

Do Index Insurance Programs Live Up to Their Promises?

Aggregating Evidence from Multiple Experiments

*Pauline Castaing
Jules Gazeaud*



WORLD BANK GROUP

Development Economics
Development Data Group
September 2022

Abstract

Despite limited uptake, index insurance is often seen as one of the most remarkable innovations of the past decades to help smallholder farmers manage risks. This paper uses a Bayesian hierarchical model to aggregate evidence from existing experiments and assess the external validity of their results. The findings show positive but highly heterogeneous responses to index insurance across experiments. Interventions expanding access to index insurance typically boost productive investments by 0.06–0.11 standard

deviation on average. However, treatment effects display substantial heterogeneity and there is no evidence that this heterogeneity can be meaningfully explained by basic household characteristics. The existing evidence base thus offers limited insights to predict the impact of index insurance in new settings. The paper concludes that governments and development agencies should remain cautious before investing in the widespread expansion of index insurance.

This paper is a product of the Development Data Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at pcastaing@worldbank.org and jgazeaud@povertyactionlab.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Do Index Insurance Programs Live Up to Their Promises? Aggregating Evidence from Multiple Experiments*

Pauline Castaing Jules Gazeaud

JEL Classification: C11, G22, G52 , 012, 013, Q12.

Keywords: Index Insurance, Agricultural Decisions, External Validity, Bayesian Hierarchical Modeling

*Pauline Castaing: The World Bank. E-mail: pcastaing@worldbank.org; Jules Gazeaud: J-PAL MENA, The American University in Cairo. E-mail: jgazeaud@povertyactionlab.org. We thank Frederick Amon-Armah, Catherine Araujo, Tanguy Bernard, Stefan Dercon, Esther Gehrke, Patrick Guillaumont, Alain de Janvry, Antoine Leblois, David McKenzie, Georgios Manalis, Rachael Meager, Muhammad Meki, Adam Osman, Elisabeth Sadoulet, Olivier Sterck, Quentin Stoeffler, Julie Subervie, Marteen Voors, Witold Wiecek, and conference, seminar, and workshop participants at CSAE, EUDN, GLaD, ICDE, J-PAL, and NOVAFRICA for helpful comments. We are grateful to all the authors who kindly agreed to share their data for this project: Erwin Bulte, Michael Carter, Francesco Cecchi, Shawn Cole, Ghada Elabed, Wouter Gelade, Xavier Giné, Catherine Guirkingier, Dean Karlan, Neha Kumar, Robert Lensink, Nicholas Magnan, Simrin Makhija, Ana Marr, Francesca de Nicola, Robert Osei, Isaac Osei-Akoto, David Spielman, Quentin Stoeffler, Christopher Udry, Marcel van Asseldonk, Ruth Vargas Hill, James Vickery and Patrick Ward. We are particularly indebted to Francesco Cecchi, Ghada Elabed, Sofia Gallo, Quentin Stoeffler, and Patrick Ward for helping us to understand the data, to Olivier Santoni for his help with the rainfall data, and to Rachael Meager and Witold Wiecek for developing the *baggr* R package and for answering our questions. Viewpoints and errors are our own. Part of this research is based on work conducted while the authors were affiliated with CERDI, CNRS, Université Clermont Auvergne (Castaing and Gazeaud), and NOVAFRICA, Nova SBE, Universidade NOVA de Lisboa (Gazeaud). Our pre-analysis plan is available at https://osf.io/zpesh/?view_only=97433abd5ea945f48c25d3963758eb6a.

1 Introduction

Index insurance has generated worldwide interest in recent decades because of its potential to help smallholder farmers manage risks. Unlike traditional indemnity-based insurance products, which compensate farmers based on the losses they experience, index insurance products delink payouts from the assessment of individual losses and compensate farmers on the basis of an exogenous, easily observable index that is highly correlated but not identical to individual losses.¹ The theoretical appeal of index insurance is that it reduces information asymmetry problems (adverse selection and moral hazard) as well as transaction and verification costs. Index insurance is particularly appealing in low- and middle-income countries because traditional forms of insurance are prohibitively expensive for smallholder farmers (Hazell, 1992; Cole and Xiong, 2017). According to Carter et al. (2017), it represents "one of the most important current opportunities to designing new institutions that can help developing countries achieve the goal of increased investment in agriculture, accelerated growth, and poverty reduction".

In this paper, we aggregate evidence from existing experiments on index insurance and explore the external validity of their results. While each individual study is commendable for the care with which it estimates impacts, critics have called into question the external validity of such estimates (Deaton and Cartwright, 2018; Peters et al., 2018; Bédécarrats et al., 2020; Ogden, 2020; Ravallion, 2020). Each study is rooted in a specific context (period, location, intervention, population) and it is not clear to which extent the estimates are informative about impacts in other contexts. We follow the seminal paper by Meager (2019) and use a Bayesian hierarchical model to aggregate data from six studies on investment response to index insurance (Elabed and Carter, 2014; Karlan et al., 2014; Cole et al., 2017; Bulte et al., 2019; Hill et al., 2019; Stoeffler et al., 2022).² This framework offers an appealing statistical tool, not only to study the external validity of the studies, but also to estimate the average impact across settings as well as the potential sources of heterogeneity. In contrast with traditional meta-analyses, which consider that individual studies estimate a common effect and that variations in observed treatment effects are solely driven by (random) sampling variation, the hierarchical model allows for the presence of heterogeneous effects across studies. More specifically, it assumes that variations in observed treatment effects reflect both sampling variation and (true) heterogeneity in treatment

¹Examples of indices include weather (e.g., rainfall, temperature), area production (e.g., average yield at the county or district level), and satellite images of vegetation.

²These studies cover a large set of regions and countries (Bangladesh, Burkina Faso, Ghana, Kenya, India and Mali), evaluate a wide range of insurance products, and thus offer an excellent opportunity to aggregate evidence and explore external validity. We explain in detail the inclusion criteria in Section 2.1.

effects. Most importantly, it provides a framework to disentangle these quantities and assess external validity.

We focus on effects on farmer investment decisions and examine five outcomes: cultivated area, fertilizers, pesticides, seeds, and a risk index.³ We construct these outcomes using original data from each study.⁴ We estimate intention-to-treat effects – that is the effects of being offered insurance regardless of the decision to subscribe or not – because it is arguably the most relevant parameter for a policymaker interested in the economy-wide effects of index insurance programs.⁵ Using the traditional pooling model, we find that farmers who are offered index insurance cultivate more land (+0.09 SD), and use more seeds (+0.10 SD), pesticides (+0.09 SD), and fertilizers (+0.07 SD). The effect on the risk index is smaller and non-significant. Using the Bayesian hierarchical model, we find similar point estimates (around 0.06–0.11 SD), however, effects are much more uncertain and can be close to zero or negative. The 95% posterior intervals are [−0.02 SD, +0.26 SD] for cultivated area, [−0.02 SD, +0.25 SD] for seeds, [−0.02 SD, +0.15 SD] for pesticides, [−0.05 SD, +0.21 SD] for fertilizers, and [−0.09 SD, +0.09 SD] for the risk index. These results tend to confirm the potential of index insurance to foster smallholder farmers’ productive investments, however, they also show that effects are more uncertain than suggested by the pooling model. In fact, the pooling model is misleading because it considers that impacts are homogeneous across studies (thus understating the uncertainty around treatment effects).

We then use estimates from the Bayesian hierarchical model to gauge the external validity of index insurance experiments. We report evidence on the conventional pooling factor, a metric which emphasizes the percentage of total variation in treatment effects stemming from sampling variation (the higher this metric, the more suggestive it is of external validity). Averaged across all outcomes and studies, we find moderate pooling of information: 52% of the observed variation in treatment effects stems from sampling variation. We obtain similar results using two complementary metrics: the brute force pooling metric and the generalized pooling factor. This suggests moderate external validity of index insurance experiments. We note however that with few experiments the pooling factor cannot be precisely estimated: the

³The risk index is defined as $R_{ik} = \sum r_{mk}d_{imk} / \sum d_{imk}$, where r_{mk} is the coefficient of variation of the yield of crop m in study k , and d_{imk} is the amount of land devoted by household i to the cultivation of crop m in study k . It captures the riskiness of crop portfolios.

⁴To derive the coefficient of variation r_{mk} of the risk index, we complement survey data from the original studies with time series on agricultural yields (see Section 3.1 below for more details).

⁵For reference, we compare the study-specific estimates of the ITT (intent-to-treat effect) and of the ATET (average treatment effect on the treated). If anything, we see larger heterogeneity in the ATET estimates (Figure A3).

95% posterior interval of the averaged pooling factor is [0.14, 0.94]. This highlights the need for more data and more accurate metrics.

To improve our understanding of the drivers of heterogeneity, we examine the effect of index insurance conditional on the following covariates: household wealth, household size, household head age, household head literacy, household predicted outcomes, and the price of the insurance product. We assess whether these characteristics are associated with larger and more heterogeneous effects (which might generate policy-relevant insights for the targeting of index insurance interventions). The results show limited heterogeneity along these covariates. We find some evidence that treatment effects are larger among households with low levels of predicted outcomes, among worse-off households, among households with a younger household head, and among households offered a lower price for the insurance product. However, in most cases, the posterior interval of the interaction includes zero comfortably. This suggests that most of the heterogeneity in treatment effects might stem from more complex household characteristics or from contextual and product characteristics.

We conclude our analysis by using estimates from the Bayesian hierarchical model to predict the effect of index insurance in a new setting. We find that the posterior intervals of the predicted effects are considerably wider than the confidence intervals from the pooling model (by a factor from 2 to 5), reflecting the substantial uncertainty stemming from treatment effect heterogeneity. For example, we estimate that in a new context the treatment effect on fertilizer has a 50% chance of being between -0.00 SD and $+0.17$ SD, a 25% chance of being negative, and a 25% chance of being larger than $+0.17$ SD. This suggests that the existing evidence base offers limited insights to predict the effects of index insurance offers in new settings, and that conducting new studies in such settings would be particularly worthwhile before widespread promotion.

This paper makes two main contributions. First, it adds to the debate as to whether and how to introduce index insurance in low- and middle-income countries. Index insurance triggered a considerable interest among academics and development practitioners. The pioneering paper by [Karlan et al. \(2014\)](#) in northern Ghana investigates the effect on investment decisions of offering capital grants and access to index insurance to smallholder farmers. The authors find strong responses of agricultural investment to the provision of index insurance, but relatively small effects of the capital grants. These results suggest that uninsured risk – and not a

lack of access to credit – is the binding constraint to investment in this context.⁶ Moreover, a handful of studies suggest that index insurance can help farmers recover in the wake of covariate shocks. Using a multi-year experiment in Mozambique and Tanzania, [Boucher et al. \(2021\)](#) find that farmers with access to index insurance and drought tolerant seeds are more resilient to large covariate shocks than farmers with only access to drought tolerant seeds.⁷ Despite these encouraging results, the delivery of index insurance to smallholder farmers has generally been disappointing, with many schemes ending up in failure or requiring heavy subsidies to induce substantial uptake ([Carter et al., 2017](#)).⁸ We contribute to the literature by aggregating evidence from existing experiments and by exploring the external validity of their results. Our results highlight the potential of index insurance to foster smallholder farmers' productive investments. However, they also reveal considerable heterogeneity in treatment effects across studies, with limited evidence that this heterogeneity can be meaningfully explained by basic household characteristics. The existing evidence base thus offers minimal insights to predict the impact of index insurance in new settings.

Second, it adds to the literature on external validity, open science, and research transparency. Randomized controlled trials are often criticized on the basis that their results are drawn from specific contexts and may therefore lack external validity.⁹ A rapidly growing literature relies on Bayesian hierarchical models to aggregate information from multiple settings and assess external validity ([Meager, 2019](#); [Romero et al., 2020](#); [Vivalt, 2020](#); [Bandiera et al., 2021](#); [Jackson and Mackevicius, 2021](#); [Kremer et al., 2022](#); [Meager, 2022](#)).¹⁰ We add to this body of literature by applying the method to a new question – the effect of index insurance on productive investments – and by showing the scope for preregistration when data are not publicly available. Study preregistration is one of the most popular tools to promote open science and

⁶Other studies on index insurance from Bangladesh ([Hill et al., 2019](#)), Burkina Faso ([Stoeffler et al., 2022](#)), China ([Cai, 2016](#)), India ([Mobarak and Rosenzweig, 2013](#); [Cole et al., 2017](#)), Kenya ([Jensen et al., 2017](#); [Bulte et al., 2019](#)) and Mali ([Elabed and Carter, 2014](#)) report qualitatively similar responses to index insurance provision. To the best of our knowledge, [Ahmed et al. \(2020\)](#) in Ethiopia is the only study reporting no investment response to index insurance. The authors find very low uptake of the insurance product at market prices (between 0.5% and 3%), and no meaningful increases in input investment when farmers were provided with small amounts of free insurance.

⁷See also: [Jensen et al. \(2017\)](#); [Bertram-Huemmer and Kraehnert \(2018\)](#); [Janzen and Carter \(2019\)](#); [del Valle et al. \(2020\)](#); [Stoeffler et al. \(2022\)](#).

⁸Several barriers have limited the uptake of index insurance. These include basis risk, liquidity constraints, lack of trust in the insurance company, lack of financial literacy of household decision makers, and limited salience of benefits ([Giné et al., 2008](#); [Giné and Yang, 2009](#); [Gaurav et al., 2011](#); [Cole et al., 2013](#); [Dercon et al., 2014](#); [Clarke, 2016](#); [Cai and Song, 2017](#); [Casaburi and Willis, 2018](#); [Stein, 2018](#); [Belissa et al., 2019](#); [Cai et al., 2020](#); [Cecchi et al., 2022](#); [Michler et al., 2022](#); [Stoeffler and Opuz, 2022](#)).

⁹The issue of external validity applies not only to experiments but also to observational studies.

¹⁰For other approaches to assess external validity, see for example: [Pritchett and Sandefur \(2014\)](#); [Gechter \(2015\)](#); [Bo and Galiani \(2021\)](#); [Dehejia et al. \(2021\)](#).

research transparency (Olken, 2015; Christensen et al., 2019). It allows researchers to bind their hands against data mining (Humphreys et al., 2013; Brodeur et al., 2016), and to mitigate publication bias arising from the under-report of null results (Casey et al., 2012). Yet, preregistration remains largely confined to experimental studies which in turn constitute only a small share of the overall economic research (Burlig, 2018). An important innovation of our work is that we preregistered our analysis prior to having the data at hand.¹¹ This paper thereby illustrates how preregistration can be credibly implemented by third-party researchers when data are already collected but not deposited in public repositories. We believe that preregistering our project was helpful to build trust with the authors of the original studies and secure access to the requested data.

The rest of the paper is organized as follows. Section 2 discusses study selection. Section 3 describes the data. Section 4 presents the methodology. Section 5 outlines the results. Section 6 concludes.

2 Study selection

2.1 Inclusion criteria

We focus on studies which (i) evaluate the effect of access to index insurance offers, (ii) measure impacts on the investment decisions of farmers, and (iii) are designed as randomized controlled trials. Focusing on randomized controlled trials has the advantage of making more credible the assumption of no systematic bias in the estimation of treatment effects (see Section 4). To maximize the number of eligible studies and limit the potential for publication bias, we do not restrict eligibility to articles published in peer-reviewed journals. Since index insurance is a relatively new product (the first products were developed in the early 2000s), we do not impose particular time constraints either. We focus on effects on farmer investment decisions for three reasons.¹² First, impacts are theoretically ambiguous: while access to index insurance can allow more risk-taking, the need to pay early premiums and the lack of financial resources

¹¹Meager (2019) also relies on a pre-analysis plan although in her setting the data were publicly available.

¹²Alternatively, we could focus on the impact of insurance on risk coping or on the determinants of insurance take-up. However, we see these aspects as less suited for the analysis we conduct in this paper. On one hand, the ex-post impacts of index insurance are not expected to be generalizable since they typically depend on idiosyncratic shocks (Rosenzweig and Udry, 2020). On the other hand, studies on the determinants of take-up look at very different dimensions such as the level of subsidy, or the presence of add-on interventions to increase trust, financial literacy, or risk-sharing. It is not clear on which dimension we should focus. For reviews on both aspects, see Binswanger-Mkhize (2012), Miranda and Farrin (2012), Marr et al. (2016), Carter et al. (2017), Cole and Xiong (2017), Jensen and Barrett (2017), Platteau et al. (2017), Clement et al. (2018), Ali et al. (2020), and Kramer et al. (2022).

of farmers could in fact crowd-out productive investments (especially in contexts with deficient credit markets). Second, the question is well-tailored to the current policy debate, as demand for index insurance is highly price sensitive and particularly low at market prices. The optimal level of subsidies remains unclear and depends on the size of the productivity gains at stake. Third, although the overall evidence suggests that access to index insurance can stimulate productive investments, some heterogeneity in treatment effects exists across studies and deserves to be investigated more systematically. For example, farmers with access to index insurance increased total cultivation expenditures by 23% in Bangladesh (Hill et al., 2019), by 13% in Ghana and Kenya (Karlan et al., 2014; Bulte et al., 2019), but in India researchers found small and non-significant effects (Cole et al., 2017).

2.2 Search process

We found ten studies satisfying our inclusion criteria. In order to identify eligible studies, we relied first on the studies listed in the review by J-PAL, CEGA, and ATAI (2016). We found an initial set of four eligible papers: Mobarak and Rosenzweig (2013); Elabed and Carter (2014); Karlan et al. (2014); Cole et al. (2017). We then relied on Google Scholar searches to screen all studies citing at least one of these four initial papers and we further assessed their eligibility by reviewing their title, abstract, and empirical analysis. This resulted in the inclusion of three additional papers: Berhane et al. (2015); Ahmed et al. (2020); Stoeffler et al. (2022). We completed this search process in April 2018. Three additional papers (Belissa and Marr, 2018; Bulte et al., 2019; Hill et al., 2019) reached our attention before registering our study, bringing the total number of papers to ten. We stopped screening potential papers to include in August 2020, following the registration of our pre-analysis plan.¹³ To confirm that this search process did not miss eligible studies, we conducted two independent literature searches on Google Scholar: (i) using the keyword “index insurance” and any of the terms “evidence”, “randomized”, or “experiment”; (ii) using the keyword “index insurance” and any of the terms “production decision”, “investment”, “risk”, or “ex-ante”. These searches yielded 116 and 327 results respectively. However, after reviewing these papers, we did not find any additional studies satisfying our inclusion criteria.¹⁴

¹³Our pre-analysis plan is available [here](#).

¹⁴These additional searches were conducted in February 2022, following the comments received by colleagues and conference participants. Naturally, we only considered a study as eligible if it was released before we registered our project.

2.3 Publication bias

As documented by [Ioannidis et al. \(2017\)](#) and [Andrews and Kasy \(2019\)](#), null effects tend to disappear from the literature because of the tendency of researchers and journal editors to prefer studies with statistically significant results. To limit such issues, as mentioned above, we did not restrict our literature search to published articles. However, a remaining concern is that researchers may not have written up papers for studies with null findings because they anticipate a harder publication process or because of their own preferences to not pursue the publication of null results. We take advantage of the AEA randomized controlled trial registry to check whether there are studies that meet our inclusion criteria but whose results have not been reported in the literature.¹⁵ In February 2022, we conducted a simple search on the registry using the keyword “index insurance”. After reviewing the description of each project, we found a total of 15 studies that met our eligibility criteria, seven of which had been identified through our search process. Among the remaining eight studies, three have results publicly released in academic papers but not available when we registered our study ([Boucher et al., 2021](#); [Mishra et al., 2021](#)),¹⁶ one analyzes an intervention that was completed only recently ([Cilliers et al., 2020](#)), and four explore interventions that are still ongoing ([Kazianga and Wahhaj, 2021](#); [Kramer et al., 2021](#); [Vieider et al., 2021a,b](#)). If anything, this provides some reassurance that concerns over publication bias are not too severe in this literature.

2.4 Data requests

Before the preparation and the registration of our pre-analysis plan, in which we specified the different variables we would use, we contacted the authors of each paper to present our project and to ask whether they would accept to share the underlying data.¹⁷ We received principle agreements for seven studies. These studies are: [Mobarak and Rosenzweig \(2013\)](#), [Elabed and Carter \(2014\)](#), [Karlan et al. \(2014\)](#), [Cole et al. \(2017\)](#), [Bulte et al. \(2019\)](#), [Hill et al. \(2019\)](#), and [Stoeffler et al. \(2022\)](#). For three studies ([Berhane et al., 2015](#); [Belissa and Marr, 2018](#); [Ahmed et al., 2020](#)), either data were not shareable or we received no answer. Since our methodology requires original data, we do not include these studies. Importantly, at this stage, although the authors already agreed to share their data for our study, we made clear that our first step

¹⁵The AEA RCT registry provides details on a study’s characteristics irrespective of whether or not it is ultimately published.

¹⁶[Boucher et al. \(2021\)](#) report the results from two registered projects. [Boucher et al. \(2021\)](#) and [Mishra et al. \(2021\)](#) do not report impacts on ex-ante production decisions.

¹⁷The email we sent is reproduced in [Appendix C](#).

would be to preregister our study and that we therefore only needed their survey instruments to start the project. After the registration of our pre-analysis plan, we followed up with the authors of each study and obtained the data for six of the seven studies whose authors gave us principle agreements. For the remaining study ([Mobarak and Rosenzweig, 2013](#)), the authors were understandably too busy with Covid-19 to prepare and send us the data to construct the requested variables.

2.5 Description of the studies

Together, the six studies cover a large set of regions and countries (Bangladesh, Burkina Faso, Ghana, Kenya, India and Mali). Five studies have been published in peer-reviewed journals. The remaining is available as an unpublished mimeo. Further characteristics of the studies are provided in Table 1. We report information on the start year, experimental design, eligibility criteria, sampling frame, sample size, and number of clusters. Start years range from 2009 in Ghana to 2016 in Kenya. Of the six studies, four are designed as price experiments. Price experiments randomly assign different levels of premium subsidies to treated clusters. Depending on the studies, clusters are villages, farmer groups or cooperatives. Other studies provide index insurance for free with randomization occurring at the household level. Two studies have additional treatments: [Karlan et al. \(2014\)](#) cross-randomize cash grants; [Cole et al. \(2017\)](#) cross-randomize coupons for discounts on locally appropriate inorganic fertilizer. Eligibility criteria vary largely across studies. [Cole et al. \(2017\)](#) target land owners, [Karlan et al. \(2014\)](#) target maize cultivators with less than 15 acres of land, and [Elabed and Carter \(2014\)](#), [Bulte et al. \(2019\)](#), [Hill et al. \(2019\)](#), and [Stoeffler et al. \(2022\)](#) target individuals belonging to specific groups (NGOs, farmer groups, cooperatives). Broadly speaking, however, these studies all cover smallholder farmers residing in low- and middle-income countries. All but one study have a baseline survey. Sample sizes vary from 780 in Kenya to 2,300 in Bangladesh. Interestingly, studies with the largest sample size are not those with the highest number of clusters. Because clusters and sample size are two of the main parameters that researchers can influence to achieve desired statistical power, this could reflect the fact that all studies maximize the same power calculation function under different budget and practical constraints.

2.6 Description of the index insurance products

Table 2 describes the products offered in each study. Three products are based on rainfall, two on area-yields, and one is hybrid (that is, based both on rainfall and area-yields). In four studies, the products are explicitly targeted towards particular crops: [Elabed and Carter \(2014\)](#) and [Stoeffler et al. \(2022\)](#) target cotton production; [Karlan et al. \(2014\)](#) target maize production; [Bulte et al. \(2019\)](#) target four crops (soya bean, sorghum, sunflower, and maize). In other studies ([Cole et al., 2017](#); [Hill et al., 2019](#)), the products are not crop-specific. Large differences exist in sale prices, both within and across studies. Within study differences are due to the price experiment design of four studies. Differences across studies reflect only to some extent differences in actuarially fair prices. Insurance purchase was generally an individual decision, except in [Elabed and Carter \(2014\)](#) and [Stoeffler et al. \(2022\)](#) where farmer groups collectively decided to purchase the insurance. In most studies, the amount of insurance that farmers could purchase depends on the amount of land they cultivate. A notable exception is the study by [Cole et al. \(2017\)](#) in India, which offered insurance products as stand-alone policies (treated households received 10 policies). Take-up of the different products vary from 29% in Mali to 100% for the products offered for free. Payoff triggers depend for example on the amount of dry/wet days, on the duration of dry spells, on cumulative rainfall, or on area-yields averages. While products based on rainfall typically use data from weather stations, area-yield products rely upon data from crop-cutting exercises or from purchasing companies.

3 Data

3.1 Primary outcomes

We focus on five outcomes: the amount of cultivated area, fertilizers, pesticides, seeds, and a risk index. This choice is guided by both theoretical and practical considerations. Conceptually, a well-established theory of change outlines that insurance provision could either increase or decrease farmer risk-taking depending on financial constraints and risk preferences.¹⁸ We therefore focus on production decisions involving a certain degree of risk-taking. Practically, because our empirical framework requires outcomes to be measured in similar ways across studies, the set of outcomes is constrained by data availability. It would be interesting to analyze outcomes such as loan subscription, irrigation, and labor expenditures, however, these

¹⁸See [Karlan et al. \(2014\)](#) for a theoretical model of investment decisions under financial and risk constraints.

outcomes are missing in many studies and measured in very inconsistent ways. We therefore limit our analysis to the amount of cultivated area, fertilizers, pesticides, seeds, and a risk index – whose components are available for all studies and measured in fairly consistent ways. We note that these outcomes are not identical from a theoretical point of view. Assuming that financial constraints are not too binding and that farmer preferences are characterized by decreasing absolute risk aversion, the provision of insurance should affect positively cultivated area, fertilizers, seeds, and crop riskiness, whereas it is not obvious that more insurance would foster the use of pesticides (more pesticides increase production costs but tend to decrease production risks).

Most of these outcomes have straightforward definitions. For the risk index, we draw on [Gehrke \(2019\)](#) and define it for household i in study k as $R_{ik} = \sum r_{mk} d_{imk} / \sum d_{imk}$, where r_{mk} is the coefficient of variation of the yield of crop m in study k , and d_{imk} is the amount of land devoted by household i to the cultivation of crop m in study k .¹⁹ We derive r_{mk} using time series of agricultural yields provided by the FAO at the country level.²⁰ For India, given the possibly large differences in crop riskiness across regions, we use sub-national data on crop yields from the Directorate of Economics and Statistics, Ministry of Agriculture.²¹ For studies in which the insurance products are based on rainfall exclusively (see [Table 2](#)), we enrich the index with the term f_{mk} to take into account the rainfall sensitivity of crop m in study k . More specifically, we use f_{mk} to multiply r_{mk} and deflate the contribution to the index of crops which are less sensitive to rainfall. The idea is to prevent the influence of other production risks, such as pests or diseases, which are not covered by rainfall insurance. We use monthly rainfall data drawn from the Climatic Research Unit, and simply define f_{km} as the correlation between total rainfall over the three wettest months and the yield of crop m .²² Overall, this index captures how insurance provision affects farmer crop portfolio choices in light of the traditional risk-productivity trade-off outlined in the literature.

[Table A1](#), taken from our pre-analysis plan, provides more details on how we derived each

¹⁹We view the allocation of land to each crop as perhaps more appropriate than the allocation of other inputs such as fertilizer because it is more ex-ante with respect to rainfall realization, and because different crops may have different fertilizer needs.

²⁰FAOSTAT provides annual statistics on harvested areas, production quantities and yields for 173 crops over the 1961-2017 period, covering the production of all primary crops for all countries – see <http://www.fao.org/faostat> for more details.

²¹See <https://aps.dac.gov.in/> for more details. To our knowledge, India is the only country in our sample providing yearly sub-national crop production statistics over a sufficiently large time frame.

²²As a robustness check, we also add a quadratic term to take into account the possibility of non-linear relationships and find similar results.

outcome, and which survey questions we used. While all outcomes are measured over the rainy season, Hill et al. (2019) also measured outcomes over the dry season. In this paper, we focus exclusively on outcomes over the rainy season because (i) data over the dry season are incomplete, (ii) insurance products only cover the rainy season, and (iii) payouts for the rainy season could influence production decisions over the dry season (in Bangladesh, rainy season payouts coincided with the planting of dry season crops). We standardize all outcomes to deal with slight differences in measurement across studies. For example, some studies collect fertilizer expenditures, while others collect only purchased quantities with no information about prices. We winsorize all outcomes at the 99th percentile to reduce the influence of outliers.

3.2 Covariates of interest

Household-level covariates may be particularly helpful to understand what drives the heterogeneity in treatment effects. In addition, they may provide useful evidence to guide the targeting of insurance products and identify subgroups for which production decisions are more responsive to the provision of index insurance.²³ We examine the following covariates: a wealth index, household size, household head age, household head literacy, and predicted outcomes.²⁴

For continuous covariates, we divide households in two groups with respect to their median.²⁵ We derive the wealth index using the methodology developed by the DHS. We predict each of our five outcomes using an OLS regression with the following set of predictors measured at baseline: chemicals usage, cultivated area, land ownership, household head gender, literacy, education, age (and age²), household size, wealth, livestock, bank account, distance to rainfall station, experience of past weather shocks, and risk preferences.²⁶ Table A2, taken from our pre-analysis plan, specifies the data we use to derive each covariate. Table A3 shows

²³Using data from seven studies on microcredit, Meager (2019) finds that microcredit only affects the profits of households with prior business experience, and that the effects for this subpopulation vary largely across contexts. This pattern suggests that business experience of microcredit beneficiaries is a necessary but not sufficient condition for positive effects.

²⁴We also specified in our pre-analysis plan that we would analyze heterogeneity with respect to household head's gender. However, very few households are in fact headed by females, as shown in Table A3.

²⁵We stratify the dummy variables by studies. However, for covariates using standardized scales across studies (household head age and household size), we show that estimates are similar with no stratification (Figure A5).

²⁶More specifically, we regress each outcome on its predictors using the full sample of control observations, and then use the coefficients from this regression to predict the outcomes of all sample units. In our pre-analysis plan, we specified that we would estimate treatment effects using a repeated split sample approach to limit the bias resulting from overfitting (Abadie et al., 2018), however this is not possible using the *baggr* R package. Our results on predicted outcomes should therefore be interpreted with caution. In particular, treatment effects for farmers with low predicted outcomes are likely to suffer from an upward bias, whereas treatment effects for farmers with high predicted outcomes are likely to suffer from a downward bias (Abadie et al., 2018).

basic descriptive statistics. We note that the list of variables available to construct the wealth index and the predicted outcomes vary somewhat between studies. This should not be an issue since we identify groups with high and low levels of wealth and predicted outcomes within each study. In a few cases, information is only available at endline. While we can legitimately rely on endline data for predetermined characteristics such as the age, gender or literacy of the household head, it may be misleading for variables potentially affected by the treatment. For predicted outcomes, we use only variables measured at baseline or predetermined. For the wealth index, when no baseline data are available, we focus on highly autocorrelated and relatively illiquid outcomes such as televisions, materials used for housing construction, and the types of water access and sanitation facilities.²⁷

Although it would be interesting to investigate heterogeneity in treatment effects using study-level covariates such as the characteristics of the insurance products, for two reasons we prefer to leave it to future research. First, with only six studies and no within-study variation, the sample size would be particularly limited for such analysis. Second, some of the covariates of interest are perfectly colinear and therefore impossible to disentangle. Insurance may be more attractive to farmers if products are provided at the group level (De Janvry et al., 2014) or if premiums are paid after harvest (Casaburi and Willis, 2018). One could therefore look at heterogeneity along the decision level for take-up (group vs. individual) or along the timing of payment (up-front vs. at-harvest). However, looking at Table 2, we see that only the insurance products in Burkina Faso and Mali have been offered at the group level, or using a pay-at-harvest contract, meaning that these two characteristics are not separable (the products in Burkina Faso and Mali are also the only two targeting cotton and based on a yield index). A promising avenue to investigate such patterns is to randomly vary product characteristics within a given study. For example, Casaburi and Willis (2018) varied the timing of payment of the premiums and find that the take-up rate is 72 percent for a pay-at-harvest product against 5 percent for a standard pay-up-front product. In this paper, we focus on the subset of studies designed as price experiments and explore heterogeneous effects with respect to assigned prices.

²⁷We do not include livestock in the wealth index because it is a highly liquid asset and in two studies (Cole et al., 2017; Bulte et al., 2019) livestock units are only available at endline such that it could be affected by the treatment.

4 Methodology

This section summarizes the Bayesian hierarchical model we use to aggregate evidence from the experiments on index insurance and explore the external validity of their results. Our approach closely follows [Gelman et al. \(2013\)](#), [Meager \(2019\)](#), and [Bandiera et al. \(2021\)](#) to which we refer readers for further technical details.

We are interested in the effect of randomly assigned offers to take up index insurance on the set of agricultural decisions described in Section 3.1. A natural starting point is the following simple descriptive model:

$$y_{ik} = \mu_k + \tau_k T_{ik} + \epsilon_{ik} \quad (1)$$

where y_{ik} is the standardized outcome variable for household i in study k ;²⁸ T_{ik} is a dummy equal to one if household i in study k is offered index insurance; τ_k is the treatment effect and the parameter of interest of each study k ; μ_k is the mean outcome in the control group in study k ; and ϵ_{ik} is the error term (clustered at the level of treatment, that is the household or the village depending on the study). Note that τ_k corresponds to the intention-to-treat effect of index insurance, that is the effect of being offered index insurance regardless of take-up decisions. This parameter is of particular interest for policymakers who need insights on the potential economy-wide effects of index insurance programs.

We define the *true* heterogeneity in treatment effects as $\sigma_\tau^2 = \text{var}(\tau_k)$. This quantity provides a natural measure to gauge the level of external validity of available evidence. The main challenge, however, is that parameters $\{\tau_1, \tau_2, \dots, \tau_K\}$ are unknown, and one can only observe estimates $\{\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_K\}$. Importantly, $\text{var}(\hat{\tau}_k)$ not only reflects the true heterogeneity in treatment effects across studies, σ_τ^2 , but also sampling variation (sometimes also called idiosyncratic or statistical variation, that is the difference between $\hat{\tau}_k$ and τ_k resulting from the use of sampling techniques).²⁹ In other words, because of sampling variation, the observed heterogeneity $\text{var}(\hat{\tau}_k)$ overestimates the true heterogeneity σ_τ^2 .

²⁸As mentioned in Section 3.1, we standardize all outcome variables because they are typically measured using different scales or units across studies. The standardized value of outcome y for household i in study k is defined as $\tilde{y}_{ik} = (y_{ik} - \bar{y}_k) / \bar{\sigma}_k$, where \bar{y}_k and $\bar{\sigma}_k$ are the mean and standard deviation in the control group.

²⁹Because of the experimental design of each study, we assume no systematic bias in estimates $\hat{\tau}_k$. For alternative approaches combining both observational studies and randomized controlled trials, see [Athey et al. \(2020\)](#) and [Gechter and Meager \(2021\)](#).

4.1 The hierarchical model

The hierarchical model offers an appealing statistical tool to disentangle the true heterogeneity (σ_τ^2) from sampling variation, to estimate the average impact of index insurance τ , and to explore the potential sources of heterogeneity. This model combines the available evidence in a structured way and is particularly suited to situations where data from multiple experimental studies are available (Andrews and Oster, 2019). In contrast with classical meta-analysis techniques (often referred to as the fixed-effect or pooling model), which consider that individual studies estimate a common effect, the hierarchical model allows for the presence of heterogeneous effects across studies. Because index insurance studies have been conducted in contexts that differ systematically from each other (see Tables 1 and 2), this model is expected to be more appropriate. The typical set-up specifies that each individual study estimates its own treatment effect τ_k , and that each individual τ_k is in turn drawn from a common distribution.³⁰ It can be described as follows:

$$\begin{aligned}\hat{\tau}_k &\sim N(\tau_k, \hat{s}e_k^2) \\ \tau_k &\sim N(\tau, \sigma_\tau^2)\end{aligned}\tag{2}$$

where $\hat{\tau}_k$ and $\hat{s}e_k^2$ are estimates of the treatment effect and sampling error in each individual study, and τ_k , τ and σ_τ^2 are defined as above. The first line of model (2) assumes that each individual $\hat{\tau}_k$ is an unbiased estimate of its own study-specific treatment effect τ_k (a reasonable assumption given the experimental design of the studies). The second line of model (2) assumes that treatment effects $\{\tau_1, \tau_2, \dots, \tau_K\}$ are normally distributed around τ and σ_τ^2 . This assumption ensures that the model recovers the results of classical meta-analysis if there is no heterogeneity in treatment effects ($\sigma_\tau^2 = 0$), and the results of initial studies if there is considerable heterogeneity in treatment effects ($\sigma_\tau^2 \rightarrow \infty$).³¹

We follow Meager (2019) and incorporate more structure to model (2) by using the original data from each study rather than just the reported estimates. The use of original data has two advantages. First, as noted earlier, it allows to standardize outcomes and to construct variables that were not always analyzed in the original studies. Second, original data allow

³⁰This approach to data aggregation was first pioneered by Rubin (1981).

³¹An additional assumption for the estimation of model (2) is that of exchangeability: the joint distribution of $\{\tau_1, \tau_2, \dots, \tau_K\}$ should be invariant to permutations of the K indices (Diaconis, 1977). The exchangeability assumption means that study characteristics are not sufficient to distinguish one study from another, and that there is no reason to expect the estimate from one study to be closer to the estimate from another study. This assumption is violated if researchers expect that the characteristics of some studies can drive their estimates close to each other.

to incorporate individual covariates in the model and thereby to explore the potential sources of heterogeneity across studies. The detailed hierarchical model with and without household-level covariates is presented in [Appendix B.1](#).

4.2 Bayesian estimation

In model (2), only $\hat{\tau}_k$ and $\hat{s}e_k^2$ are observable. Other parameters τ_k , τ , σ_τ^2 are unknown and should be estimated. In line with the recent literature, we use a Bayesian estimation method. The Bayesian methodology considers τ and σ_τ^2 as random variables and combines existing evidence with information from outside the data (the so-called "priors") to jointly estimate posterior distributions.³² In practice, we conduct inference using the *baggr* (short for "Bayesian aggregator") package – an *R* package designed by Rachael Meager and Witold Wiecek with the objective of facilitating the implementation and tractability of Bayesian meta-analysis ([Meager and Wiecek, 2022](#)).³³ We specify the following set of priors for the hypermean τ and the hyper-standard-deviations σ_τ .³⁴

$$\begin{aligned}\tau &\sim N(0, 100^2) \\ \sigma_\tau &\sim \text{Cauchy}(0, 5)\end{aligned}\tag{3}$$

5 Results

We use the Bayesian hierarchical model to provide evidence on (i) the average effect of index insurance, (ii) the degree of heterogeneity in effects, (iii) the potential sources of this heterogeneity, and (iv) the predicted effect of index insurance in a new study.

³²Technical details are provided in [Appendix B.2](#). This estimation method has several advantages compared to frequentist alternatives. Frequentist alternatives include methods such as empirical Bayes and maximum likelihood. See [Meager \(2019\)](#) for a summary of why Bayesian inference is preferable for data aggregation when the number of studies is limited. In short, Bayesian inference can reduce mean squared error and the risk of overfit.

³³In our pre-analysis plan, we specified we would use the "mutau" model, which incorporates data on the mean outcome of the control group and thereby increases precision. However, this is not possible because of the standardization of the outcomes by study (by definition the control mean is set to zero in all studies).

³⁴We assume a normal distribution with a zero mean assumption for the prior on τ because it has been shown to perform well in other recent applications ([Meager, 2019](#); [Bandiera et al., 2021](#)). The half-Cauchy prior on σ_τ is motivated by its capacity to allow the scaling to vary widely ([Meager, 2019](#)). We show that results are robust to the use of the default, data-driven priors of the *baggr* package.

5.1 The average effect of index insurance on production decisions

We first present evidence on τ , the average effect of index insurance on farmer production decisions. Estimates of τ provide important indications on the effect of index insurance in contexts that are comparable to the current set of contexts.

Figure 1 shows the posteriors obtained from the Bayesian hierarchical model as well as the estimates from the pooling model for comparisons. We find evidence from the pooling model that farmers cultivate more land and invest more in productive inputs when they are offered access to index insurance. The average amount of cultivated land is +0.09 SD higher among farmers receiving insurance offers (p-value = 0.013). Farmers use also more seeds (+0.10 SD), more pesticides (+0.09 SD), and more fertilizers (+0.07 SD) when they receive insurance offers. The effect on the risk index is positive but much smaller and not statistically significant, suggesting that access to index insurance does not lead farmers to cultivate more risky crops, or that such effects are too peculiar to be captured by our generic risk index. The latter explanation is perhaps more plausible because individual studies report subtle effects on crop choices.³⁵ In addition, our risk index ignores the imperfect correlation between crops, and some farmers might increase the number of crops in their portfolio to reduce their exposure to risk. Looking at these aspects in a systematic way is beyond the scope of this paper but is a fruitful area for future research.

Estimates from the Bayesian hierarchical model confirm that the effects of index insurance on productive investments are likely positive. The posterior mean of τ is positive for all outcomes but the risk index, and similar to estimates from the pooling model. Farmers cultivate more land (+0.11 SD), and use more seeds (+0.11 SD), more pesticides (+0.06 SD), and more fertilizers (+0.08 SD). However, the distributions of the posterior suggest that effects are more uncertain than in the pooling model. All 95% posterior intervals now include zero, and, for some outcomes, there is a sizable probability of negative effects. The 95%-interval is [-0.02 SD, +0.25 SD] for cultivated area, [-0.09 SD, +0.09 SD] for the crop riskiness index, [-0.05 SD, +0.21 SD] for fertilizers, [-0.02 SD, +0.15 SD] for pesticides, and [-0.02 SD, +0.25 SD] for seeds. Results are broadly robust to using the automatic priors (Figure A1) and to excluding any single study (Figure A2).

³⁵For example, in India, [Cole et al. \(2017\)](#) show that the provision of index insurance induces farmers to shift production towards cash crops. In Burkina Faso, [Stoeffler et al. \(2022\)](#) show that the provision of an index insurance contract designed for cotton farmers led them to cultivate more sesame (an attractive cash crop in this context because it requires lower input and time investments).

These results tend to confirm the potential of index insurance to foster smallholder farmers' productive investments, but they also suggest that effects are more uncertain than suggested by the pooling model. The pooling model is misleading because it considers that impacts are homogeneous across studies (thus leading to minimize the uncertainty around treatment effects). The Bayesian hierarchical model offers a practical framework to detect heterogeneity in treatment effects and to take it into account while estimating the average effect of index insurance. In what follows, we use the Bayesian hierarchical model to provide evidence on the true degree of heterogeneity in treatment effects across studies, to explore the possible sources of this heterogeneity, and to predict the likely effect of index insurance in a new context.

5.2 The heterogeneity in treatment effects

Figure 2 shows the posteriors of study-specific treatment effects τ_k as well as the no-pooling OLS estimates. While most point estimates suggest positive effects, there is considerable heterogeneity in the individual coefficients. Estimates for cultivated land vary from a minimum of -0.05 SD in Burkina Faso to a maximum of $+0.44$ SD in Ghana.³⁶ The 95% confidence intervals also vary largely: from $[-0.32$ SD, $+0.23$ SD] in Burkina Faso to $[+0.22$ SD, $+0.67$ SD] in Ghana. It is not clear however how much of this variation stems from true heterogeneity in treatment effects vs. random variation due to sampling errors. The Bayesian hierarchical model (BHM) allows to separate the true heterogeneity in treatment effects from the sampling variation. It is therefore not surprising that BHM estimates display less variation than OLS estimates (Figure 2). Nevertheless, substantial variation persists. The posterior means for cultivated area vary from $+0.03$ SD in Burkina Faso to $+0.28$ SD in Ghana. The corresponding 95% posterior intervals are $[-0.07$ SD, $+0.12$ SD] and $[+0.07$ SD, $+0.53$ SD]. Overall, these results suggest that τ_k is heterogeneous across studies and that pooling the data to estimate a common treatment effect, as done in traditional meta-analyses, is misleading. The Bayesian hierarchical model offers a more reasonable framework.

We now use estimates of σ_τ^2 , the true heterogeneity in treatment effects across studies, to derive metrics to gauge the external validity of index insurance experiments. If $\hat{\sigma}_\tau^2$ is close to

³⁶As mentioned above, these estimates correspond to the ITT effects of index insurance offers. One potential driver of heterogeneity in treatment effects is the variation in take-up rates across studies (take-up rates vary from 29% in Mali to 100% in Ghana and India). In Figure A3, we use the random assignment to insurance offers to instrument actual take-up and to compare ITT effects with average treatment effects on the treated (ATET). If anything, we see larger heterogeneity in the ATET point estimates, which suggests that the variation in take-up rates alone does not explain heterogeneity in treatment effects, and that the external validity of the ITT effect might actually be larger than the external validity of the ATET.

zero, there is almost no heterogeneity across studies and $\hat{\tau}$ provides a better estimate of each τ_k than its corresponding $\hat{\tau}_k$ – external validity is high. Alternatively, if $\hat{\sigma}_\tau^2$ is large, the heterogeneity across studies is important and each $\hat{\tau}_k$ provides a better measure of its corresponding τ_k – external validity is low. An important question, however, is that of what constitutes a large or small value of $\hat{\sigma}_\tau^2$. To interpret the magnitude of $\hat{\sigma}_\tau^2$, a range of pooling metrics have been developed in the literature, the most prominent of which is perhaps the pooling factor, defined as $\lambda_k = \hat{s}e_k^2 / (\hat{s}e_k^2 + \hat{\sigma}_\tau^2)$. The main advantage of the pooling factor is that it has a straightforward interpretation: the potential values of λ_k are restricted to the interval $[0, 1]$, and any values above 0.5 indicate a dominance of sampling variation over heterogeneity in treatment effects. For each outcome, we report $\lambda(\tau)$, the pooling factor averaged across all studies, which can be interpreted as the percentage of total variation in treatment effects stemming from sampling variation. We also report evidence on two complementary metrics: the generalized pooling factor and the brute force.³⁷

Table 3 reports evidence on these pooling metrics. We see some differences depending on the outcome considered, with 38% of the observed variation in treatment effects due to sampling variation for cultivated area and fertilizers, 40% for seeds, 69% for pesticides, and 76% for the crop index. Averaged across all outcomes, we find moderate pooling of information across studies: 52% of the observed variation in treatment effects stems from sampling variation. We note however the high level of uncertainty around the estimates of the pooling factor: averaged across all outcomes, the interval around the pooling factor is $[0.14, 0.94]$. This moderate and imprecise level of pooling helps to understand why in Figure 1 the BHM posteriors of τ display more uncertainty than the pooled estimates, and why in Figure 2 substantial differences subsist in the BHM estimates of τ_k (especially for the outcomes with the smallest estimated pooling factor: cultivated area, fertilizers, and seeds). Results are of a similar magnitude using the brute force and the generalized pooling factor. Overall, this suggests that offering index insurance to smallholder farmers generates effects that are fairly heterogeneous across contexts.

5.3 The sources of heterogeneity in treatment effects

We look at five household-level covariates that could help to explain the relatively high level of heterogeneity in treatment effects we document in the previous subsection: a household wealth index, household head age, household head literacy, household size, and household predicted

³⁷More details on these metrics are provided in [Appendix B.3](#).

outcomes. We define dummy variables equal to one if the age of the household head is above the median, if the head is literate, if the wealth index is above the median, if the household size is above the median, and if the predicted outcome is above the median.³⁸ For each outcome and each covariate, we report the BHM posteriors of the average effect of index insurance offers for the group with the covariate equal to zero as well as the additional effect for the group with the covariate equal to one (Figure 3). In addition, using the four studies designed as price experiments, we report the effect of receiving insurance offers at a price below vs. above average. The results show limited heterogeneity along the covariates we study. We find some evidence that treatment effects are larger for households with low levels of predicted outcomes. The effect on seeds is +0.12 SD higher for households who would have used less seeds in the absence of index insurance offers. The effect on the risk index is +0.18 SD higher for households who would have cultivated less risky crops in the absence of index insurance offers.³⁹ Treatment effects also seem larger among worse-off households (according to the wealth index), among households with a younger head, and among households offered a lower price for the index insurance product – but posterior distributions include zero comfortably for these covariates. Results for household head literacy and household size display little heterogeneity. Overall, we find limited evidence that the heterogeneity in treatment effects can be explained using basic household and product characteristics. This suggests that most of the heterogeneity in treatment effects stems from more complex household/product characteristics (e.g. preferences or trust levels) as well as from contextual variables. Exploring the role of these covariates is a fruitful – and challenging – area for future research.

5.4 The predicted effect of index insurance in a new study

We use estimates from the Bayesian hierarchical model on σ_τ^2 to predict the effect of index insurance offers in a new context (τ_{k+1}). Figure 4 reports the posterior means and intervals of τ_{k+1} as well as the estimates from the pooling model for comparisons. Overall, we find very uncertain predicted effects. The posterior intervals of τ_{k+1} are considerably wider than the confidence intervals from the pooling model, reflecting the sizable heterogeneity in treatment effects documented in Section 5.2, especially for outcomes with the largest heterogeneity (seeds, fertilizers

³⁸See Section 3.2 for more details.

³⁹These patterns are visible across all the studies, although the 95% posterior distributions sometimes include zero (Figure A6). While interesting, these patterns must be interpreted with caution due to the high number of parameters estimated and to the potential bias resulting from overfitting (see footnote 26). In addition, for these analysis, we could not cluster standard errors at a higher level than the household in *baggr*.

and cultivated area). If we were to run a new experiment, the chances of obtaining null or negative effects would be non-negligible. For example, the treatment effect on fertilizer has a 50% chance of being between -0.00 SD and $+0.17$ SD, a 25% chance of being negative, and a 25% chance of being larger than $+0.17$ SD. The posterior intervals of the predicted effects are considerably wider than the confidence intervals from the pooling model (by a factor from 2 to 5). Findings are similar using automatic priors (Figure A7), and estimates of τ_{k+1} along the different covariates are very uncertain too (Figure A8). Overall, this evidence suggests the existing evidence base offers limited insights to predict the effects of index insurance offers in new settings because of the sizable heterogeneity in treatment effects across studies. Conducting new studies in such settings is particularly worthwhile.

6 Conclusion

Farmers in low- and middle-income countries are exposed to severe risks and live in areas where formal insurance markets are generally absent. In recent years, a solution that has attracted considerable attention to help smallholder farmers manage risks is index insurance. In this paper, we aggregated evidence from six experiments on index insurance and explored the external validity of their results. We find positive but highly heterogeneous effects on farmers' productive investments, with no evidence that this heterogeneity can be meaningfully explained by basic household characteristics. This suggests that most of the heterogeneity in treatment effects might stem from more complex household characteristics or from contextual variables such as study characteristics or product design. Overall, we uncover moderate external validity and limited insights to predict the impact of index insurance in new settings. We estimate a sizable likelihood of finding null or even negative effects in new settings.

These findings should in no way be interpreted as evidence that efforts to promote index insurance should be abandoned. In fact, they tend to confirm that index insurance has the potential to boost smallholders' productive investments. We estimate that interventions expanding access to index insurance typically boost productive investments by 0.06–0.11 standard deviation on average. However, our results also suggest that effects are highly uncertain, and, together with evidence on the limited uptake of index-based products, they emphasize that governments and development agencies should remain cautious before investing in their widespread promotion. Recent innovations in the design and implementation of index-based

products have raised expectations that new products might overcome some of the problems faced by first-generation products as well as traditional indemnity-based products. For example, products taking advantage of the latest advances in remotely sensed data systems and digital technologies could further reduce transaction costs and basis risk. However, more evidence is required before index insurance can be promoted as a cost-effective solution to help smallholder farmers manage risks at scale. Moreover, recent studies in South Asia have found promising results for alternative risk-management tools, including stress-tolerant crops (Emerick et al., 2016), emergency credit lines (Lane, 2018), and access to weather forecasting (Rosenzweig and Udry, 2019). Assessing their cost-effectiveness as well as their complementarity with insurance products constitute promising areas for further research.

Methodologically, our paper illustrates both the potential and limitations of Bayesian hierarchical models for economic research. Such models provide a powerful tool to aggregate evidence from an existing evidence base by allowing for the presence of heterogeneous effects across studies. Access to the underlying micro-data can further help to examine dimensions that are not always reported in the original studies, either because of a lack of statistical power or because of researchers' own preferences. In addition, Bayesian hierarchical models provide a unique tool to analyze heterogeneity across contexts and explore questions around external validity. Our paper, however, suggests that assessing external validity with few studies can be challenging, as shown by the considerable uncertainty around the pooling factor. We believe that this result highlights the need for more data and more accurate metrics. The ongoing accumulation of evidence from various contexts and the rapid transition towards more open norms around data sharing give rise to cautious optimism. However, given the typical size of evidence bases in economics, our expectation is that accurately answering questions around external validity will remain difficult using current metrics.

References

- Abadie, A., Chingos, M. M., and West, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, 100(4):567–580.
- Ahmed, S., McIntosh, C., and Sarris, A. (2020). The impact of commercial rainfall index insurance: Experimental evidence from Ethiopia. *American Journal of Agricultural Economics*, 102(4):1154–1176.
- Ali, W., Abdulai, A., and Mishra, A. K. (2020). Recent advances in the analyses of demand for agricultural insurance in developing and emerging countries. *Annual Review of Resource Economics*, 12:411–430.
- Andrews, I. and Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, 109(8):2766–94.
- Andrews, I. and Oster, E. (2019). A simple approximation for evaluating external validity bias. *Economics Letters*, 178:58–62.
- Athey, S., Chetty, R., and Imbens, G. (2020). Combining experimental and observational data to estimate treatment effects on long term outcomes. Mimeo.
- Bandiera, O., Fischer, G., Prat, A., and Ytsma, E. (2021). Do women respond less to performance pay? Building evidence from multiple experiments. *American Economic Review: Insights*, 3(4):435–54.
- Bédécarrats, F., Guérin, I., and Roubaud, F. (2020). *Randomized control trials in the field of development: A critical perspective*. Oxford University Press.
- Belissa, T., Bulte, E., Cecchi, F., Gangopadhyay, S., and Lensink, R. (2019). Liquidity constraints, informal institutions, and the adoption of weather insurance: A randomized controlled trial in Ethiopia. *Journal of Development Economics*, 140:269–278.
- Belissa, T. and Marr, A. (2018). Effects of bundled index-based insurance with credit and agricultural inputs on uptake, investment, productivity and welfare of smallholder farmers in Ethiopia. Mimeo.
- Berhane, G., Dercon, S., Hill, R. V., and Taffesse, A. (2015). Formal and informal insurance: Experimental evidence from Ethiopia. Mimeo.
- Bertram-Huemmer, V. and Kraehnert, K. (2018). Does index insurance help households recover from disaster? Evidence from IBLI Mongolia. *American Journal of Agricultural Economics*, 100(1):145–171.
- Binswanger-Mkhize, H. P. (2012). Is there too much hype about index-based agricultural insurance? *Journal of Development Studies*, 48(2):187–200.
- Bo, H. and Galiani, S. (2021). Assessing external validity. *Research in Economics*, 75(3):274–285.

- Boucher, S. R., Carter, M. R., Flatnes, J. E., Lybbert, T. J., Malacarne, J. G., Marenya, P., and Paul, L. A. (2021). Bundling genetic and financial technologies for more resilient and productive small-scale agriculture. NBER Working Paper No. 29234.
- Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.
- Bulte, E., Cecchi, F., Lensink, R., Marr, A., and Van Asseldonk, M. (2019). Do crop insurance-certified seed bundles crowd-in investments? Experimental evidence from Kenya. *Journal of Economic Behavior and Organization*.
- Burlig, F. (2018). Improving transparency in observational social science research: A pre-analysis plan approach. *Economics Letters*, 168:56–60.
- Cai, J. (2016). The impact of insurance provision on household production and financial decisions. *American Economic Journal: Economic Policy*, 8(2):44–88.
- Cai, J., de Janvry, A., and Sadoulet, E. (2020). Subsidy policies and insurance demand. *American Economic Review*, 110(8):2422–53.
- Cai, J. and Song, C. (2017). Do disaster experience and knowledge affect insurance take-up decisions? *Journal of Development Economics*, 124:83–94.
- Carter, M., de Janvry, A., Sadoulet, E., and Sarris, A. (2017). Index insurance for developing country agriculture: A reassessment. *Annual Review of Resource Economics*, 9:421–438.
- Casaburi, L. and Willis, J. (2018). Time versus state in insurance: Experimental evidence from contract farming in Kenya. *American Economic Review*, 108(12):3778–3813.
- Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics*, 127(4):1755–1812.
- Cecchi, F., Lensink, R., and Slingerland, E. (2022). Loss aversion, ambiguity attitudes, and willingness to pay for index insurance: Experimental evidence from rural Kenya. Mimeo.
- Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.
- Cilliers, J., Jack, W., and Zeitlin, A. (2020). Demand for and impacts of mobile phone-based index insurance in agriculture: Experimental evidence from Kenya. AEA RCT Registry. March 04. <https://doi.org/10.1257/rct.5170-1.0>.
- Clarke, D. J. (2016). A theory of rational demand for index insurance. *American Economic Journal: Microeconomics*, 8(1):283–306.
- Clement, K. Y., Botzen, W. W., Brouwer, R., and Aerts, J. C. (2018). A global review of the impact of basis risk on the functioning of and demand for index insurance. *International Journal of Disaster Risk Reduction*, 28:845–853.

- Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., and Vickery, J. (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1):104–35.
- Cole, S., Giné, X., and Vickery, J. (2017). How does risk management influence production decisions? Evidence from a field experiment. *The Review of Financial Studies*, 30(6):1935–1970.
- Cole, S. A. and Xiong, W. (2017). Agricultural insurance and economic development. *Annual Review of Economics*, 9:235–262.
- De Janvry, A., Dequiedt, V., and Sadoulet, E. (2014). The demand for insurance against common shocks. *Journal of Development Economics*, 106:227–238.
- Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210:2–21.
- Dehejia, R., Pop-Eleches, C., and Samii, C. (2021). From local to global: External validity in a fertility natural experiment. *Journal of Business & Economic Statistics*, 39(1):217–243.
- del Valle, A., de Janvry, A., and Sadoulet, E. (2020). Rules for recovery: Impact of indexed disaster funds on shock coping in Mexico. *American Economic Journal: Applied Economics*, 12(4):164–95.
- Dercon, S., Hill, R. V., Clarke, D., Outes-Leon, I., and Taffesse, A. S. (2014). Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in Ethiopia. *Journal of Development Economics*, 106:132–143.
- Diaconis, P. (1977). Finite forms of de Finetti’s theorem on exchangeability. *Synthese*, 36(2):271–281.
- Elabed, G. and Carter, M. (2014). Ex-ante impacts of agricultural insurance: Evidence from a field experiment in Mali. Mimeo.
- Emerick, K., De Janvry, A., Sadoulet, E., and Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, 106(6):1537–61.
- Gaurav, S., Cole, S., and Tobacman, J. (2011). Marketing complex financial products in emerging markets: Evidence from rainfall insurance in India. *Journal of Marketing Research*, 48(SPL):S150–S162.
- Gechter, M. (2015). Generalizing the results from social experiments: Theory and evidence from Mexico and India. Mimeo.
- Gechter, M. and Meager, R. (2021). Combining experimental and observational studies in meta-analysis: A mutual debiasing approach. Mimeo.
- Gehrke, E. (2019). An employment guarantee as risk insurance? Assessing the effects of the NREGS on agricultural production decisions. *The World Bank Economic Review*, 33(2):413–435.

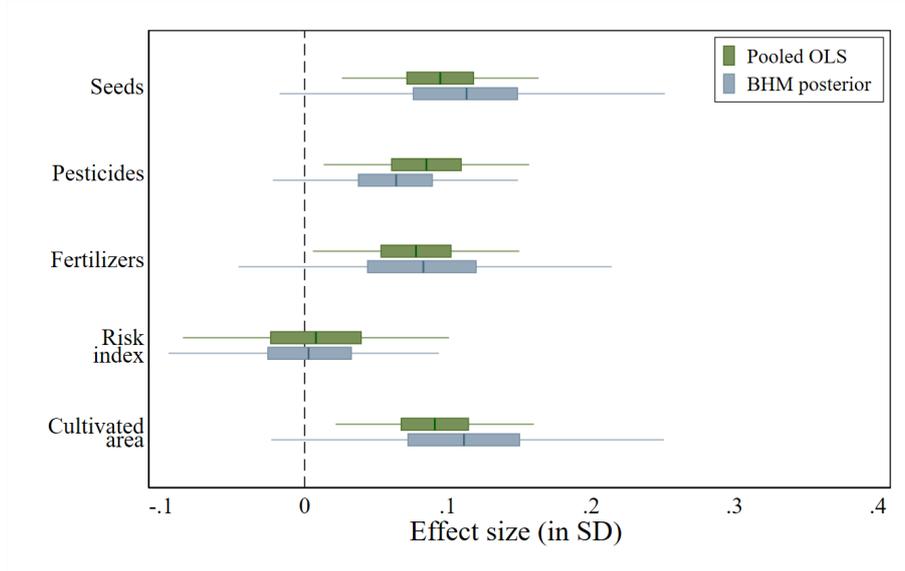
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Third Edition. Chapman and Hall/CRC.
- Giné, X., Townsend, R., and Vickery, J. (2008). Patterns of rainfall insurance participation in rural India. *The World Bank Economic Review*, 22(3):539–566.
- Giné, X. and Yang, D. (2009). Insurance, credit, and technology adoption: Field experimental evidence from Malawi. *Journal of Development Economics*, 89(1):1–11.
- Hazell, P. B. (1992). The appropriate role of agricultural insurance in developing countries. *Journal of International Development*, 4(6):567–581.
- Hill, R. V., Kumar, N., Magnan, N., Makhija, S., de Nicola, F., Spielman, D. J., and Ward, P. S. (2019). Ex ante and ex post effects of hybrid index insurance in Bangladesh. *Journal of Development Economics*, 136:1–17.
- Humphreys, M., De la Sierra, R. S., and Van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 21(1):1–20.
- Ioannidis, J., Stanley, T., and Doucouliagos, C. (2017). The power of bias in economics research. *The Economic Journal*, 127(605):236–265.
- J-PAL, CEQA, and ATAI (2016). Make it rain. Policy Bulletin, Cambridge, MA: Abdul Latif Jameel Poverty Action Lab, Center for Effective Global Action, and Agricultural Technology Adoption Initiative.
- Jackson, C. K. and Mackevicius, C. (2021). The distribution of school spending impacts. NBER Working Paper No. 28517.
- Janzen, S. A. and Carter, M. R. (2019). After the drought: The impact of microinsurance on consumption smoothing and asset protection. *American Journal of Agricultural Economics*, 101(3):651–671.
- Jensen, N. and Barrett, C. (2017). Agricultural index insurance for development. *Applied Economic Perspectives and Policy*, 39(2):199–219.
- Jensen, N. D., Barrett, C. B., and Mude, A. G. (2017). Cash transfers and index insurance: A comparative impact analysis from northern Kenya. *Journal of Development Economics*, 129:14–28.
- Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, 129(2):597–652.
- Kazianga, H. and Wahhaj, Z. (2021). Enhancing access to index-based weather agricultural insurance: a new marketing approach in Burkina Faso. AEA RCT Registry. January 13. <https://doi.org/10.1257/rct.5180-1.0>.

- Kramer, B., Cecchi, F., Kivuva, B., and Waithaka, L. (2021). Improving agricultural productivity and resilience with cellphone imagery to scale climate-smart crop insurance. AEA RCT Registry. September 27. <https://doi.org/10.1257/rct.7435-1.4000000000000001>.
- Kramer, B., Hazell, P., Alderman, H., Ceballos, F., Kumar, N., and Timu, A. G. (2022). Is agricultural insurance fulfilling its promise for the developing world? A review of recent evidence. *Annual Review of Resource Economics*, 14.
- Kremer, M., Luby, S., Maertens, R., Tan, B., and Więcek, W. (2022). Water treatment and child mortality: A meta-analysis and cost-effectiveness analysis. Becker Friedman Institute for Economics Working Paper.
- Lane, G. (2018). Credit lines as insurance: Evidence from Bangladesh. Mimeo.
- Marr, A., Winkel, A., van Asseldonk, M., Lensink, R., and Bulte, E. (2016). Adoption and impact of index-insurance and credit for smallholder farmers in developing countries. *Agricultural Finance Review*.
- Meager, R. (2019). Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, 11(1):57–91.
- Meager, R. (2022). Aggregating distributional treatment effects: A bayesian hierarchical analysis of the microcredit literature. *American Economic Review*. Forthcoming.
- Meager, R. and Wiecek, W. (2022). *Baggr*, an R package for Bayesian meta-analysis using Stan. Available at <https://github.com/wwiecek/baggr>.
- Michler, J. D., Viens, F. G., and Shively, G. E. (2022). Risk, crop yields, and weather index insurance in village India. *Journal of the Agricultural and Applied Economics Association*.
- Miranda, M. J. and Farrin, K. (2012). Index insurance for developing countries. *Applied Economic Perspectives and Policy*, 34(3):391–427.
- Mishra, K., Gallenstein, R. A., Miranda, M. J., Sam, A. G., Toledo, P., and Mulangu, F. (2021). Insured loans and credit access: Evidence from a randomized field experiment in northern Ghana. *American Journal of Agricultural Economics*, 103(3):923–943.
- Mobarak, A. M. and Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review: Papers & Proceedings*, 103(3):375–80.
- Ogden, T. (2020). RCTs in development economics, their critics and their evolution. In Bédécarrats, F., Guérin, I., and Roubaud, F., editors, *Randomized control trials in the field of development: A critical perspective*, chapter 4, pages 126–151. Oxford University Press, Oxford.
- Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29(3):61–80.

- Peters, J., Langbein, J., and Roberts, G. (2018). Generalization in the tropics—development policy, randomized controlled trials, and external validity. *The World Bank Research Observer*, 33(1):34–64.
- Platteau, J.-P., De Bock, O., and Gelade, W. (2017). The demand for microinsurance: A literature review. *World Development*, 94:139–156.
- Pritchett, L. and Sandefur, J. (2014). Context matters for size: why external validity claims and development practice do not mix. *Journal of Globalization and Development*, 4(2):161–197.
- Ravallion, M. (2020). Should the randomistas (continue to) rule? In Bédécarrats, F., Guérin, I., and Roubaud, F., editors, *Randomized control trials in the field of development: A critical perspective*, chapter 1, pages 47–78. Oxford University Press, Oxford.
- Romero, M., Sandefur, J., and Sandholtz, W. A. (2020). Outsourcing education: Experimental evidence from Liberia. *American Economic Review*, 110(2):364–400.
- Rosenzweig, M. R. and Udry, C. (2019). Assessing the benefits of long-run weather forecasting for the rural poor: Farmer investments and worker migration in a dynamic equilibrium model. NBER Working Paper No. 25894.
- Rosenzweig, M. R. and Udry, C. (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies*, 87(1):343–381.
- Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6(4):377–401.
- Stein, D. (2018). Dynamics of demand for rainfall index insurance: Evidence from a commercial product in India. *The World Bank Economic Review*, 32(3):692–708.
- Stoeffler, Q., Carter, M., Guirking, C., and Gelade, W. (2022). The spillover impact of index insurance on agricultural investment by cotton farmers in Burkina Faso. *The World Bank Economic Review*, 36(1):114–140.
- Stoeffler, Q. and Opuz, G. (2022). Price, information and product quality: Explaining index insurance demand in Burkina Faso. *Food Policy*, 108:102213.
- Vieider, F., Epper, T., Vieider, F., Abdellaoui, M., and Kemel, E. (2021a). Go for it while you can: Insurance panel experiment. AEA RCT Registry. June 13. <https://doi.org/10.1257/rct.7084-1.2000000000000002>.
- Vieider, F., Epper, T., Vieider, F., Verschoor, A., and D’Exelle, B. (2021b). Go for it: Effects of insurance on investment. AEA RCT Registry. June 14. <https://doi.org/10.1257/rct.7639-1.0>.
- Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association*, 18(6):3045–3089.

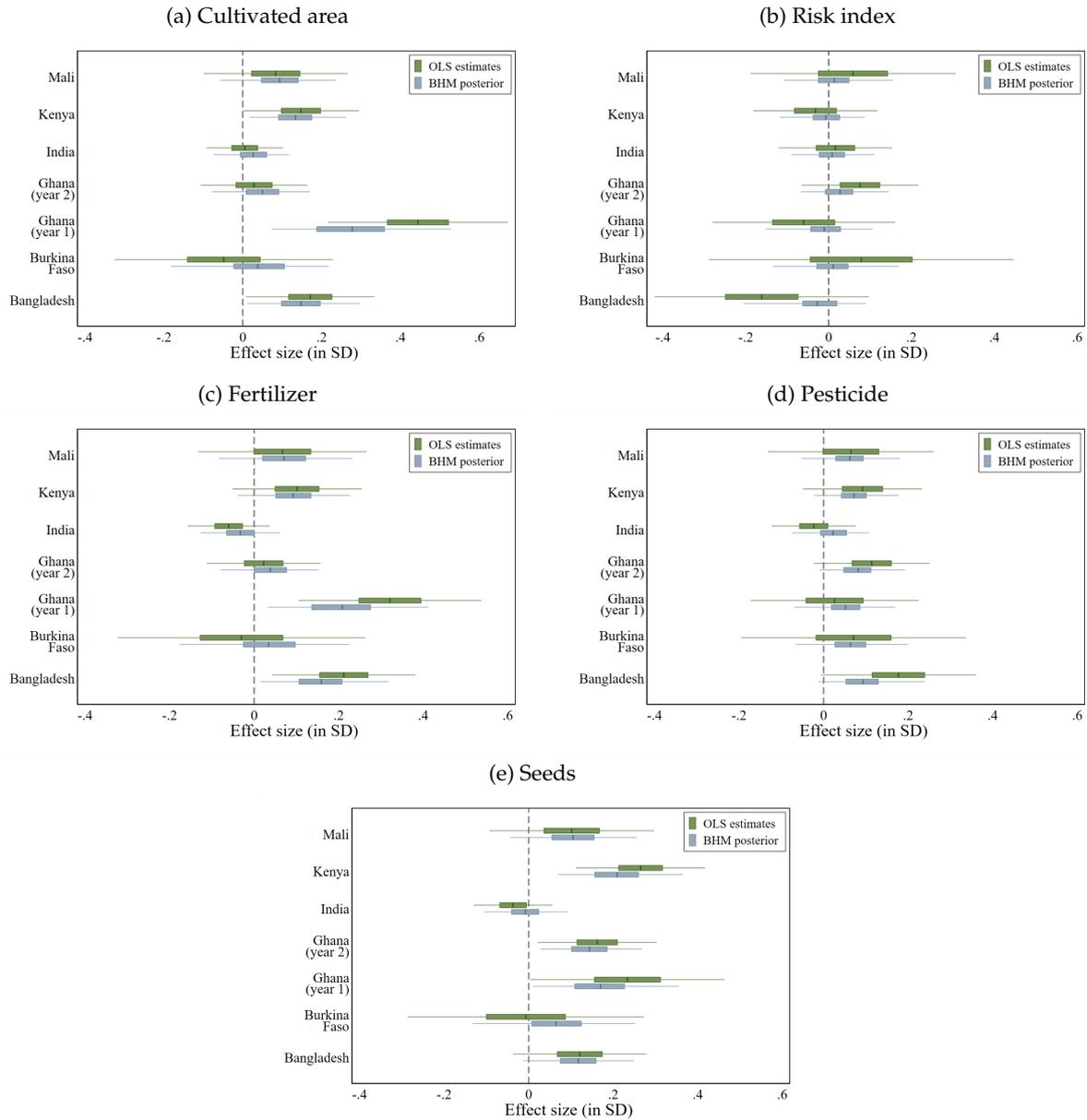
Main figures and tables

Figure 1: The average effect of index insurance on production decisions (τ)



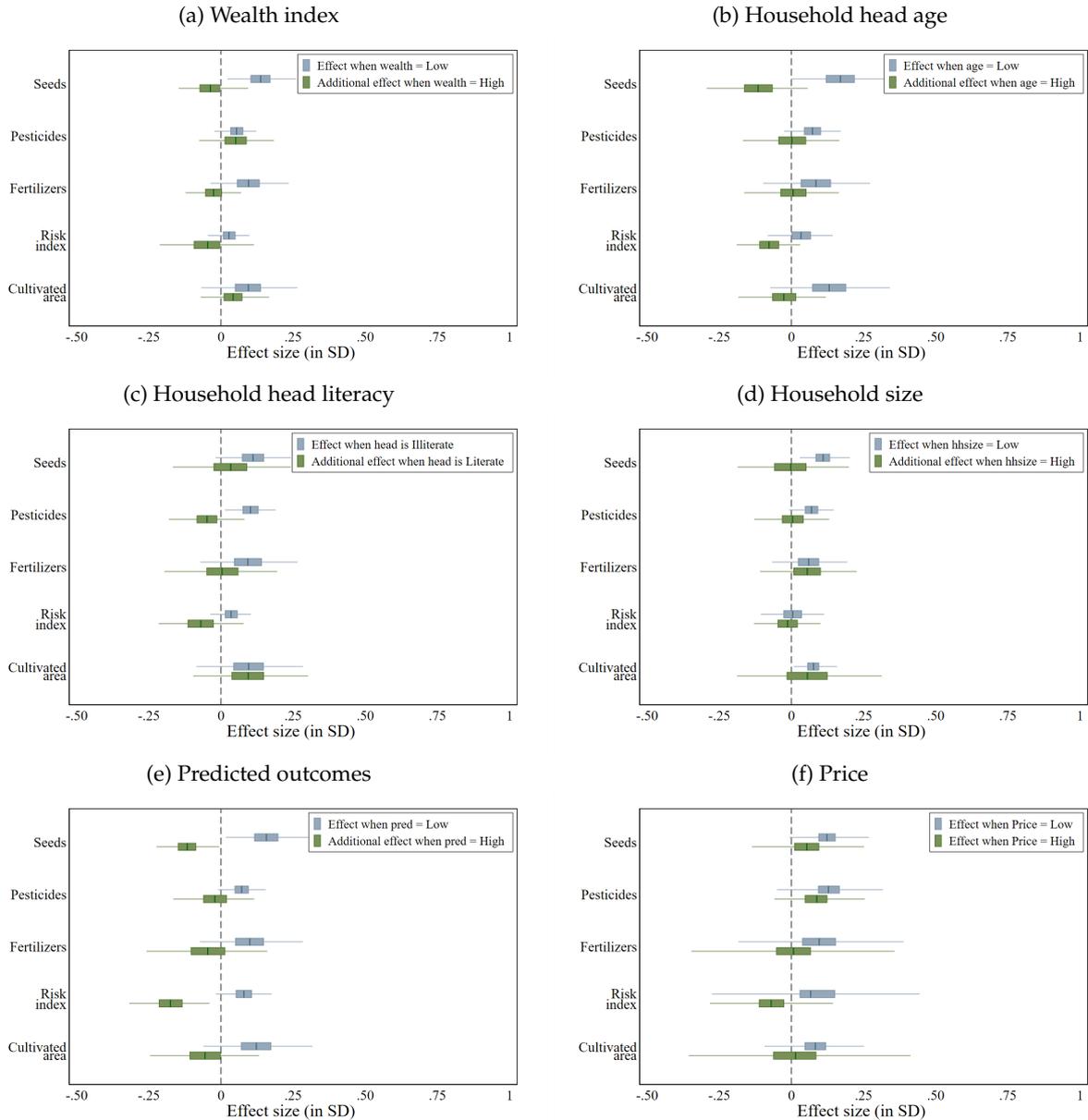
Notes: This figure presents evidence on the average effect of randomly assigned offers. Estimates from pooled OLS assume that impacts are homogeneous across studies whereas BHM estimates allow for the presence of heterogeneous effects. Thin (resp. thick) lines represent the 95% (resp. 50%) confidence/posterior intervals. All outcomes have been standardized using data from the original studies. Standard errors are clustered at the level of the randomization (i.e., household or village depending on the study). Source: Authors' elaboration using data from [Elabed and Carter \(2014\)](#); [Karlan et al. \(2014\)](#); [Cole et al. \(2017\)](#); [Bulte et al. \(2019\)](#); [Hill et al. \(2019\)](#); [Stoeffler et al. \(2022\)](#).

Figure 2: The heterogeneity of study-specific treatment effects (τ_k)



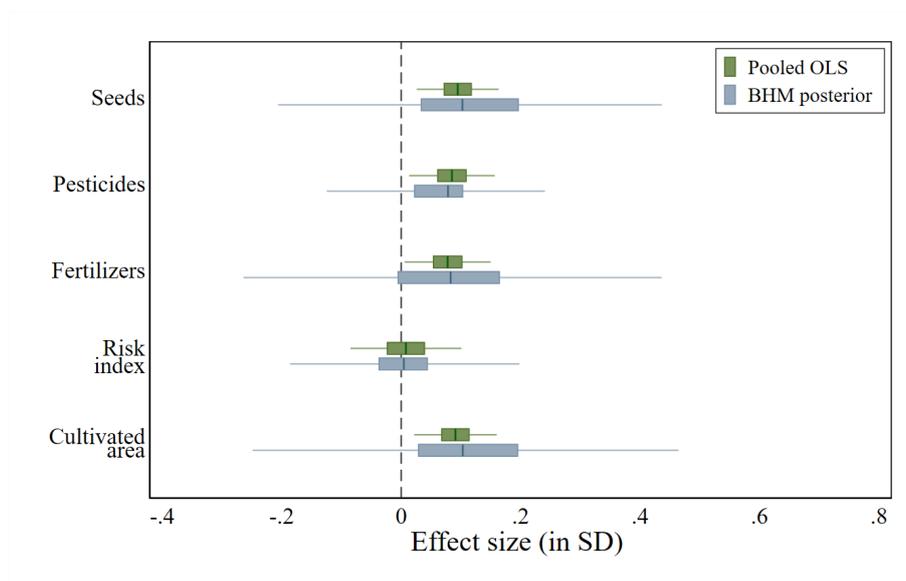
Notes: These figures present the BHM posteriors of study-specific treatment effects τ_k as well as the no-pooling OLS estimates for comparisons. The Bayesian hierarchical model allows to separate the genuine heterogeneity in treatment effects from the sampling variation. Thin (resp. thick) lines represent the 95% (resp. 50%) confidence/posterior intervals. All outcomes have been standardized using data from the original studies. Standard errors are clustered at the level of the randomization (i.e., household or village depending on the study). Source: Authors' elaboration using data from [Elabed and Carter \(2014\)](#); [Karlán et al. \(2014\)](#); [Cole et al. \(2017\)](#); [Bulte et al. \(2019\)](#); [Hill et al. \(2019\)](#); [Stoeffler et al. \(2022\)](#).

Figure 3: The heterogeneity of average treatment effects by key covariates



Notes: These figures present the BHM posteriors of the average effect of index insurance offers by key household covariates. We define dummy variables equal to one if the age of the household head is above the median, if the head is literate, if the wealth index is above the median, if the household size is above the median, and if the predicted outcome is above the median. Effects for the group with the covariate equal to zero are shown in blue. Additional effects for the group with the covariate equal to one are shown in green. In addition, using the four studies designed as price experiments, we report the effect of receiving insurance offers at a price below (blue) vs. above (green) average. Thin (resp. thick) lines represent the 95% (resp. 50%) confidence/posterior intervals. All outcomes have been standardized using data from the original studies. Standard errors are clustered at the household level. Source: Authors' elaboration using data from [Elabed and Carter \(2014\)](#); [Karlan et al. \(2014\)](#); [Cole et al. \(2017\)](#); [Bulte et al. \(2019\)](#); [Hill et al. \(2019\)](#); [Stoeffler et al. \(2022\)](#).

Figure 4: The predicted effect of index insurance in a new study (τ_{k+1})



Notes: This figure presents evidence on the predicted effect of increasing access to index insurance. Estimates from pooled OLS assume that impacts are homogeneous across studies. BHM estimates allow for the presence of heterogeneous effects and use evidence on σ_{τ}^2 to predict the effect of index insurance in a new context. Thin (resp. thick) lines represent the 95% (resp. 50%) confidence/posterior intervals. All outcomes have been standardized using data from the original studies. Standard errors are clustered at the level of the randomization (i.e., household or village depending on the study). Source: Authors' elaboration using data from [Elabed and Carter \(2014\)](#); [Karlan et al. \(2014\)](#); [Cole et al. \(2017\)](#); [Bulte et al. \(2019\)](#); [Hill et al. \(2019\)](#); [Stoeffler et al. \(2022\)](#).

Table 1: Studies description

	Bangladesh	Burkina Faso	Ghana		India	Kenya	Mali
	(1)	(2)	Year 1 (3)	Year 2 (4)	(5)	(6)	(7)
Citation	Hill et al. (2019)	Stoeffler et al. (2022)	Karlan et al. (2014)	Ibid.	Cole et al. (2017)	Bulte et al. (2019)	Elabed and Carter (2014)
Published	Yes	Yes	Yes	Ibid.	Yes	Yes	No
Data publicly available ^a	No ^b	No	No	Ibid.	Yes	No	No
Start year	2013	2013	2009	2010	2009	2016	2011
Experimental design	Price experiment ^c	Price experiment	Free provision	Price experiment	Free provision	Free provision ^d	Price experiment
Randomization level	Cluster	Cluster	Household	Cluster, household	Household	Household	Cluster
Additional treatment	None	None	Cash grants	Ibid.	Fertilizer coupons	None	None
Eligibility criteria	Membership to GUK ^e	Membership to a farmer group operating with Sofitex (cotton purchasing company)	Maize cultivation, less than 15 acres	Ibid.	Land ownership	Membership to a farmer group	Membership to a cotton cooperative
Sampling frame	Villages of the Bogra region	Farmer groups of the Houndé region	Villages in Northern Ghana included in the GLSS5+ survey	Expansion to villages within 30 kilometers of rain gauges	Villages in two districts in Andhra Pradesh, Mahbubnagar and Anantapur	Farmer groups of the Meru region	Cooperatives of the Bougouni region
Baseline survey	Yes	Yes	Yes	Ibid.	Yes	Yes	No
Sample size	2,300	1,015	385 ^f	1406 ^f	1,479	780	981
Number of clusters	120	80	60	72	45	40	87

Source: Authors' elaboration.

^a As of August 2020 (when we registered our study and pre-analysis plan).

^b The published article contains an online appendix with supplementary data. However, the data set only includes the variables analyzed in the original study.

^c Some of the subsidies took the form of rebates.

^d Free provision was conditional on purchasing improved seeds.

^e GUK: Gram Unnayan Karma – local NGO providing a range of services to households in Bogra region, including microfinance, non-formal primary education, primary healthcare, and women's empowerment activities.

^f We exclude households receiving only cash grants.

Table 2: Design of the insurance products

	Bangladesh	Burkina Faso	Ghana		India	Kenya	Mali
	(1)	(2)	Year 1 (3)	Year 2 (4)	(5)	(6)	(7)
Type of insurance	Hybrid	Area-yield	Rainfall	Ibid.	Rainfall	Rainfall ^a	Area-yield
Targeted crop	Not crop-specific ^b	Cotton	Maize	Ibid.	Not crop-specific	Four crops	Cotton
Sale price (in PPP USD) ^c	0.3 to 3.3 USD	12.6 to 50.5 USD	Free	1.1 to 15.3 USD	Free	Free (conditional on purchasing improved seeds)	15.8 to 31.6 USD
Actuarially fair price (in PPP USD) ^{c,d}	3.6 USD	28.9 USD	38.8 USD	7.9 to 10.3 USD	18.0 USD	5.2 to 13.8 USD	31.6 USD
Purchase decision level	Individual	Cluster	Individual	Ibid.	Individual	Individual	Cluster
Unit	Stand-alone policy ^b	1 ha	0.4 ha	Ibid.	Stand-alone policy	0.4 ha	1 ha
Timing of payment	Up-front	After harvest	NA	Up-front	NA	Up-front	After harvest
Take-up ^e	87%	45%	100%	63%	100%	59%	29%
Payoff triggers	Dry spell duration, low average of area yields	Low average of farmer group yields	Number of monthly dry/wet days	Ibid.	Cumulative rainfall	Rainfall excess/deficit	Low average of cooperative yields
Source of data	Weather station, crop-cutting exercise	Purchasing company	Weather station	Ibid.	Weather station	Weather station	Purchasing company
Insurer	NGO	NGO	NGO	Ibid.	Insurance company	Insurance company	Insurance company

Source: Authors' elaboration.

^a The insurance product also includes an indemnity component covering against other risks such as hail, frost, fire, windstorm, and uncontrollable pests and diseases. Indemnities are released after crop stand checks conducted by field inspectors.

^b While not explicitly tied to a particular crop, each policy was meant to cover revenue from 0.1 acre (0.04 ha) of land cultivated under transplanted aman rice. Households could purchase multiple units of insurance based on the amount of land they cultivate during the monsoon season. According to Hill et al. (2019), this should reduce incentives to view the insurance as a gamble.

^c Indexed to 2015 dollars. Source: World Bank.

^d In some studies, the actuarially fair price varies depending on the community (Karlan et al., 2014) or crop (Bulte et al., 2019). For example, in Bulte et al. (2019), the price per unit insured is 5.2 USD for sunflower, 10.9 USD for soya bean, 11.1 USD for sorghum, and 13.8 USD for maize.

^e Take-up is defined as the share of treated households that subscribe to at least one unit of insurance.

Table 3: Pooling factors from the Bayesian hierarchical model

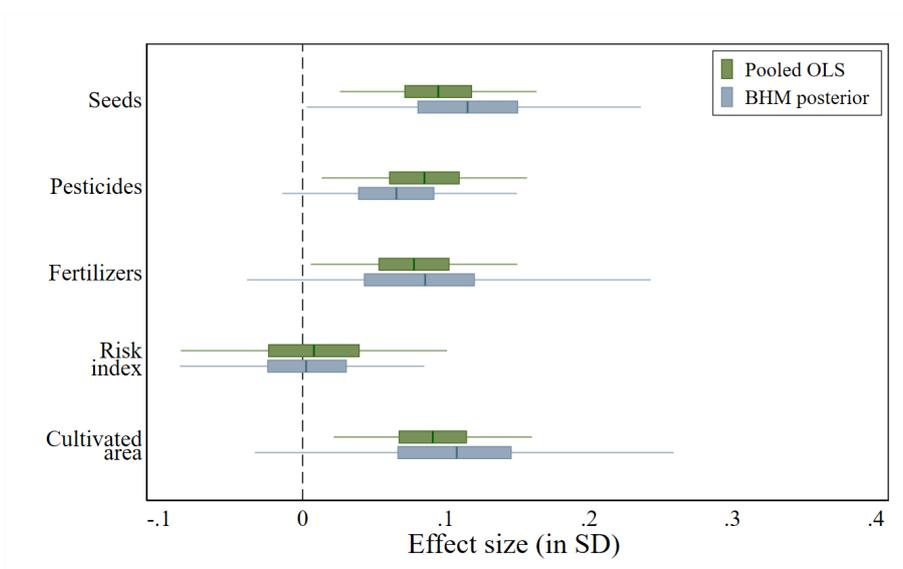
	(1) Fertilizer	(2) Cultivated area	(3) Seeds	(4) Pesticides	(5) Crop risk index
Pooling factors					
Conventional pooling factor	0.38 [0.09, 0.87]	0.38 [0.06, 0.95]	0.40 [0.08, 0.86]	0.69 [0.19, 1.00]	0.76 [0.26, 1.00]
Generalized pooling factor	0.41	0.41	0.44	0.72	0.79
Brute force	0.47	0.45	0.51	0.81	0.90

Notes: Pooling factors are averaged across studies and belong to the interval [0,1], with 0 indicating no pooling and 1 indicating full pooling. 95% posterior intervals are shown in brackets. Computational details are provided in [Appendix B.3](#). Source: Authors' elaboration using data from [Elabed and Carter \(2014\)](#); [Karlan et al. \(2014\)](#); [Cole et al. \(2017\)](#); [Bulte et al. \(2019\)](#); [Hill et al. \(2019\)](#); [Stoeffler et al. \(2022\)](#).

Online Appendices

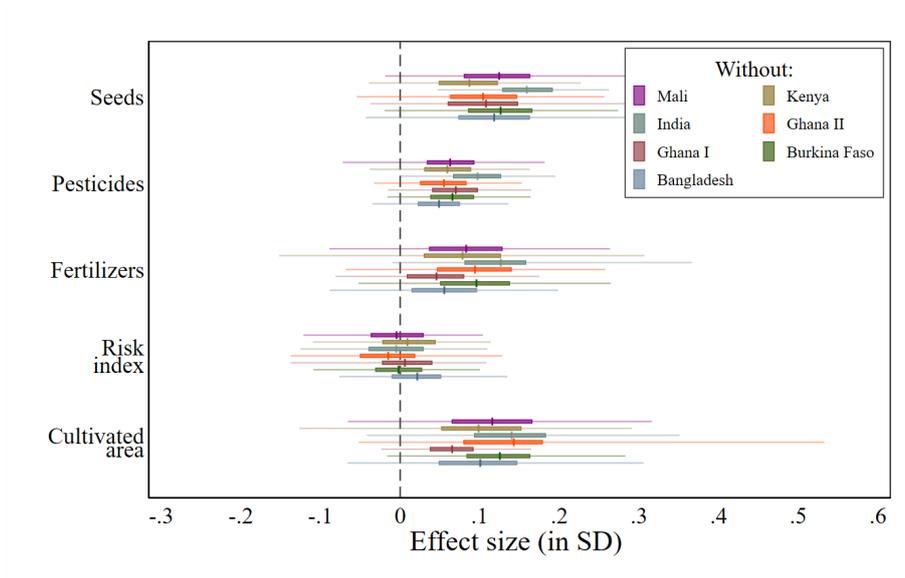
Appendix A Additional tables and figures

Figure A1: Sensitivity to using automatic priors: the average effect of index insurance on production decisions τ



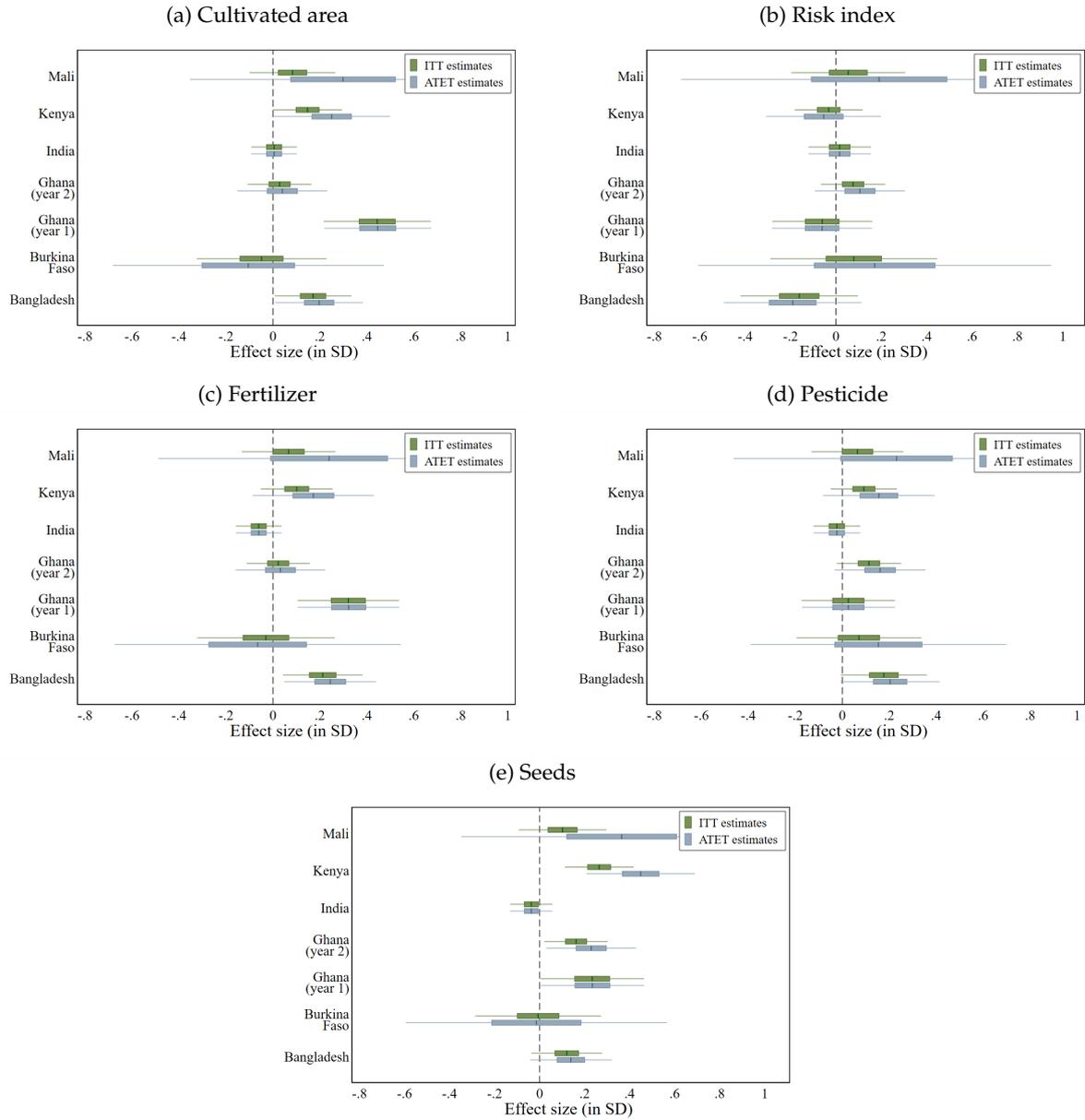
Notes: This figure presents evidence on the average effect of randomly assigned offers, using *baggr*'s automatic priors. See notes to Figure 1 for more details.

Figure A2: Sensitivity to the omission of any single study: the average treatment effect of index insurance on production decisions τ



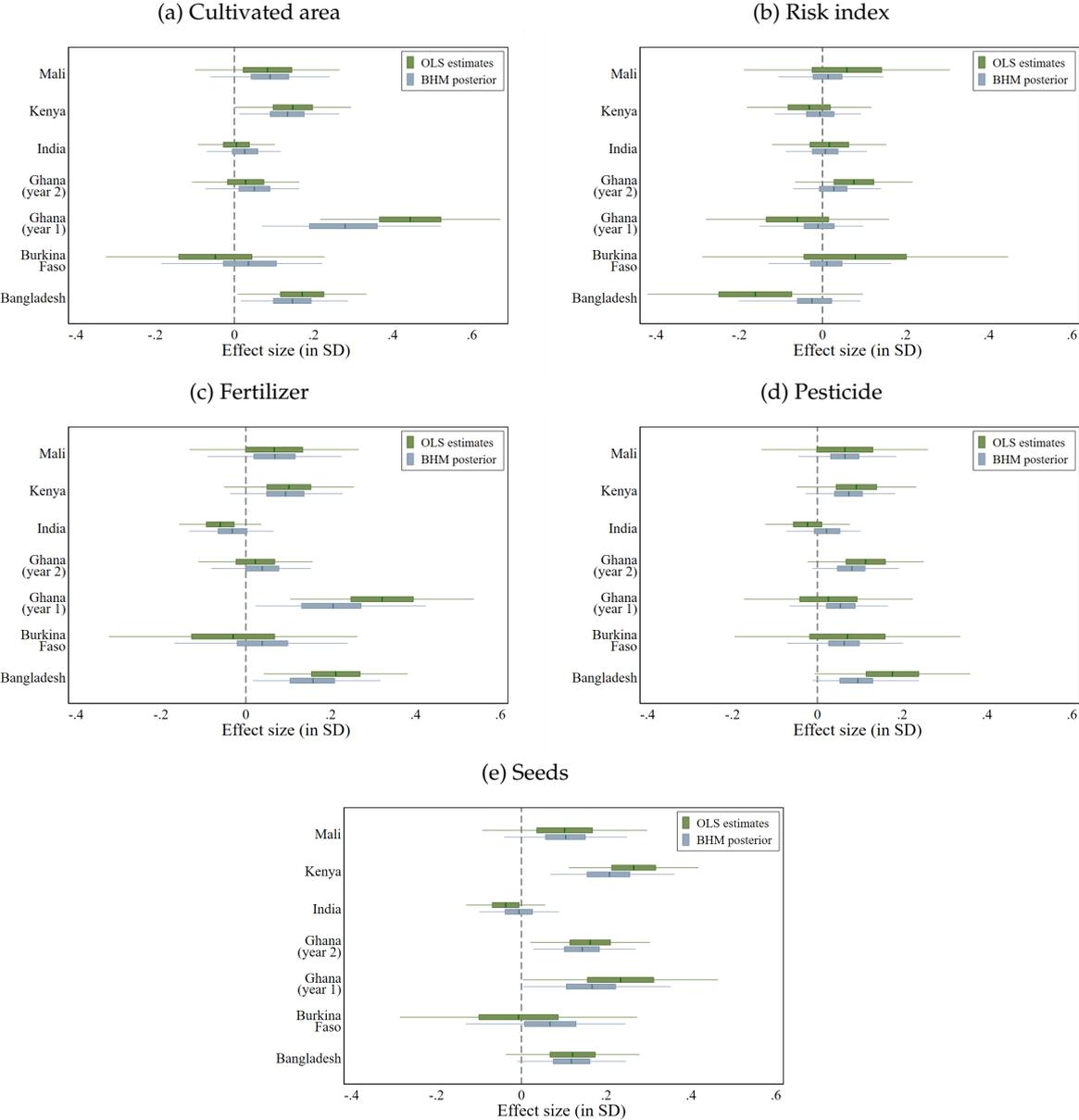
Notes: This figure presents BHM estimates of the average effect of randomly assigned offers, omitting any of the individual studies. See notes to Figure 1 for more details.

Figure A3: Comparing the heterogeneity in ITT and ATET estimates



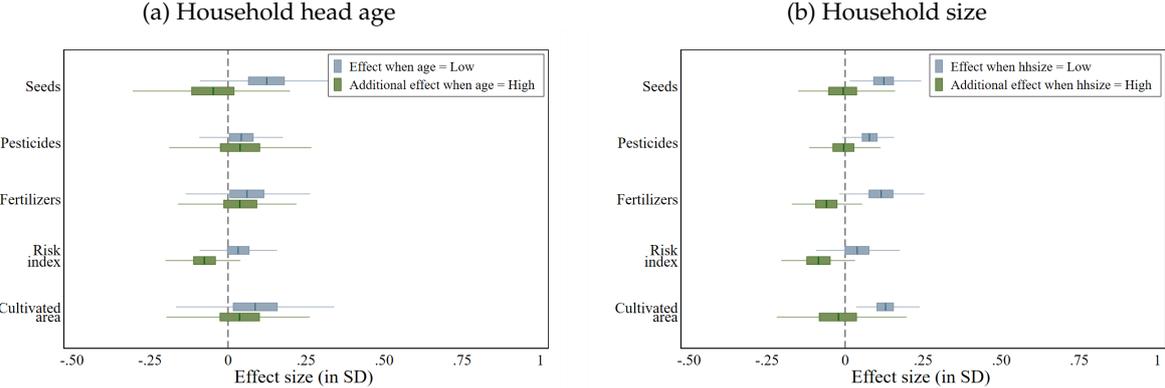
Notes: These figures compare study-specific estimates of the ITT (intent-to-treat effect) and of the ATET (average treatment effect on the treated). ITT estimates are obtained using a no-pooling OLS model (as in Figure 2). ATET estimates are obtained using random assignment to the insurance offer as an instrument for actual take-up. Thin (resp. thick) lines represent the 95% (resp. 50%) confidence intervals. All outcomes have been standardized using data from the original studies. Standard errors are clustered at the level of the randomization (i.e., household or village depending on the study). Source: Authors' elaboration using data from Elabed and Carter (2014); Karlan et al. (2014); Cole et al. (2017); Bulte et al. (2019); Hill et al. (2019); Stoeffler et al. (2022).

Figure A4: Sensitivity to using automatic priors: the heterogeneity of study-specific treatment effects τ_k



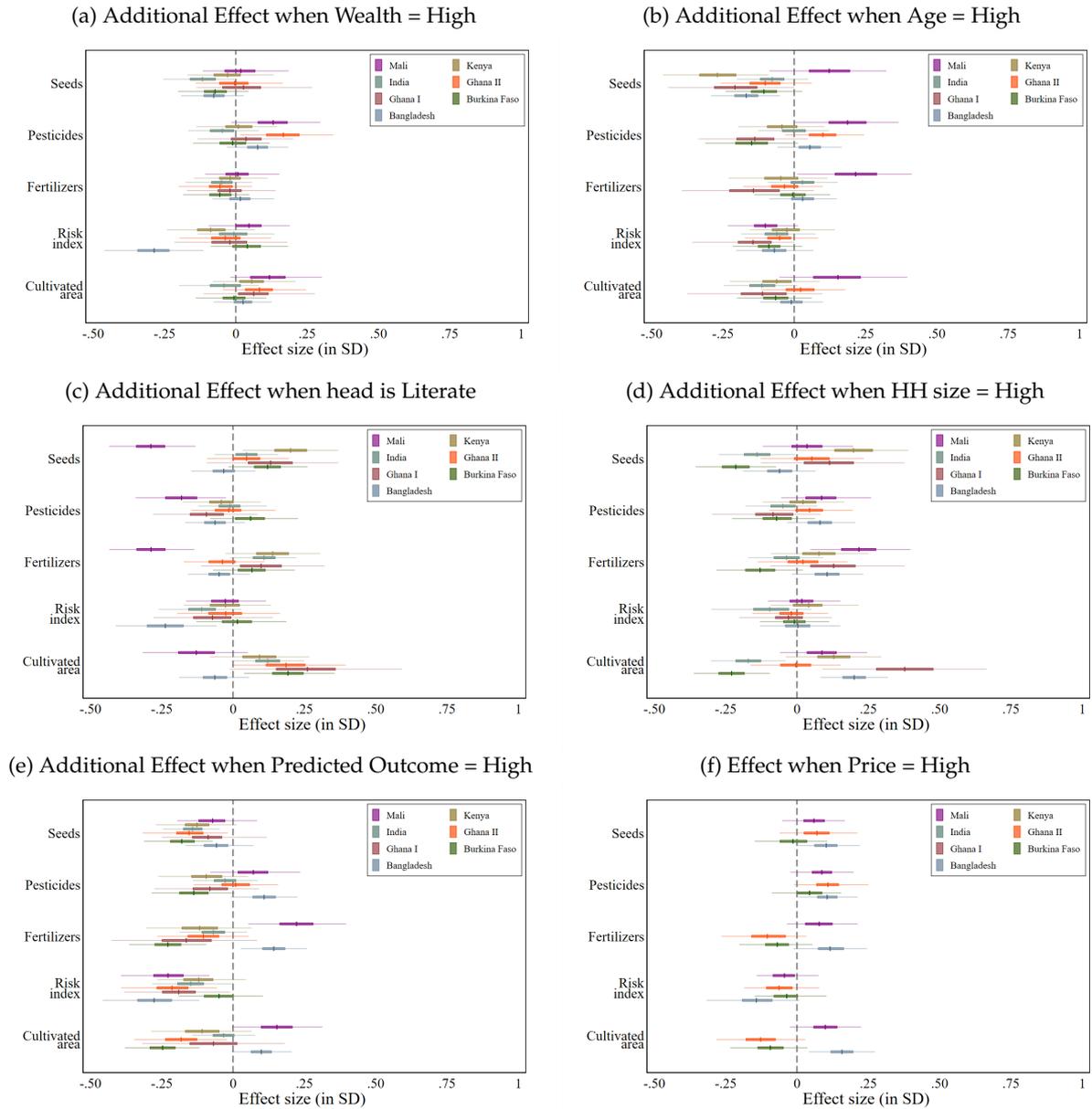
Notes: These figures present the BHM posteriors of study-specific treatment effects τ_k , using *baggr*'s automatic priors. The no-pooling OLS estimates are shown for comparisons. See notes to Figure 2 for more details.

Figure A5: Sensitivity to using no stratification to construct the covariates: the heterogeneity of average treatment effects by key covariates



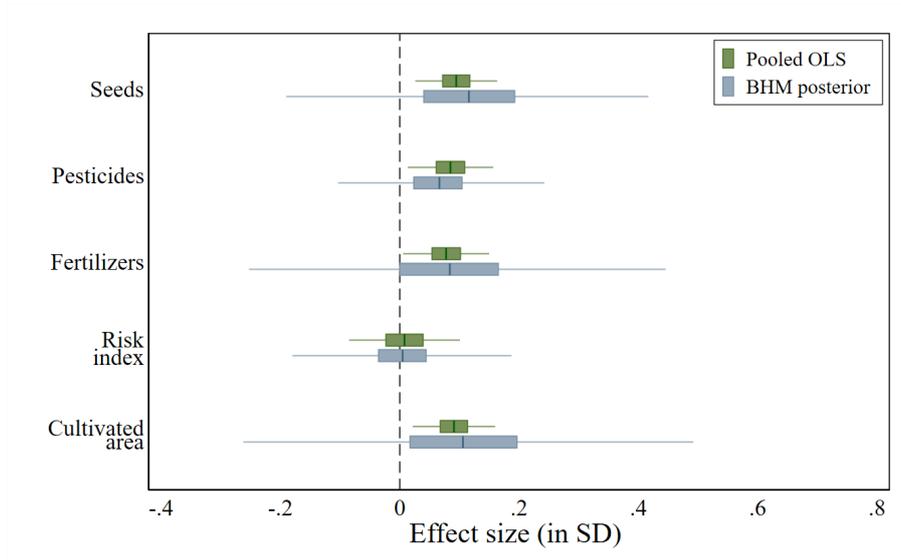
Notes: These figures present the BHM posteriors of the average effect of index insurance offers by household covariates, when no stratification by studies is used to construct the covariate dummies. Results are restricted to covariates with standardized units. See notes to Figure 3 for more details.

Figure A6: Estimates of study-specific treatment effects τ_k for all outcomes split by key covariates



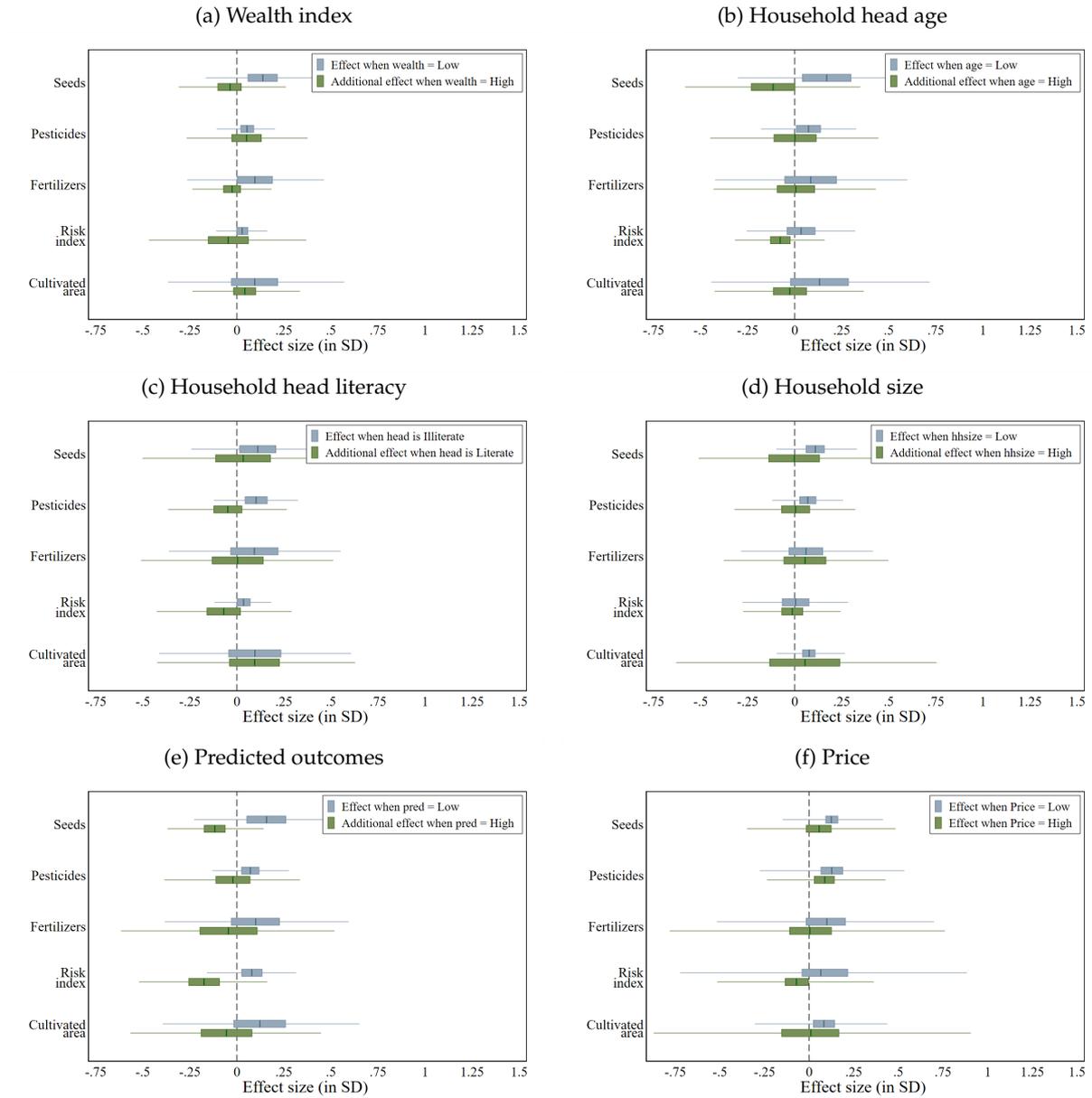
Notes: These figures present the BHM posteriors of study-specific effects of index insurance offers by key household covariates. For each study and outcome, we report the additional effects for the group with the covariate equal to one. See notes to Figure 3 for more details.

Figure A7: Sensitivity to using automatic priors: the predicted effect of index insurance in a new study τ_{k+1} (automatic priors)



Notes: This figure presents evidence on the predicted effect of increasing access to index insurance in a new setting, using *baggr*'s automatic priors. The no-pooling OLS estimates are shown for comparisons. See notes to Figure 4 for more details.

Figure A8: Estimates of τ_{k+1} for all outcomes split by key covariates



Notes: This figure presents evidence on the predicted effect of increasing access to index insurance by key household covariates. Thin (resp. thick) lines represent the 95% (resp. 50%) confidence/posterior intervals. All outcomes have been standardized using data from the original studies. Standard errors are clustered at the household level. Source: Authors' elaboration using data from Elabed and Carter (2014); Karlan et al. (2014); Cole et al. (2017); Bulte et al. (2019); Hill et al. (2019); Stoeffler et al. (2022).

Table A1: Questions that will be used to derive outcomes

	Bangladesh	Burkina Faso	Ghana		India (Cole et al.)	India (Mobarak and Rosenzweig)	Kenya	Mali
	(1)	(2)	Year 1	Year 2	(5)	(6)	(7)	(8)
Outcomes^a								
Cultivated area	E5	C1, D1.2, F1.3	H6	H8	E5	320, 322, 328	M7.1.1	H2, I3
Fertilizer use	E2C	D1.9-11, F1.9-11	L4-7	L4-7	E1.c-d	413-415, 430-431	M7.5.1	H6, H8, H9, I7, I9
Pesticide use	E2C	D1.12-13, F1.12-13	L4-7	L4-7	E1.e	434-436	M7.5.1	H10-11, I10-11
Seeds use	E17	D1.8, F1.8	K2	K3, K6	E1.a-b	405-408	M7.3	H5, I6
Crop portfolio riskiness ^b	E3	C10-23	H8	H9-10	E5	305	M7.1.1	I2

^a We report the questions numbering from endline survey instruments provided by the authors. Mobarak and Rosenzweig (2013) have three endline survey instruments corresponding to the states of Andhra Pradesh, Tamil Nadu, and Uttar Pradesh. The codes for the outcomes are the same across survey and, thus, we write it only once. Although Mobarak and Rosenzweig (2013) and Hill et al. (2019) registered the outcomes for both dry and monsoon seasons, we analyze the outcomes for the monsoon (*Kharif*) season only.

^b The risk index requires information on the type of crop that are cultivated on plots. Thus, codes identified here refer to questions about names of crops cultivated. When codes are redundant with *cultivated area*, it means that the distinction about crop types was already included in this question.

Table A2: Questions that will be used to derive covariates

	Bangladesh	Burkina Faso	Ghana Year 1	Ghana Year 2	India (Cole et al.)	India (Mobarak and Rosenzweig)	Kenya	Mali
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covariates^a								
Age of the household head	A3	A1.4	1.5	B3	A4*	104.1	202	A4
Sex of the household head	A2	A1.1	1.2	B2	A3*	103	201	A2
Literacy of the household head ^b	A6	A1.6	2.c	C11-14	A6-7*	105	yrs_edu ^c	C1.6
Household size	A0	A2	1.1	B1	A1 ^d	101	hhsz	C1.2
Wealth index ^e	H	H1	12.A-B	T	D1-2*	922-926*	M3*	E, E2
Predictors^f								
Land ownership	C8	C1.21	7C	I	D6*	211	301	G1.5
Livestock units	H1-3	H2	7A	P	B*	901-909**	M3.4*	E3
Chemicals usage	E30	D1, F1,	7.I	L	A15	628-635**		
Bank account	L5		11A-C	X4	E1		401	
Attitude towards risk	L	T1, T2*			H			N
Years of schooling of the household head	A7	A2.25	2A-B	C4	A5*	105	yrs_edu	C1.6
Weather shock	K	D1.21, F1.22**	7N	N2.8		1302**		L5-6

^a We report the questions numbering from baseline survey instruments provided by the authors. If not available, we report them from endline survey instruments and differentiate them with single asterisks (*).

^b In Mobarak and Rosenzweig (2013), Elabed and Carter (2014) and Bulte et al. (2019), because there is no question on literacy, we proxy it using household head's years of schooling and divide the sample in two groups (above or below average years of schooling).

^c This variable indicates the educational level of the respondent who is the main farmer in the household but not necessarily the head of the household.

^d Household members below 6 years old are not registered.

^e To compute the wealth index, we make use of the maximum of information available at the level study (durable assets owned by the household, housing and sanitary conditions) and refer to their variable codes here.

^f Some of the predictors are already reported as covariates. If the predictors are not available in baseline questionnaires, we report them from endline survey instruments and differentiate them through single asterisk (*). When there are double asterisks, the variables are found in endline survey instruments but refer to the baseline period through recall questions (**).

Table A3: Descriptive statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Bangladesh	Burkina Faso	Ghana		India	Kenya	Mali
			Year 1	Year 2			
Household size ^a	4.330 (1.388)	10.417 (6.245)	7.039 (3.531)	6.398 (3.897)	5.151 (2.050)	5.655 (1.981)	19.135 (13.738)
Head is male	0.961 (0.194)	0.995 (0.070)	0.948 (0.222)	0.709 (0.454)	0.913 (0.282)	0.913 (0.282)	0.999 (0.032)
Head age	42.723 (11.767)	43.800 (12.928)	44.361 (17.306)	32.097 (23.341)	49.596 (12.404)	46.205 (13.916)	55.387 (14.950)
Head education (in years)	3.506 (3.941)	1.175 (2.465)	.	.	3.748 (4.759)	6.288 (3.700)	0.984 (3.709)
Head literacy ^b	0.490 (0.500)	0.329 (0.470)	0.242 (0.429)	0.256 (0.437)	0.430 (0.495)	0.587 (0.493)	0.415 (0.493)
Has a bank account	0.293 (0.455)	.	0.078 (0.268)	0.073 (0.259)	0.283 (0.450)	0.262 (0.440)	.
Owens land	0.617 (0.486)	.	.	.	1.000 (0.000)	1.000 (0.000)	0.892 (0.311)
Land owned (in acres)	0.362 (0.642)	.	.	.	5.368 (5.471)	3.183 (2.633)	13.342 (12.663)
Livestock ^c	0.852 (0.810)	6.391 (9.304)	1.982 (3.850)	1.978 (5.473)	1.542 (2.641)	3.622 (3.733)	1.8e+06 (4.1e+06)
Weather shock	0.155 (0.362)	.	0.506 (0.501)	0.577 (0.494)	.	.	0.784 (0.412)
Cultivated area (in acres) ^d	0.707 (0.592)	20.703 (15.614)	9.123 (6.244)	9.140 (9.806)	3.996 (3.592)	2.880 (3.114)	16.384 (10.159)
Used fertilizers	0.967 (0.180)	0.962 (0.190)	0.691 (0.463)	0.684 (0.465)	0.931 (0.253)	0.851 (0.356)	1.000 (0.000)
Used pesticides	0.760 (0.427)	0.992 (0.089)	0.395 (0.489)	0.470 (0.499)	0.637 (0.481)	0.841 (0.366)	0.994 (0.078)
Purchased seeds	0.585 (0.493)	.	0.496 (0.501)	0.447 (0.497)	0.974 (0.158)	0.494 (0.500)	.
N	1974	1010	385	1406	1479	780	971

Source: Authors' elaboration using data from Elabed and Carter (2014); Karlan et al. (2014); Cole et al. (2017); Bulte et al. (2019); Hill et al. (2019); Stoeffler et al. (2022).

^a In Cole et al. (2017), household members with age below or equal to 5 are not enumerated. We proxy total household size using a scaling factor defined as the total household size divided by the number of household members with age above 5 (we use data from the DHS surveys and restrict the sample to rural households).

^b In Elabed and Carter (2014) and Bulte et al. (2019), because there is no question on literacy, we proxy it using household head's years of schooling and divide the sample in two groups (above or below average years of schooling).

^c This variable captures total livestock population in tropical units, with the exception of Elabed and Carter (2014) in which it represents the value of livestock owned by the household.

^d We harmonize the unit of cultivated areas across studies by converting all values into acres (1 hectare = 2.4711 acres).

Appendix B Technical appendix

Appendix B.1 The Bayesian Hierarchical model with and without household covariates

Consider some outcome of interest y_{ik} for an individual $i = 1, 2, \dots, N_k$ in study $k = 1, 2, \dots, K$. Let Y_k represent the N_k -length vector of observed outcomes from group k . Denote the binary indicator of treatment status by T_{ik} , and denote by T_k the N_k -length vector of all treatment status indicators from group k .

Suppose that y_{ik} varies randomly around its mean $\mu_k + \tau_k T_i$. τ_k is the treatment effect in group k . The random variation in y_{ik} may be the result of sampling variation or measurement error or it may be the result of unmodelled heterogeneity or uncertainty in outcomes for individuals within the group. We allow the variance of the outcome variable y_{ik} to vary across sites, so σ_{yk}^2 may differ across k .

The evidence aggregation model from Rubin (1981) consists of a hierarchical likelihood as follows:

$$\begin{aligned}\hat{\tau}_k &\sim N(\tau_k, \hat{s}e_k^2) \\ \tau_k &\sim N(\tau, \sigma_\tau^2)\end{aligned}\tag{4}$$

When we add covariates at the individual level like in Meager (2019), the model becomes:

$$\begin{aligned}y_{ik} &\sim N\left(\sum_{p=1}^{2^L} [\mu_k^p + \tau_k^p T_{ik}] X_{ik}^{\pi(p)}, \sigma_{yk}^2\right) \\ \tau_k^p &\sim N(\tau^p, \sigma_{\tau^p}^2)\end{aligned}\tag{5}$$

where X_{ik} is the covariate for household i in study k and L is the number of covariates included in the analysis. Overall, this results into 2^L intercept terms and 2^L slope terms indexed by p . Because X_{ik} are dummy variables, we obtain the bijection of the full set of interactions of these

variables $\pi(p) : \{1, 2, \dots, 2^L\} \rightarrow \{0, 1\}^L$. For $p \in \{0, 1\}^L$, we consider $X_{ik}^p = \prod_{p=1}^L [X_{ik}^p]^{\mathbb{1}\{I_p=1\}}$.

Appendix B.2 Bayesian estimation

The Bayesian estimation method makes draws from the joint posterior distribution $p(\{\tau_k\}_{k=1}^K, \tau, \sigma_\tau^2 | y)$.

We know from Baye's rule that $p(\{\tau_k\}_{k=1}^K, \tau, \sigma_\tau^2 | y)$ is equal to $p(\{\tau_k\}_{k=1}^K | \tau, \sigma_\tau^2, y) p(\tau | \sigma_\tau^2, y) p(\sigma_\tau^2 | y)$.

During the process of estimation, the hyperparameters σ_τ^2 and τ are successively drawn from their marginal posterior distributions and used to draw $\{\tau_k\}_{k=1}^K$ from their posterior distribution.⁴⁰

Given the equations for the posteriors, estimating the distributions of parameters relies on simulations. The Bayesian computation relies on the use of Markov chain Monte Carlo. For instance, consider s steps in the process of simulation. We first stimulate $\sigma_\tau^{2(s)}$ thanks to its distribution and compute $p(\sigma_\tau^2 | y)$. $\sigma_\tau^{2(s)}$ then constitutes the input to calculate $p(\tau | \sigma_\tau^2, y)$ and get $\tau^{(s)}$ from its normal distribution. Eventually, $\tau^{(s)}$ helps to sample $p(\{\tau_k\}_{k=1}^K | \tau, \sigma_\tau^2, y)$ leading to independent $\tau_k^{(s)}$. It clearly appears that inferences on $\{\tau_k\}_{k=1}^K$ stem from inferences on (τ, σ_τ^2) and vice versa, propagating the underlying uncertainty at each step of the model (Betancourt and Girolami, 2013).

Appendix B.3 Pooling metrics

The pooling factor The following equation, proposed first by Box and Tiao (1973), characterizes the main study-level pooling statistic:

$$\lambda_k^1 = \frac{\hat{s}e_k^2}{\hat{s}e_k^2 + \hat{\sigma}_\tau^2} \quad (6)$$

For each study k , this pooling metric shows the decomposition between the true heterogeneity across treatments captured by $\hat{\sigma}_\tau^2$ and sampling variation estimated by $\hat{s}e_k^2$. Intuitively,

⁴⁰Gelman et al., 2013 describes the marginal posterior of the hyperparameters as $p(\tau, \sigma_\tau^2 | y) \propto p(\tau, \sigma_\tau^2) \prod_{k=1}^K N(\hat{\tau}_k | \tau, \sigma_\tau^2 + \hat{s}e_k^2)$. This can be simplified through integration over τ leading to $p(\tau, \sigma_\tau^2 | y) = p(\tau | \sigma_\tau^2, y) p(\sigma_\tau^2 | y)$ (see e.g. Bandiera et al., 2021).

when λ_k^1 becomes smaller, there is little chance that results from one context can inform us about the expected impacts in a new setting. $\lambda_k^1 > 0.5$ indicates a domination of sampling variation over true heterogeneity.

The brute force The use of sampling variation in the first metric makes it vulnerable to the fluctuation of sample size across studies. Thus, we complete the analysis with an alternative pooling factor at the study-level, which directly measures the extent to which the posterior mean of the treatment effect in study k , denoted $\tilde{\tau}_k$, is determined by the posterior mean of the treatment effect $\tilde{\tau}$ versus the estimate $\hat{\tau}_k$. Meager (2019) calls it "the brute force" and defines it as follows:

$$\lambda_k^2 = \frac{\tilde{\tau}_k - \hat{\tau}_k}{\tilde{\tau} - \hat{\tau}_k} \quad (7)$$

Both the pooling factor and the brute force are eventually averaged across studies and denoted $\lambda^1(\tau)$ and $\lambda^2(\tau)$.

The generalized pooling factor Gelman and Pardoe (2006) develops further the analysis to compute a "generalized pooling factor". Let E be the posterior mean and $\epsilon_k = \tau_k - \tau$. The generalized pooling factor follows:

$$\lambda = 1 - \frac{\frac{1}{K-1} \sum_{k=1}^K (E[\epsilon_k] - E[\bar{\epsilon}])^2}{E[\frac{1}{K-1} \sum_{k=1}^K (\epsilon_k - \bar{\epsilon})^2]} \quad (8)$$

Gelman and Pardoe (2006) also considers $\lambda > 0.5$ to be the threshold above which there is a higher degree of information at the "population level" rather than at the "site level". At the extreme value $\lambda = 0$, pooling data is not relevant since the broader population contains no information to the true effect in a specific context.

Appendix C E-mail sent to the authors

Object: Data query for a research project on Index Insurance

Dear [NAMES OF AUTHORS],

We are Pauline Castaing and Jules Gazeaud, two PhD candidates at CERDI (France) both supervised by Catherine Araujo-Bonjean.

We recently came up with the attached research project that partly depends on data that you collected in [COUNTRY] for your paper [TITLE]. We would like to know whether you could provide us with an access to this data and if yes under which conditions. We invite you to look at the attached document to better capture the essence of our project.

For the moment, we only need to know whether you would be willing to provide us an access to your survey instrument and data. Then, we will write and register a detailed pre-analysis plan on the Open Science Framework platform (see <https://osf.io/>) depending on your answer and potential feedbacks. It is only then that we will make formal data requests to the authors of selected studies.

Our main objective is to quantify the average impact of index insurance on farmers' production decisions, aggregating the evidence found by existing randomized controlled trials. In addition, we seek to understand how much of the variation in index insurance treatment effect directly stems from the sampling variation to eventually conclude on the external validity of such programs. In other words, our idea is to leverage the evidence across index insurance studies to better predict the expected impacts of index insurance in other contexts.

We believe that the attached document clarifies our intentions and make them as transparent as possible. However, please do not hesitate to let us know if any of this is unclear or if you need any further information.

We are aware that sharing such data is sensitive. Therefore, in case you decide to support our research project and make your data available, we will commit to the following:

1. We will not share the data with anyone.
2. We will not re-evaluate your results, nor overstate them. As mentioned, our intention is to aggregate the evidence found by RCTs testing the effect of index insurance on production decisions to derive the average effect of such intervention across different contexts (and how much stems from sampling variation).
3. The use of your data will serve the unique purpose of the project attached. No additional exploration will be made without your consent.
4. Naturally, you will be thanked for collecting and sharing the data underlying our research, and we will acknowledge that errors and viewpoints are our own.

Of course, we would be very happy to chat more about these or any additional conditions that you may need to share your data. Also, please let us know if you have comments or feedbacks on our research project. We are very open to discussions!

Many thanks in advance for your consideration.

Best regards,

Pauline and Jules

References

- Bandiera, O., Fischer, G., Prat, A., and Ytsma, E. (2021). Do women respond less to performance pay? Building evidence from multiple experiments. *American Economic Review: Insights*, 3(4):435–54.
- Betancourt, M. J. and Girolami, M. (2013). Hamiltonian Monte Carlo for hierarchical models. *arXiv:1312.0906 [stat]*.
- Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis*. Wiley Classics.
- Bulte, E., Cecchi, F., Lensink, R., Marr, A., and Van Asseldonk, M. (2019). Do crop insurance-certified seed bundles crowd-in investments? Experimental evidence from Kenya. *Journal of Economic Behavior and Organization*.
- Cole, S., Giné, X., and Vickery, J. (2017). How does risk management influence production decisions? Evidence from a field experiment. *The Review of Financial Studies*, 30(6):1935–1970.
- Elabed, G. and Carter, M. (2014). Ex-ante impacts of agricultural insurance: Evidence from a field experiment in Mali. Mimeo.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Third Edition. Chapman and Hall/CRC.
- Gelman, A. and Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, 48(2):241–251.
- Hill, R. V., Kumar, N., Magnan, N., Makhija, S., de Nicola, F., Spielman, D. J., and Ward, P. S. (2019). Ex ante and ex post effects of hybrid index insurance in Bangladesh. *Journal of Development Economics*, 136:1–17.
- Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, 129(2):597–652.
- Meager, R. (2019). Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, 11(1):57–91.
- Mobarak, A. M. and Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review: Papers & Proceedings*, 103(3):375–80.
- Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6(4):377–401.
- Stoeffler, Q., Carter, M., Guirkinge, C., and Gelade, W. (2022). The spillover impact of index insurance on agricultural investment by cotton farmers in Burkina Faso. *The World Bank Economic Review*, 36(1):114–140.