i$		$potash_i$	
	OLS	IV	OLS	IV	OLS	IV
B_i^H	-29.437* (15.309)		-69.119** (28.456)		-86.644*** (23.795)	
$\hat{B}_i^H$		-129.783* (73.708)		-378.751** (154.835)		-637.270*** (165.671)
$price_{i,t-5}$	-0.001 (0.020)	-0.024 (0.029)	-0.047 (0.047)	-0.116* (0.063)	-0.144*** (0.042)	-0.270*** (0.068)
own_i	-35.142* (18.412)	-40.687** (18.583)	-11.753 (31.912)	-30.172 (34.309)	12.860 (25.664)	-21.520 (34.912)
inc_i	0.711*** (0.201)	0.588** (0.236)	2.052*** (0.415)	1.644*** (0.492)	1.473*** (0.330)	0.746 (0.498)
irr_i	44.849*** (13.492)	19.044 (23.775)	16.173 (33.865)	-66.933 (61.902)	-12.554 (28.729)	-156.458** (62.724)
$caste_i$	21.882 (16.280)	24.655 (16.505)	48.182 (31.310)	55.446* (33.190)	34.371 (27.793)	47.217 (35.261)
edu_i	-3.677*** (1.234)	-4.213*** (1.308)	-10.775*** (2.510)	-12.442*** (2.807)	-6.762*** (1.914)	-9.705*** (2.842)
$size_i$	-13.813 (27.969)	-9.122 (29.186)	-59.865 (54.491)	-49.791 (57.791)	13.851 (47.135)	30.944 (59.758)
N	659	659	659	659	659	659
R^2	0.063	-	0.072	-	0.069	-
Instrument Relevance (F-Stat)	-	13.35***	-	13.48***	-	13.82***

Notes: *** p<0.01 ** p<0.05 * p<0.10 | Standard errors are robust to heteroskedasticity.

Table 5.8: First-Stage regression modelling farmers' nitrogen, phosphorus and potassium application as a function of $price_{i,t-5}$ and other controls.

Explanatory Variable(s)	Dependent Variables		
	$nitro_i$	$phos_i$	$potash_i$
Over Perceiving Sub-Sample			
Instrument(s)			
$price_{i,t-5}$	-0.03** (0.01)	-0.09*** (0.03)	-0.11*** (0.03)
Controls	$own_i, caste_i, edu_i, ag_i, irr_i, size_i$		
Under Perceiving Sub-Sample			
Instrument(s)			
$price_{i,t-5}$	-0.01 (0.01)	-0.015 (0.03)	-0.009 (0.02)
Controls	$own_i, caste_i, edu_i, ag_i, irr_i, size_i$		

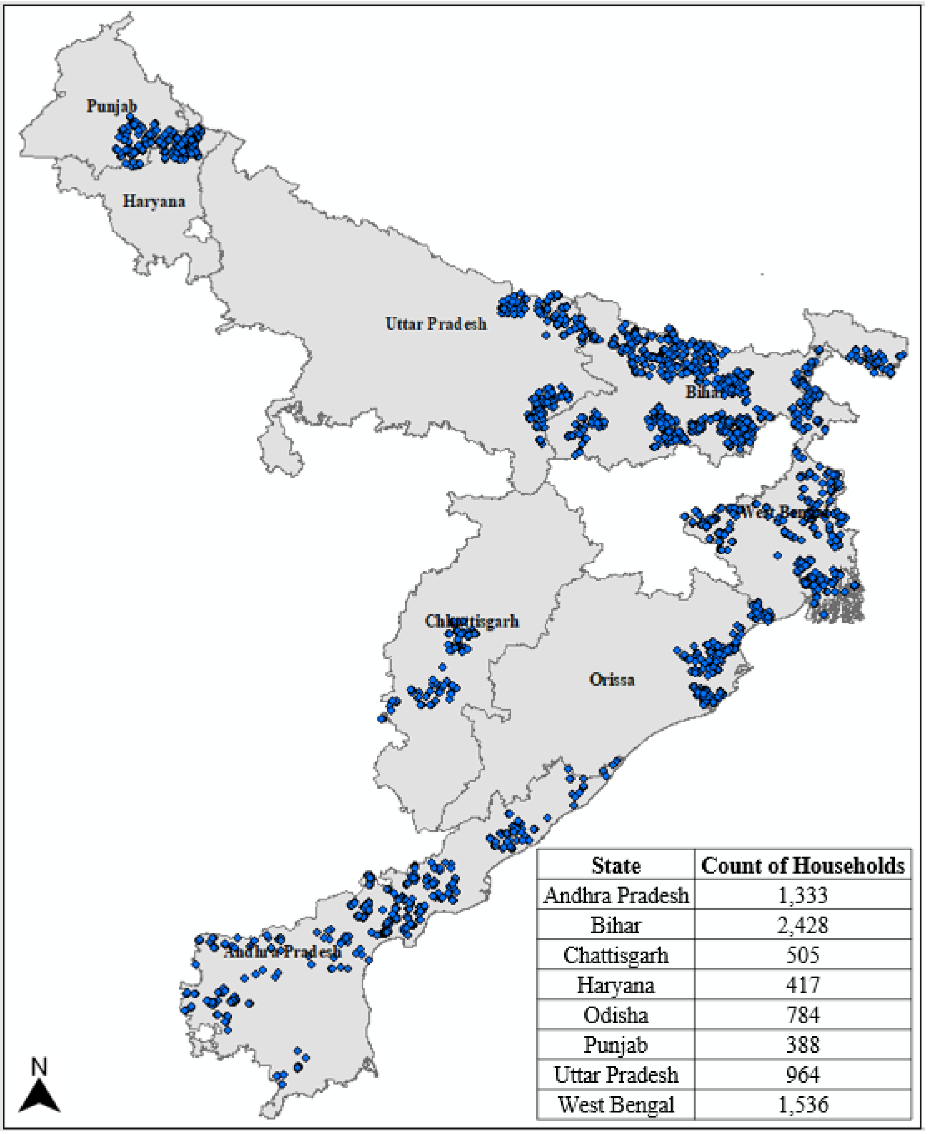
Notes: *** p<0.01 ** p<0.05 * p<0.10 | Standard errors are robust to heteroskedasticity.

Table 5.9: Influence of fertilizer application on propensity of Over-Perceptions, Regression Results (Probit and IV-Probit).

Explanatory Variables	**Dependent Variable: Over-Perceptions B_i^L**					
Model Description	$nitro_i$		$phos_i$		$potash_i$	
	Probit	**IV**	**Probit**	**IV**	**Probit**	**IV**
Fertilizer Application (kg/acre)	0.0014*** (0.0003)		0.0006*** (0.0001)		0.0007*** (0.0001)	
Predicted Fertilizer Application (kg/acre)		-0.004*** (0.001)		-0.002*** (0.001)		-0.001* (0.001)
own_i	0.25* (0.15)	0.034 (0.148)	0.21 (0.15)	0.197 (0.127)	0.21 (0.15)	0.276** (0.129)
ag_i	0.009*** (0.001)	0.009*** (0.002)	0.008*** (0.001)	0.010*** (0.002)	0.01*** (0.001)	0.009*** (0.002)
irr_i	0.19 (0.19)	0.337** (0.133)	0.21 (0.19)	0.275* (0.146)	0.25 (0.19)	0.250 (0.157)
$caste_i$	0.12 (0.14)	0.222** (0.111)	0.09 (0.13)	0.322** (0.134)	0.05 (0.14)	0.284* (0.154)
edu_i	-0.01 (0.01)	-0.024*** (0.008)	-0.01 (0.01)	-0.027*** (0.010)	-0.01 (0.01)	-0.023** (0.010)
$size_i$	0.23 (0.24)	0.138 (0.222)	0.25 (0.25)	0.222 (0.231)	0.24 (0.25)	0.326 (0.231)
$tempcv_i$	1.13*** (0.16)	0.111 (0.352)	1.06*** (0.16)	0.401 (0.318)	0.92*** (0.16)	0.695*** (0.206)
insect$_i$	0.39** (0.16)	0.327** (0.136)	0.38** (0.16)	0.383*** (0.141)	0.34** (0.16)	0.444*** (0.141)
N	651	651	651	651	651	651
R^2	0.12	-	0.11	-	0.11	-
Instrument Relevance (F-Stat)	-	5.42**	-	11.33***	-	18.63***

Notes: *** p<0.01 ** p<0.05 * p<0.10 | Standard errors are robust to heteroskedasticity.

5.7 Figures



State	Count of Households
Andhra Pradesh	1,333
Bihar	2,428
Chattisgarh	505
Haryana	417
Odisha	784
Punjab	388
Uttar Pradesh	964
West Bengal	1,536

Figure 5.1: Map representing the spatial distribution of 8355 farm households in our sample across eight Indian states.

Chapter 6: Conclusion

The essays in this thesis collectively advance our understanding of how farm-level constraints, input management, and behavioural perceptions shape yield outcomes and production risk management in semi-arid agriculture. Across these analyses, a consistent theme emerges: agricultural productivity and risk management in smallholder systems cannot be fully understood through average effects alone. Instead, distributional approaches that capture heterogeneity across yield outcomes, structural constraints, and subjective decision-making processes are essential for designing effective decision support systems and agricultural policy frameworks (Bedoya et al. 2018). The first paper operationalizes this argument by developing an ex-ante metric of credit rationing incidence that informs on the distribution of credit constraints across the sample population providing a sharper lens for rural credit targeting. Combining quantile regression with propensity score matching, the analysis reveals that while access to credit generally enhances yields, the benefits likely accrue disproportionately. The odds of rationing are characterised using a state-contingent technology framework that reveals the impact of rationing induced credit constraints on farm-level adaptive capacity under production uncertainty.

Building on this, the second paper quantifies how input choices shape yield distributions illuminating complementarities between input application, land use decisions, and institutional risk management instruments such as crop insurance - offering insights into persistent puzzles like low insurance uptake in developing regions. Employing Beta regression models with non-linear instrumental variables, the analysis indicates that inputs such as irrigation can simultaneously raise average yields and reduce downside yield risk, while other inputs like phosphorus exhibit risk-enhancing characteristics depending on the complementarities. The case study on cotton–pigeonpea mixed cropping further reveals that diversified inter-species cropping can reduce yield variability and enhance welfare through improved risk-return trade-offs. Collectively, these findings underscore that the design of tools for production risk management must account for farmers' on-farm production risk management strategies.

The final paper highlights divergence between farmers' subjective soil perceptions that are critical objects of farm-level decision-making, and data-based measures of soil quality (typically utilized in policy making) suggesting that effective land management policies must integrate behavioural insights regarding how farmers interpret their production constraints.

Taken together, these essays point to the need for a multidimensional understanding of production risk. Structural constraints (such as credit), technological choices (in inputs and land use), and subjective perceptions of the production environment jointly shape the yield distributions that determine agricultural welfare (Hennessey 2009; Waldman et al. 2025). Policy initiatives aimed at strengthening agricultural resilience in semi-arid regions must therefore operate on several fronts: improving credit targeting, promoting input use in a yield-stabilizing manner and aligning information systems with farmers' perceptions of their production environment (Braverman and Guasch 1986; Beaman et al. 2023; Bino et al. 2023). Beyond specific empirical findings, this thesis demonstrates the value of high-resolution microdata for agricultural policy design (Blandford, 2007; Mullen 2016; Carletto 2021). These methodological contributions (including the data paper) provide standardized frameworks that can be applied to investigate broader economic issues in the wider developing-country contexts that face rising production uncertainty.

6.1 References

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2023. “Selection Into Credit Markets: Evidence From Agriculture in Mali.” *Econometrica* 91 (5): 1595–1627. https://doi.org/10.3982/ecta18916.

Bedoya, Guadalupe, Arguelles, Luca Bittarello, Jon Davis, Nikolas Mittag, Jonathan Martin Villars Davis, and Nikolas Karl Mittag. 2018. “Distributional Impact Analysis: Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3006213.

Paul, Bino, Kamal K. Murari, Unmesh Patnaik, Chandra S. Bahinipati, and Subash Sasidharan. 2023. "Sustainability Transition for Indian Agriculture." *Scientific Reports 13, no. 1*: 7290. Accessed December 22, 2025. https://doi.org/10.1038/s41598-023-34092-0.

Blandford, D. 2007. “Information Deficiencies in Agricultural Policy Design, Implementation and Monitoring”, *OECD Food, Agriculture and Fisheries Papers, No. 6,* OECD Publishing, Paris. http://dx.doi.org/10.1787/067228574571

Braverman, Avishay, and J.Luis Guasch. 1986. “Rural Credit Markets and Institutions in Developing Countries: Lessons for Policy Analysis from Practice and Modern Theory.” *World Development* 14 (10–11): 1253–67. https://doi.org/10.1016/0305-750x(86)90104-x.

Carletto, Calogero. 2021. "Better Data, Higher Impact: Improving Agricultural Data Systems for Societal Change." *European Review of Agricultural Economics 48*, no. 4: 719-740. https://doi.org/10.1093/erae/jbab030.

Hennessy, David A. 2009. “Crop Yield Skewness Under Law of the Minimum Technology.” *American Journal of Agricultural Economics* 91 (1): 197–208. https://doi.org/10.1111/j.1467-8276.2008.01181.x.

Mullen, J. D. 2016. Impact assessment of ICRISAT Village Level Studies: 1975 to 2013. Tech. Rep., An independent external review commissioned by ICRISAT. Accessed here.

Waldman, Kurt, Jordan Blekking, Meha Jain, Chewe Nkonde, and Noemi Vergopolan. 2025. “What Discrepancies Between Farmer Perceptions and Observational Data Teach Us About Smallholder Decision Making.” *Environmental Research Letters* 20 (8): 081002. https://doi.org/10.1088/1748-9326/adeff3.

Sumedha Shukla CV

SUMEDHA SHUKLA

sumedhashukla.5@gmail.com +91-7982553558 New Delhi, India

RESEARCH INTERESTS

Agricultural Economics, Natural Resource Economics, Applied Econometrics & Spatial Econometrics

EDUCATION

PhD in Economics, IIIT-Delhi — July 2026
Advisor: Gaurav Arora
MSc. Economics (Specialization in Resource Economics), TERI University — 2016-18
Advisor: Soumendu Sarkar
Bachelors with Honours in Economics, Delhi University — 2013-16

WORK EXPERIENCE

Senior Associate Researcher: Tata-Cornell Institute for Agriculture and Nutrition — June 2026 - Present
Project: *Zero-Hunger Zero-Carbon Project*
Role:*Design & conduct large-scale primary survey in rural Bihar; enumerator training; data handling; & scientific reporting.*

Research Associate: Econometrics Lab, IIIT-Delhi — November 2018 - June 2021
Project: *Analyzing competing agricultural risk management policies in India*
Role:*Handle observational microdata & geo-spatial climate data; conduct literature reviews; write for popular press*

Research Assistant: Ashoka University — April 2020 - March 2022
Project: *Role of historic institutions & long-term effects on women's outcomes*
Role: *Combine historical administrative data with geo-spatial information; handle microdata from large-scale national sample surveys*

Intern: Centre for Policy Research, IISc., Bangalore — May 2017 - July 2017
Project: *Role of Science, Technology & Innovation in R&D Policy*
Role: *Prepare literature reviews*

PEER-REVIEWED PUBLICATIONS

- Maity B., Shukla S., 2025, **"Political Regime and Women's Marriage Outcomes"**, *Economica*, Volume 93, Issue 369, pp. 130-159, linked here.
- Sarkar, S., Shukla, S. 2020, **"Predatory Pricing and the Notion of Multi-Market Dominance: A Case of the Indian Telecom Industry"**. In: Datta M., Mahajan R., Bose M. (eds.) Business, Economics and Sustainable Development: The Emerging Issues, New Delhi: Bloomsbury, pp. 237-249.

POPULAR PRESS PUBLICATIONS

- Shukla S., Plavsik M., 2025, **"Selling Bihar's Livestock Farmers on Climate-Smart Technologies"**, Tata-Cornell Institute (TCI) Blogs.

- Shukla S., Plavsik M., 2025, **"Assessing Livestock Farmers' Willingness to Pay for Sex-Sorted Semen"**, **Tata-Cornell Institute (TCI) Blogs**.
- Shukla S., Plavsik M., 2025, **"Reducing Methane Emissions from Livestock in Bihar"**, **Tata-Cornell Institute (TCI) Policy Briefs**.
- Shukla S., Arora G., 2020, **"No Hand to Lend"**, **Down to Earth Magazine**.
- Shukla S., Arora G., 2020, **"The Credit Burden"**, **Indian Express (Opinion)**.
- Shukla S., Arora G., 2020 & 2021, "छोटे और भूमिहीन किसानों को कर्ज़ मिलना हुआ मुश्किल", **Down to Earth Magazine(Hindi) (2020), Tathyabharti (Hindi) (2021)**
- Shukla S., Maity B., 2021, **"Environment & Agriculture Related Conflicts in India: Insights from the Armed Conflict Location And Event Data (ACLED)"**, **Invited Article (ISPP Policy Review)**
- Kumar, K., Shukla S., 2018, **"SAARC Countries Cooperating on Energy could Yield Economic Benefits for All Nations"**, **Financial Express(Opinion)**
- Kumar, K., Shukla S., 2018, **"Rashtriya Swasthya Bima Yojana: RSBY Lessons for Ayushman Bharat"**, **Financial Express (Opinion)**
- Shukla S., Kumar K., 2018, **"Ayushman Bharat: A Critical Perspective"**, **LiveMint (Opinion)**
- Kumar, K., Shukla S., 2018, **"National Merit Order List for Electricity Generation to Help in Reducing Power Cost"**,**Financial Express (Opinion)**

WORKING PAPERS & CONFERENCE PRESENTATIONS

- Shukla, S., Arora, G., **"Heterogeneous crop yield impacts of agricultural credit: A quantile regression approach embedded in the state-contingent framework"**.
  Working Paper presented (accepted#) at:
  - **11th Research Scholars' Workshop** organized by **Calcutta University**,February 11, 2026.
  - **Young Scholars' Seminar**# organized by **Delhi School of Economics**,February 15-17, 2026.
  - **Annual Economics Conference** organized by **Ahmedabad University**,January 8-9, 2026.
  - **Advanced Graduate Workshop on Poverty, Globalization, and Development** organized by **Azim-Premji University** and **Institute of New Economic Thinking (INET)**, July 8-20, 2024.
  - **Research Symposium on Finance & Economics** organized by **Krea University**, 14-16 June 2023.
  - **IIT-Tirupati & YSI-INET** Workshop on Credit Markets in South Asia, 7-10 December 2022.
  - **Asian Economic Development Conference**, 14-15 July 2022.
  - **97th Western Economics Association** International Conference, Portland, June 29-July 3, 2022.
  - Annual Conference on Economic Growth and Development organized by **Indian Statistical Institute**, Delhi, New Delhi, December 20-22, 2021.
  - **DSE Winter School**, New Delhi, December 15-18, 2021.

- Invited Talk at the Policy Praxis Lab, **Indian School of Public Policy.**
- **AAEA & WAEA** Joint Annual Meeting, Austin, TX, August 1-3, 2021.
- **8th Triennial International Conference on Agricultural Statistics**, New Delhi, India, November 18-21, 2019.

- Shukla, S., Arora, G., Agarwal S.K., **"Estimation of crop yield density conditional on input choices for smallholder farms using a *BetaIV* framework."**
  Working Paper presented at:
  - **PSU Altoona International Conference on Empirical Economics**, August 3–4, 2024.
  - **Asian Meeting of the Econometrics Society**, IIT-Bombay, 10-12 January 2023.
  - **DSE Winter School**, New Delhi, December 14-17, 2022.
  - **AAEA Annual Meeting**, Anaheim, CA, July 31-August 2, 2022.

- Shukla, S., Arora, G., Agarwal S.K., **"Role of mixed-cropping in production risk management on small farms (An application of the *BetaIV* framework for yield density estimation)"**
  Working Paper presented (accepted#) at:
  - **The World Congress of Environmental and Resource Economists** organized by **World Council of Environmental and Resource Economists Associations (WCEREA)**, June 29 - July 3, 2026.
  - **AAEA & WAEA Joint Annual Meeting**, Denver,Colorado, 27-29 July 2025.
  - Conference on **Climate Change Adaptation and Resilience in Agriculture: Strategies for Sustainable Development in South Asia#** organized by **YSI-INET & Krea University**, March 17–19, 2025.

- Shukla, S., Arora, G. **"Assessing systematic biases in farmers' soil quality perceptions and their linkage with farming decisions: Evidence from rice-growing farmers in India"**
  Working Paper presented at:
  - Invited Talk at the **Uma Lele Workshop** organized by the **University of Oxford** in conjunction with **Tufts University**, 24-25 August 2026.
  - **PSU Altoona International Conference on Empirical Economics**, August 2–3, 2025.
  - **International Conference of Agricultural Economists**,New Delhi, August 2-7 2024.
  - **International Conference on Recent Developments in Economics Research: Theory and Evidence** organized by **Centre for International Trade and Development (CITD), JNU**, March 7-8, 2024.
  - **AAEA Annual Meeting**, Washigton DC, 21-23 July, 2023.

- Shukla, S., Arora, G. **"Plot-level land quality data pertaining to semi-arid tropics in India: A construction using the Village Dynamics Studies in South Asia Dataset"**
  Working Paper presented at:
  - **7th International Conference on Applied Econometrics (ICAE VII)** at **ICFAI**, Hyderabad, July 11-12 2024.

- Shukla, S., Arora, G. **"A rank similarity test for quantile treatment effects in conjunction with propensity score matching: An application to heterogeneous crop yield impacts of agricultural credit."**
  Working Paper presented at:
  - **Asian Meeting of the Econometrics Society**, IIT-Bombay, 10-12 January 2023
  - **AAEA Annual Meeting**, Anaheim, CA, July 31-August 2, 2022.
- Maity, B., Shukla, S. **"The Long-Term Effect of Exposure to the Taliban Regime on Women's Age at Marriage in Afghanistan"**
  Working Paper presented at:
  - Annual Conference on Economic Growth and Development organized by **Indian Statistical Institute**, Delhi, New Delhi, December 19-22, 2022.
  - **UNU-WIDER Development Conference**: The Puzzle of Peace – Towards Inclusive Development in Fragile Contexts

ACADEMIC AWARDS

2025	Completed summer training on **Remote Sensing & GIS** organized by **India Space Academy**
2025|2024	Award for **Outstanding Teaching Assistance (Dean's List)**, **IIIT-Delhi.**
2024	Conference participation grant by **International Association of Agricultural Economists**.
2023	Travel grant by **Applied Agricultural Economics Association (AAEA)**.
2019	Qualified for **Assistant Professor** under **NTA-NET exam**.
2018	**Gold Medalist**, MSc. Economics, TERI University.
2017	**Encouragement Award**, Annual Paper Presentation, ARSD College, Delhi University.
2013-16	**Annual Scholarship for Academic Excellence** at ARSD College, Delhi University.

TEACHING EXPERIENCE

Role: Preparing evaluations & grading; designing & teaching data analysis on selected software.

Graduate Teaching Assistant: IIIT-Delhi — August 2021 - Present
Courses: *Econometrics, Game Theory, Microeconomics, Econometrics with Machine Learning* — Class Size: 100-250
Software: R, Python | *Selected student projects: IIIT-D Econometrics Blog.*

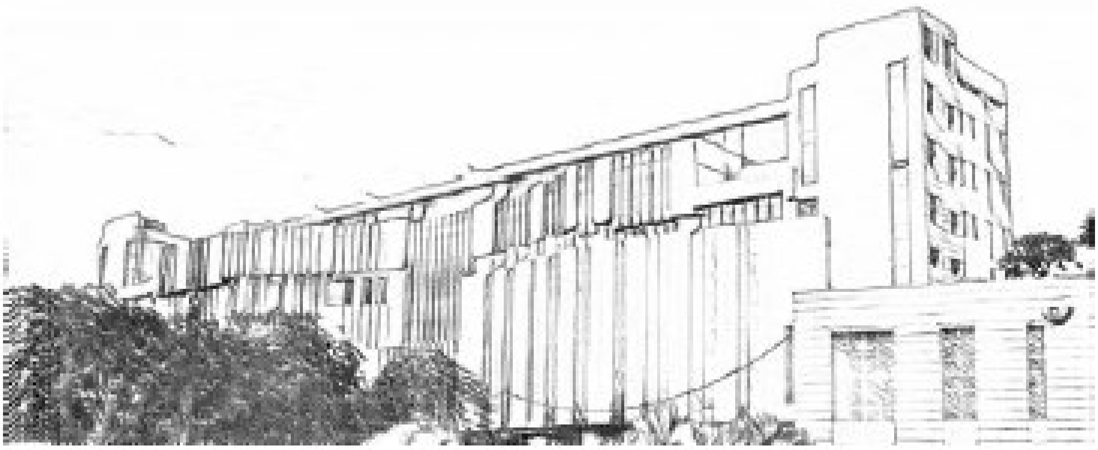
Guest Faculty: Indian School of Public Policy — March 2024 - June 2024
Course: *Introductory Econometrics* | **Software:** R — Class Size: 60

Teaching Assistant: Post-Graduate Diploma in Data Science in Health and Climate Change for Social Impact — October 2023-2024, Class Size: 85-100
Modules: *Climate Data Science, Spatial Data Analysis* | **Software:** R, QGIS

Learning Associate (Quantitative): Indian School of Public Policy — August 2020 - 2021
Courses: *Statistics, Game Theory, Law and Economics, Impact Evaluation* — Class Size: 50
Software: R, Excel

Part-Time Teaching Fellow: IIIT, Delhi — July 2019 - June 2020
Courses: *Econometrics, Game Theory* | **Software:** R, Mathematica — Class Size: 120-180

Instructor: Inkpot Online — May 2018 - May 2020
Courses: *Undergraduate Mathematical Economics & Econometrics* — Class Size: 25-400



INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY DELHI