DEMAND AND DESIGN CONSIDERATIONS FOR SMALLHOLDER FARMERS' WEATHER INDEX INSURANCE PRODUCTS

Dissertation
to obtain the Ph. D. degree
in the International Ph. D. Program for Agricultural Sciences in Goettingen (IPAG)
at the Faculty of Agricultural Sciences,
Georg-August-University Göttingen, Germany

presented by

Francisco Ceballos

born in Buenos Aires, Argentina

D7

1. Name of supervisor: **Prof. Dr. Matin Qaim**

2. Name of co-supervisor: **Dr. Maximo Torero**

Date of dissertation: **22 November 2017**

Summary

This cumulative dissertation discusses a range of topics related to the demand for and considerations around product design of weather index insurance instruments aimed at smallholder farmers' risk management.

The second chapter, "Demand for a Simple Weather Insurance Product in India: Theory and Evidence," uses survey and administrative data from a project conducted in Madhya Pradesh, India during the 2010-2012 period. As part of this project, a new index insurance product was introduced to protect smallholder farmers of rainfed soybean from both deficit and excess rainfall during two consecutive Kharif (summer) seasons. In order to assess the drivers behind the demand for insurance the study induced exogenous variation along three dimensions: (1) Spatial basis risk / Distance to the weather station (by installing three new randomly-positioned weather stations), (2) Insurance premium (by offering random discounts), and (3) Product understanding (by randomly varying the intensity of training across villages).

The chapter relies on a standard expected utility theory framework by Clarke (2016) to develop a series of hypotheses about the responsiveness of demand to price, spatial basis risk, and farmer's risk aversion. Demand is found to behave as predicted: falling with price and basis risk and hump-shaped in risk aversion. Moreover, there is evidence of differential price sensitivity at different levels of basis risk, as predicted by the model. With respect to product understanding, the evidence suggests that increased incentives to learn or learning by using are more effective at increasing both understanding and demand.

Furthermore, the chapter contributes to the scarce evidence on the dynamics of the demand for insurance by analyzing a two-year panel of insurance purchases. While the effect of premium subsidies persists over time, that of investments in new weather stations diminishes and the effect of increased training in the first season seems to disappear altogether during the second season. Importantly, while having previously purchased insurance does not encourage future uptake, receiving a payout does, potentially reflecting issues of trust in the product or the insurance company.

The third chapter, "Estimating Spatial Basis Risk in Rainfall Index Insurance: Methodology and Application to Excess Rainfall Insurance in Uruguay," tackles the

important topic of basis risk in weather index insurance in more depth. In particular, the chapter sets out to estimate the actual extent of spatial or geographic basis risk and compare this to farmers‘ perceptions, as captured from survey data.

A novel methodology is developed to estimate the degree of spatial basis risk for an arbitrary rainfall index insurance instrument. The methodology relies on a widely-used stochastic rainfall generator by Wilks (1998), extended to accommodate non-traditional dependence patterns through a copula function. In particular, the model intends to capture spatial upper tail dependence in rainfall, or the tendency for extreme rainfall (as that related to extensive, large-area storms) to be more spatially correlated than milder rainfall. This feature is empirically shown to occurr in available historical rainfall data. The extent of basis risk is then captured by simulating from the calibrated model and calculating the fraction of cases in which the insurance product would not pay even when rainfall at the farmer’s plot is within the payout region.

The methodology is applied to an index product insuring against excess rainfall in Uruguay. To calibrate the model for this case study, the chapter uses historical daily rainfall data from the national network of weather stations, complemented with a unique, high-resolution dataset from a dense network of 34 automatic weather stations around the study area. The degree of downside spatial basis risk is then estimated by Monte Carlo simulations and the results are linked to both a theoretical model for the demand of index insurance and to farmer perceptions about the product.

The results indicate that basis risk is not negligible in our case study. Depending on the farmer’s location, basis risk is such that the insurance product would fail to pay between 1 to 5 times out of 10 in which a farmer faces critical crop losses. Moreover, while spatial basis risk naturally increases with distance to the insurance reference gauge, it does so at a decreasing rate. In turn, farmers seem to overestimate the rate of increase, pointing to the presence of information asymmetries regarding the spatial properties of rainfall.

In terms of the comparison to the theoretical model by Clarke (2016), spatial basis risk generally remains within the theoretical range in which a risk-averse farmer would demand a positive amount of insurance, even for plots located at a considerable distance from the reference weather station at which the index is measured. Finally, the results point to the importance of taking into consideration geographic variation in precipitation patterns—even within relatively small regions—when designing an index insurance product. This element is shown to considerably increase (or decrease) the degree of spatial basis risk, depending on the exact location of a farmer’s plot and its insurance reference weather station. This calls for a much more careful consideration of local climatologies before launching an index insurance product based on nearby weather stations.

The fourth chapter, “Demand Heterogeneity for Index-Based Insurance: The Case for Flexible Products,” discusses a generally-overlooked yet important design issue in weather index insurance. Notably, most existing index insurance products are characterized by a relatively rigid payout structure, intended for a representative farmer's standard risk profile. Albeit convenient, this one-size-fits-all structure comes at the cost of ignoring considerable heterogeneity in agricultural risk profiles, potentially lowering the product's worth for many farmers.

The chapter provides unique evidence on the ways in which heterogeneity in farmers' risk exposure affects their demand for agricultural index insurance. To achieve this, it analyzes a set of flexible insurance products against excess rainfall recently marketed in Uruguay during the 2013-2014 period to cover horticultural losses around harvest. The products were designed as independent *insurance units*—separately covering against the risk of excess rainfall across different calendar months and at different rainfall intensities—and were intended to be freely combined by the farmers to form optimal insurance portfolios that suited their particular risk profiles. The analysis exploits the substantial variation observed in the insurance portfolios demanded by farmers.

The relevance of alternative sources of heterogeneity is explored by extending a simple expected utility decision model and relying on structural estimation to test the significance of each of these sources. The results show evidence for the presence of important aspects of farmer heterogeneity that directly affect their demand for risk-coping instruments, including the particular mix of crops chosen by the farmer, ex-post planting dates, soil drainage, product understanding, and spatial basis risk. The chapter concludes by quantifying the benefits of a flexible scheme by comparing farmer welfare to that achieved under alternative counterfactual insurance options. Overall, the value of providing flexibility in the form of an insurance units scheme is substantial. The findings underscore the need to provide flexibility when implementing index-based tools for hedging agricultural risks.

To Mila and Felix, with all their light

Acknowledgments

First of all, I am deeply grateful to Maximo Torero for his initial encouragement to pursue this path. His overall trust and subsequent support along the way have made this dissertation possible.

My supervisor, Matin Qaim, kindly allowed me the opportunity to embark on these studies under his guidance and deposited trust in me.

My thanks also go to Prof. Oliver Musshoff (whose course inspired me for one of the present chapters) for his time and effort in serving as third supervisor and for providing thorough comments on a preliminary draft of Chapter 3.

This dissertation would not have been possible absent the teachings, guidance, trust, and kindness from many people along my professional path: Federico Sturzenegger, Germán Fermo, Sergio Schmukler, Miguel Robles, Manolo Hernandez. My appreciation goes to all of them. Special thanks go to Miguel, who through countless conversations helped me mature a lot of ideas related to insurance and other topics.

This dissertation would have also not been possible absent a great deal of luck, coincidence, and overall mystery in the workings of the world.

My family, knowingly and unknowingly, provided me with the support and time to carry out this work, patiently dealing with the inevitable emotional upheavals that accompanied it.

I thank my dog, Eko, for our daily walks through the forest, which gave me invaluable time and mental space to gradually advance the ideas materialized in this dissertation.

Finally, infinite thanks go to my mom, Ana Inés, who provided me with the personal values, educational foundations, and sense of humour that allowed me to grow well after she left our side.

Table of Contents

Summary ... iii
Acknowledgments ... vii
Table of Contents ... ix
List of Tables ... xii
List of Figures ... xiii
1 General Introduction ... 1
1.1 The pernicious effects of weather risk ... 1
1.2 Formal insurance market failure ... 3
1.3 The promise of index insurance ... 3
1.4 Problem statement ... 4
1.5 Objectives of the study ... 5
1.6 Dissertation outline ... 6
2 Demand for a Simple Weather Insurance Product in India – Theory and Evidence ... 9
2.1 Theoretical Framework and Predictions on the Demand for Index Insurance ... 12
2.1.1 Maximization Problem ... 13
2.1.2 Basis Risk ... 14
2.1.3 Price ... 15
2.1.4 Demand Price Elasticity and Basis Risk ... 16
2.1.5 Risk Aversion ... 16
2.2 Empirical Design ... 17
2.2.1 Basis Risk ... 18
2.2.2 Price ... 19
2.2.3 Understanding ... 20

2.3 Data ... 21
2.4 Analysis ... 24
2.4.1 The Initial Impact of Marketing Interventions ... 25
2.4.2 Impact on Insurance Knowledge and Attitudes ... 28
2.4.3 The Longer-Run Impact of Marketing Interventions ... 30
2.4.4 Is Demand Hump-Shaped in Risk-Aversion? ... 32
2.5 Conclusions ... 36
3 Estimating Spatial Basis Risk in Rainfall Index Insurance: Methodology and Application to Excess Rainfall Insurance in Uruguay ... 39
3.1 Precipitation Model and Extension through Copulas ... 44
3.1.1 Chain-dependent stochastic precipitation model ... 45
3.1.2 Extension ... 46
3.1.3 Copulas ... 46
3.2 Basis Risk: Definition, Measure, and Theoretical Upper Bound ... 48
3.3 Study Context and Data ... 50
3.3.1 Survey and Farmers' Perceptions ... 50
3.3.2 Data ... 53
3.4 Estimation ... 53
3.4.1 Rainfall occurrence process ... 54
3.4.2 Rainfall amount process ... 56
3.4.3 Spatial Interpolation of Parameters ... 59
3.5 Simulations and Results ... 64
3.6 Conclusions ... 71
4 Demand Heterogeneity for Index-Based Insurance: The Case for Flexible Products ... 73
4.1 Context and Data ... 79
4.2 Model ... 83
4.2.1 Farmer's Utility ... 83
4.2.2 Insurance Units ... 84
4.2.3 Decision Problem ... 84
4.2.4 Calibration ... 85
4.3 Sources of Heterogeneity ... 87
4.3.1 Crop Composition ... 87
4.3.2 Planting Dates ... 87

4.3.3 Soil Type 88
4.3.4 Product Understanding 88
4.3.5 Distance to the Weather Station 89
4.4 Structural Estimation 90
4.5 Results 91
4.5.1 Model Estimates 91
4.5.2 The Value of Flexibility 93
4.6 Discussion and Policy Implications 96
5 General Conclusions 99
5.1 Key findings 100
5.2 Policy implications 101
5.3 Final comments 104
Appendix Chapter 2 107
Appendix Chapter 3 111
References 113

List of Tables

Table 2.1 — Joint Probability of Income Loss and Index Payout......12
Table 2.2 — Weather Station Assignment......19
Table 2.3 — Tests of Balance between Villages Assigned to New and Old Weather Stations 20
Table 2.4 — Tests of Balance between Villages Offered Intensive and Basic Insurance Literacy Training......23
Table 2.5 — Summaries of Insurance Purchases by District and Sample......24
Table 2.6 — Take-Up Among Sampled Households, 2011......26
Table 2.7 — Log of Units Bought, 2011......26
Table 2.8 — Impact of Offering Insurance on Insurance Knowledge and Attitudes......28
Table 2.9 — Impact of Training, Discounts, and Weather Stations on Insurance Knowledge and Attitudes......29
Table 2.10 — Take-Up Among Sampled Households, 2012......30
Table 2.11 — Take-Up Among Households, 2012, Including Price in 2011......31
Table 2.12 — 2012 Uptake Among Sampled Households, Including 2011 Uptake and Payouts 33
Table 2.13 — Pooled Results......34
Table 2.14 — Testing the Relationship between Risk Aversion, Price, and Demand......35
Table 3.1 — Farmers' Perception of Geographic Variability in Rainfall Patterns......52
Table 3.2 — Estimated First-Order Markov Chain Probabilities and Model Selection......56
Table 3.3 — Estimated Gamma Parameters......57
Table 3.4 — Downside Basis Risk and Theoretical Upper Bound......69
Table 4.1 — Insurance Units Triggers and Purchases by Coverage Month and Degree – 2014-15 Season......80
Table 4.2 — Summary Statistics of Farmer-Level Survey Data......82
Table 4.3 — Model Parameter Estimates using Average Planting Dates......92
Table 4.4 — Model Parameter Estimates using Actual Planting Dates......94
Table 4.5 — Results from Alternative Policy Experiments......96

List of Figures

Figure 2.1 — Income with and without insurance 13
Figure 2.2 — Optimal insurance under basis risk 15
Figure 2.3 — Timeline of activities 21
Figure 2.4 — Price sensitivity of demand as distance to the weather station increases, 2011 ... 27
Figure 2.5 — Probability of purchase against coefficients of relative risk aversion, ... 34
Figure 2.6 — Probability of purchase against coefficients of relative risk aversion 35
Figure 3.1 — Perception of similarity in rainfall and excess rainfall patterns 52
Figure 3.2 — Location of Rainfall Gauges around Canelones, Uruguay 54
Figure 3.3 — Dependence Structure of Precipitation Occurrence Process 56
Figure 3.4 — Overall Correlation and Tail Dependence of Rainfall at Sample Sites 60
Figure 3.5 — Kendall $\boldsymbol{\tau}$ *Correlation and Upper Tail Dependence* $\boldsymbol{\lambda U}$ 61
Figure 3.6 — Dependence Structure of Precipitation Amount Process 62
Figure 3.7 — Interpolated Gamma Distribution Parameters 63
Figure 3.8 — Interpolated Markov-Chain Probabilities for Rain Occurrence 64
Figure 3.9 — Downside Basis Risk - 85th Percentile Product 66
Figure 3.10 — Basis Risk - 95th Percentile Product 67
Figure 3.11 — Downside Basis Risk and Direction to Reference Weather Station 70
Figure 4.1 — Insurance Coverage under Composite Product Scheme 75
Figure 4.2 — Insurance Coverage under Insurance Units Scheme 77
Figure 4.3 — Farmer-Level Portfolios of Insurance Units 81

CHAPTER 1

General Introduction

Climate variability has raised to the forefront of the public agenda during the past decade. While the scientific consensus is not set on whether the frequency or intensity of extreme weather events has indeed increased, the perception of the general public is progressively shifting towards an overall fear of future climate unpredictability.

Looking ahead, the most conservative scientific predicitions are currently forecasting global temperature increases of +1º Celsius within the next 3 decades (IPCC, 2014). Increases in the frequency and intensity of extreme weather events is also deemed very likely, for instance in terms of the duration, intensity, and spatial extent of heat waves and warm spells, heavy precipitation, and coastal flooding events, albeit with heterogeneity across regions. In addition to an increase in the frequency and intensity of events, socioeconomic and climate assessment models point to an increase in exposure among both urban and rural populations, particularly in low-latitude, less-developed areas (IPCC, 2014; World Bank, 2013).

In this context, public and private sectors alike are actively considering scenarios and devising possible actions to cope with and mitigate the foreseen consequences of climate change. In particular, climate adaptation (or the ability of governments and populations to adapt to the new contexts brought about by a changing climate) is deemed crucial. This is particularly the case for the rural agricultural sector, where climate change has the potential to radically shift the landscape of crops suitable for a given area, an effect that may be most disruptive in the case of smallholder farmers with low levels of education and little outside options and coping capacity (Nelson et al., 2013; World Bank, 2013).

1.1 The pernicious effects of weather risk

In agriculture, weather constitutes a major source of risk. In developed and developing countries alike, agricultural production is directly tied to the weather. High and low temperatures, hail, and winds regularly induce damage to agricultural crops with a direct effect on production. Moreover, both excess and scarcity of rainfall constitutes a substantial hazard in rainfed regions without access to irrigation and other coping mechanisms. Weather extremes provide suitable conditions for the flourishing of pests and diseases, which can result in serious consequences on regional crops. Finally, large-scale natural hazards such as floods or hurricanes, can have a direct and lasting impact over livelihoods and infrastructure across broad geographic areas and large groups of people.

While agricultural producers are generally regarded as the most obvious victims of negative weather events—as these induce direct losses in their income—, most actors in the rural economy are affected. Large disruptive effects of extreme weather events on production are subsequently transmitted to other layers of the agricultural value chain (e.g. traders, wholesalers, processors, suppliers) and to rural financial markets (through loan defaults, illiquidity). In such a scenario a general dampening of economic activity may ensue, resulting in a negative feedback loop with further impacts over non-farm incomes. In addition, when production losses are considerable for a specific agricultural commodity, the undersupply can be reflected in price increases affecting most households in the economy. Ultimately, local, regional, and national governments face pressure to respond to extreme weather shocks through agricultural disaster assistance and other social and economic emergency programs (Clarke and Dercon, 2016).

Crucially, weather risk does not equally affect all actors in the rural economy. *Poor* rural farmers are disproportionately affected by it. Faced with a shock to current income, and in the absence of proper financial instruments to smooth these shocks, the poor regularly resort to costly coping strategies which may turn a temporary shock into a permanent one. For instance, following a temporary income shock very poor households may be forced to choose between either liquidating a fraction of their productive assets (i.e. livestock, machinery, land) —thus trumping their future growth potential— or reducing current consumption —a strategy which implicitly puts the burden on future human capital development, particularly for young children in the household.[1]

In this context, households have traditionally resorted to a number of risk coping mechanisms. Some examples include holding savings (either in cash or in kind), investing into semi-liquid assets (such as machinery or livestock), borrowing from informal sources such as moneylenders, relying on social or family networks, and diversification—both in terms of agricultural activities and in terms of the mix between agricultural and non-agricultural labor. Most of these strategies, however, are costly, and their risk-mitigation potential is limited (Townsend, 1994). For instance, loans or gifts from other households have the potential to protect from idiosyncratic shocks (i.e. unexpected losses that affect a reduced number of households within a locality or social network), but are ill-suited to protect against systemic (or common) shocks, which affect all households in a given region and thus undermine their capacity to support each other. Importantly, a diversification strategy may come at an efficiency cost, impeding individuals from capturing the full range

[1] These dynamics are commonly known as, respectively, nutrition-based (Dercon and Hoddinott, 2004) or asset-based (Barrett and McPeak, 2006) poverty traps.

of benefits from specialization or holding back investment into riskier opportunities with a higher expected income.

1.2 Formal insurance market failure

Formal insurance is a natural solution to some of these problems, and may function in a complementary way to existing informal mechanisms. Since insurance markets can pool risks across a much broader scope of activities and a much larger geographic area, they can manage systemic risks more efficiently.

The most common type of insurance is known as indemnity insurance, whereby compensations rely on identifying specific losses and indemnifying the individual against them. From a farmer's perspective, access to insurance can reduce the costs associated to some of the informal strategies outlined above, for instance by improving access to credit. Most importantly, income protection from potential bad years stemming from insurance may enable poor households to unlock investment opportunities previously shunned away as being too risky.

But while formal insurance markets for certain risks (e.g. life, automotive) are fairly developed across the world, insurance against weather risks is very limited (with the exception of certain developed countries or large subsidized systems in few developing ones, with a high level of public intervention in most cases). In the case of crops, multiple peril crop insurance, which can protect a farmer against any source of risk affecting their crop yields, has been historically unsucessfully commercialized among small farmers. Single peril crop insurance, which covers against a specific factor affecting the crop (e.g. hail or winds), has had better success, though still only at modest scales (Smith and Goodwin, 2010).

There are a number of reasons for which agricultural indemnity insurance has failed to develop successfully. Providing agricultural insurance is difficult in rural areas of developing countries because of the small size of farms and lack of proper rural infrastructure, resulting in costly claim verification processes. Moreover, indemnity insurance is prone to information asymmetry problems such as adverse selection (where only the most risky farmers purchase insurance) and moral hazard (where a farmer exerts sub-optimal amounts of effort when being insured).[2] Both of these derive in an increased cost of providing insurance, further trumping its development.

1.3 The promise of index insurance

As a result of these market failures, a promising innovation has gained traction during the past two decades. Weather *index* insurance is an alternative type of crop insurance more adequately suited to smallholder farmers for managing their agricultural risk (Hazell et al., 2010). This type of insurance makes payouts based on whether a given index —for

[2] Hazell, Pomareda, and Valdes (1986).

instance, a particular measure of a weather variable— is above or below a pre-specified threshold.[3] As way of example, a hypothetical index insurance product against drought would pay when the amount of millimeters of rainfall (as measured by a specific weather station or satellite) is under a certain predefined 'trigger'.

Index-based insurance products have been regarded as having enormous potential to reach smallholder farmers in developing countries. Because payouts are based only on publicly-observed data rather than on private information reported by the individual filing claims, all index products reduce the adverse selection and moral hazard problems that often plague insurance markets.[4] In addition, index insurance products drastically reduce the costs of loss verification. This makes insurance easier and cheaper to distribute, potentially lowering premiums and making it more affordable to poor farmers. An added advantage of this type of instrument is that payouts can be calculated and disbursed quickly and automatically without the need for households to formally file a claim, thus supporting farmers' incomes when most needed.

1.4 Problem statement

The potential of index insurance to offer quick and direct support against shocks at a low cost has attracted donors and governments alike. Many local and international organizations have carried out projects in developing countries with the expectation of private insurers stepping in for scaling these up. In India alone, more than nine million farmers purchase these hedging products to insure their risk (Clarke et al., 2012). In the U.S. a large federal insurance index-based scheme protects farmers against a variety of weather risks.

Nevertheless, most index insurance products in developing countries have suffered from dishearteningly-low take-up levels. Some of the explanations put forth for this lack of demand include issues of trust in the insurance company, lack of understanding of the product, liquidity constraints, and crowding-out of insurance by implicit public guarantees (e.g. governments providing emergency relief in the case of an adverse weather event).[5] While all of these obstacles are also applicable to traditional indemnity insurance, weather index insurance suffers from one unique disadvantage: basis risk.

Basis risk can probably be regarded as the Achilles' heel of index insurance. It arises due to an index's inadequacy to perfectly capture the individual losses of an insured

[3] A different type of index insurance, 'area-yield insurance', does not rely on a weather variable as its index, but instead focuses on whether the average yield over a specified area is above or below a specified threshold.

[4] Adverse selection could in theory still exist if farmers in an area have information not available to the insurance company (such as better seasonal forecasts) that allows them to purchase insurance only during more risky seasons. However, while theoretically possible, this is unlikely.

[5] See, for instance, Cole et al. (2013) and Matul et al. (2013).

farmer. This imperfect relationship can be related to a number of factors. First, the weather index may be imperfectly measured because of natural variation of weather between a measurement station and the farmer's plot (a component of basis risk commonly known as *spatial* basis risk), or because of an imperfect remote estimation in the case of satellites. Second, a weather index cannot capture the full complexity of the effect of weather on a crop, which generally involves the interplay of a number of weather variables (temperature, rainfall, humidity, evapotranspiration, winds, etc.) and other factors such as crop variety, soil quality, and farming practices. This is commonly referred to as *design* basis risk. Third, other non-weather events may impact crop growth, such as pest attacks and diseases, which would not be captured by a weather-index product.

Against this backdrop, a primary objective of the research agenda should be to gain a comprehensive understanding of the main obstacles for insurance uptake, while at the same time ensuring that the instruments offered are indeed welfare-enhancing for the targeted farmers. On the former, it is crucial to evaluate the gap between insurance supply and farmer needs and perceptions. This involves both assessing the external validity of existing findings and digging deeper on the determinants of basis risk and on the perceptions that farmers have about it. On the latter, more emphasis should be put in adding to the scarce evidence on the benefits of being insured, both from an ex-ante (i.e. farmer decisions taken before weather realizations) and an ex-post (i.e. insurance outcomes across states of the world and its effect on shock-coping and consumption smoothing) perspective.

1.5 Objectives of the study

The objective of this dissertation is twofold: (i) to provide novel perspectives on the extent and determinants of demand for weather index insurance instruments among smallholder farmers in developing countries; and (ii) to provide insights on the design of index insurance products that can help improve future products. In particular, the dissertation intends to answer the following questions:

I. What is the sensitivity of insurance demand to price, product understanding, and spatial basis risk? How does insurance demand relate to farmer risk aversion? How do these compare to those found in other studies?
II. What are the dynamics of insurance demand over time? How does future demand depend on past experiences with the product?
III. Is it possible to derive ex-ante measures of basis risk by relying on available historical weather information? Can these measures contribute to the improvement of future products at the design phase?
IV. What is the actual extent of spatial basis risk in a case study of rainfall index insurance? How does it compare to farmer perceptions?
V. Is flexible index insurance a viable alternative to existing one-size-fits-all products? Can farmers adapt the insurance to their particular risk profile? How does this impact their welfare?

The evidence presented comes from two weather index insurance projects implemented in very different contexts: India and Uruguay. In the first case, an index insurance product was introduced in Madhya Pradesh, India during the 2010-2012 period to protect smallholder farmers of rainfed soybean from both deficit and excess rainfall during two consecutive Kharif (summer) seasons. The second project was conducted in Canelones, the main horticultural producing region in Uruguay, during the 2013-2014 period and it aimed to cover horticultural farmers against excess rainfall around harvest. In both contexts, we rely on extensive farmer survey data together with administrative insurance purchase data. In the case of Uruguay, we additionally exploit long-term historical weather information together with unique, 3-year data from a high density network of rainfall gauges.

1.6 Dissertation outline

The dissertation proceeds as follows.

The second chapter, "Demand for a Simple Weather Insurance Product in India: Theory and Evidence," analyzes the demand for the rainfall-based weather insurance product in rural India. It presents a standard expected utility theory framework on the nature of demand for index insurance (Clarke, 2016) and explores its predictions in terms of price, basis risk, and risk aversion using survey and administrative data from India, relying on experimental, random variation across the first two dimensions. The results are contrasted against the evidence presented in previous studies (e.g. Cole et al., 2013; Mobarak and Rosenzweig, 2012; Hill, Hoddinott, and Kumar, 2013; and Clarke and Kalani, 2011; among many others). In addition, it contributes to the literature by presenting evidence on the price sensitivity of the demand for insurance at different levels of basis risk and on the impact of insurance training relative to other mechanisms for increasing understanding. Finally, it presents unique evidence on the demand of insurance over time, an important topic that has only been seldom analyzed (Cai, de Janvry, and Sadoulet, 2013; Karlan et al., 2012; and Cole, Stein, and Tobacman, 2014).

The third chapter, "Estimating Spatial Basis Risk in Rainfall Index Insurance: Methodology and Application to Excess Rainfall Insurance in Uruguay," develops a novel methodology to estimate the degree of spatial basis risk for an arbitrary rainfall index insurance instrument. While basis risk has overall been one of the most important issues discussed in the index insurance literature and debates, a concerted effort to understand and analyze its full extent and characteristics has been lacking in the literature and studies directly tackling the subject are scarce. The methodology relies on a widely-used stochastic rainfall generator (Wilks, 1998), extended to accommodate non-traditional dependence patterns—in particular spatial upper tail dependence in rainfall—through a copula function. By doing so, the methodology contributes to the literature by describing a weather generator model suited to analyzing the spatial properties of *extreme* rainfall in a given region.

The methodology is applied to the index product insuring against excess rainfall in Uruguay. The model is first calibrated using historical daily rainfall data from the national network of weather stations, complemented with a unique, high-resolution dataset from a dense network of 34 automatic weather stations around the study area. The degree of downside spatial basis risk is then estimated by Monte Carlo simulations and the results are linked to both a theoretical model for the demand of index insurance and to farmers' perceptions about the product. This application enriches the broader index insurance debate by presenting the first direct empirical exploration of spatial basis risk, relying on an appropriate operational definition (beyond correlation in rainfall), and pointing to the importance of a directional element, generally disregarded as a relevant factor. Moreover, the study contributes additional evidence on behavioral frictions in the insurance market by indicating a relative gap between the real and perceived extents of spatial basis risk.

The fourth chapter, "Demand Heterogeneity for Index-Based Insurance: The Case for Flexible Products," discusses a generally-overlooked yet important design issue in weather index insurance. Notably, most existing index insurance products are characterized by a relatively rigid payout structure, intended for a representative farmer's standard risk profile. Albeit convenient, this one-size-fits-all structure comes at the cost of ignoring considerable heterogeneity in agricultural risk profiles, potentially lowering the product's worth for many farmers. This chapter relates to Chapter 3 in that the issue described can also be interpreted as a subcomponent of basis risk—known as *design* basis risk—, related to the elements in the design of an index product that contribute to the mismatch between payouts and losses.

The chapter provides unique evidence on the ways in which heterogeneity in farmers' risk exposure affects their demand for agricultural index insurance. To achieve this, it analyzes a set of flexible insurance products against excess rainfall marketed in Uruguay and exploits the substantial variation in insurance portfolios demanded by farmers. The relevance of alternative sources of heterogeneity is explored by extending a simple expected utility decision model and relying on structural estimation to test the significance of each of these sources. To our knowledge, this study is the first one to provide theory-based empirical estimates of real-world farmers' risk management behavior. Finally, the chapter quantifies the benefits of a flexible scheme by conducting policy experiments and comparing farmer welfare under this scheme to that achieved under alternative counterfactual insurance options.

The dissertation concludes by discussing policy implications of the findings and discussing potential avenues for future research.

CHAPTER 2

Demand for a Simple Weather Insurance Product in India – Theory and Evidence*

Income risk is substantial for farmers in developing countries. Formal insurance markets for this risk are poorly developed, and much of this risk remains uninsured with significant consequences for investment in productive activitiesp and the welfare of individuals. A large share of the income risk is agricultural, yet crop insurance markets are difficult in this setting (Hazell, Pomareda, and Valdes 1986) because of the small size of farms and limited formal records, and significant potential for adverse selection and moral hazard. As a result an increasing trend has been to sell weather index insurance products (weather hedges) to smallholder farmers to manage their risk (Hazell et al. 2010). In India alone, more than nine million farmers purchase these hedging products to insure their risk (Clarke et al. 2012).

In this chapter we explore the predictions of a standard expected utility theory framework on the nature of demand for such products, in particular testing whether demand behaves as predicted with respect to price, the basis of the hedge, and risk aversion. We use the model of Clarke (2016) to develop a series of hypotheses that we test using data from a randomized control trial in which price and basis risk was varied for a series of hedging products offered to farmers in Madhya Pradesh in India. We find that demand behaves as predicted, with demand falling with price and basis risk, and appearing hump-shaped (that is, increasing and then decreasing) in risk aversion.

Typically, hedging products are not sold to individuals, given their complexity. The market for hedging products that has recently emerged for smallholder farmers in low-

---

* This chapter is the Author's Original Version (AOV) of an article that was published in the American Journal of Agricultural Economics, Volume 98, Issue 4, 1 July 2016, Pages 1250–1270, https://doi.org/10.1093/ajae/aaw031. The chapter is co-authored with Ruth V. Hill and Miguel Robles. The authors' contributions are as follows. RVH and MR conceptualized and obtained funding for the project, designed the insurance product, carried out fieldwork, and collected the data. FC and RVH conceptualized and conducted the analyses. FC, RVH, and MR interpreted the results and wrote the paper. All authors read and approved the final manuscript.

income countries is quite unusual in that it offers complex risk management products to individuals with limited formal education and, in most cases, no prior experience with insurance products. In this chapter, we also examine how individuals learn about these products over time: We look at the impact of increased investments in training on hedging products as well as evidence for learning by doing among farmers. We find evidence that suggests that learning by doing is more effective than increasing the intensity of training.

During the past decade, index-based insurance has received considerable attention as a promising solution to the problem of imperfect insurance markets for rural households in developing countries. As a result, a number of pilot programs—generally coupled with evaluations—have been conducted throughout the world. Cole et al. (2013) report results on the determinants of demand in a number of randomized control trials conducted in India. They find that demand falls as price increases (with a price elasticity of –0.66 to –0.88), tighter credit constraints, and higher distrust of the insurance provider. Mobarak and Rosenzweig (2012) also find that demand for a weather index insurance product is decreasing in prices and distance to the weather station. This chapter contributes to this literature. We also find that demand for an innovative index insurance product falls as price increases and as distance to the weather station increases. We estimate a negative price elasticity of 0.58 and find that doubling a household's distance to a reference weather station decreases demand by 20 percent.

We place these findings in the context of a theoretical model of demand for hedging products and test two other predictions implied by this framework: that demand is more price-elastic when basis risk is lower and that demand is hump-shaped in risk aversion. We find that farmers located less than 5 kilometers (km) away from the reference weather station are four times as sensitive to prices as farmers located at more than 12 km (for whom the basis is higher).

A number of empirical papers have found a puzzling negative relationship between risk aversion and demand for index insurance (Cole et al. 2013; Hill, Hoddinott, and Kumar 2013; and Clarke and Kalani 2011). However, Clarke (2016) shows that in the presence of basis risk under index-based insurance, when premiums are above actuarially fair, a hump-shaped demand with respect to risk aversion is expected. Moreover, when premiums are actuarially fair or below, a downward-sloped relationship is to be expected. We explicitly test for these predictions and find results consistent with the above: for products with a multiple above 1, the demand increases at low levels of risk aversion but again falls at higher levels; for products with a multiple at or below 1, the intensity of demand is negatively related to risk aversion. Nevertheless—arguably because of limited power—the changes in demand at different levels of risk aversion are not statistically significant.

An additional contribution of our study to the existing literature on the determinants of demand for index insurance is our analysis of understanding and demand for hedging products over time. Our study is among few to analyze the demand for index insurance over time (among them Cai, de Janvry, and Sadoulet 2013; Karlan et al. 2012; and Cole, Stein, and Tobacman 2014, which focus on different aspects of the dynamics of demand for insurance). We find that a higher intensity of insurance literacy training has a weakly significant, short-run effect on demand. Households that received more intense training are 5 percentage points more likely to purchase the insurance product in the season immediately following the training. However, we find that intensive training has no significant effect on general insurance knowledge when tested months later, probably due to the fact that households in both control and treatment villages increased their levels of insurance knowledge similarly by the time of the follow-up survey. As a result, unsurprisingly, we find it does not have an effect on insurance demand in the subsequent season. Instead we find that tests on general knowledge of insurance are still consistent with learning by doing: Farmers who received higher price discounts have a greater level of understanding about the insurance product, despite the overall increase in insurance understanding among all farmers.

We also examine whether insurance purchases in 2011 help to explain insurance purchase decisions in 2012, as would be expected in a model of learning by doing. The evidence shows no stand-alone impact of previous purchases; however, receiving a payout in 2011 has an impact on demand during the following season. Purchasing insurance and receiving a payout is strongly positively correlated with the decision to purchase insurance in the subsequent season. A payout in 2011 increases the probability of purchasing in 2012 by around 7 percentage points.

Given that in reality most—if not all—index insurance schemes suffer from recurrent low demand, it is highly relevant to understand and quantify how demand responds to different factors. The overall take-up for the index insurance product that we consider was low in both seasons, at 6.8 and 4.0 percent during 2011 and 2012, respectively. This is in line with the take-up rates found in several other studies for index insurance in India (Giné, Townsend, and Vickery, 2008, report lower than 5 percent take-up in Andhra Pradesh). As considerable government funds are being used to invest in index insurance markets expecting that will help farmers manage uninsured risk, it is crucial to understand why demand is low. Possible reasons are their high price (multiples on index insurance products range from 1.75 to 3.03, according to Cole et al. 2013, which is very high in comparison to a multiple of about 1.2 for indemnity insurance in more developed markets); low levels of investment in weather station infrastructure or financial literacy; or simply an inherent mismatch between the product's attributes and farmers' risk management needs. The results of this study suggest that the price and basis risk are key drivers of demand and that weather hedges will prove to be a useful tool for farmers only if these two elements are substantially reduced.

The rest of the chapter is structured as follows: The next section discusses theoretical predictions on the nature of demand for index insurance. Section 2.2 discusses the different components of the design in our study. Section 2.3 describes the data used for the subsequent analysis. Section 2.4 presents and discusses the results from the empirical analysis and contrasts them with the theory. The final section concludes.

2.1 Theoretical Framework and Predictions on the Demand for Index Insurance

In this section we present a simplified version of Clarke's (2016) index-based insurance model to provide an intuition on the characteristics of demand for insurance products with substantial basis risk, such as the weather hedges sold to farmers in India. The model entails testable theoretical predictions that motivate our empirical work.

Consider a representative farmer who faces an agricultural income stream, Y, that depends on two states of nature, $S = \{Loss = L, Loss = 0\}$ such that $Y(S) = W - Loss$ where $Loss = L$ with probability p and $Loss = 0$ with probability $1 - p$. In the absence of any insurance opportunity, expected income is $E(Y) = W - pL$ and the variance of income is $Var\ (Y) = p(1 - p)L^2$. The farmer's preferences over income are represented by the indirect utility function $V(.)$, and the farmer is a strictly risk-averse agent; thus $V'(.) > 0$ and $V''(.) < 0$.

Now consider an index insurance product such as a weather hedge ($Index$) that pays L with probability p and 0 with probability $1 - p$. The random loss ($Loss$) and the random index payout ($Index$) have identical marginal distributions6 but are not independent. The index product is considered a good insurance instrument if it provides a payout when there is a loss and does not provide one when there is no loss. However, there is a joint probability r that there is a loss L ($Loss = L$) and the index payout is 0 ($Index = 0$) and, symmetrically, a joint probability r that there is no loss ($Loss = 0$) together with a positive index payout ($Index = L$). This joint probability r is the basis risk associated with the product. As a result, there are four states of nature $S = \{(Loss = L, Index = 0), (Loss = L, Index = L), (Loss = 0, Index = 0), (Loss = 0, Index = L)\}$ with probabilities $P(S)$ where $P(L, 0) = r, P(L, L) = p - r, P(0,0) = 1 - p - r, P(0, L) = r$. We summarize this in Table 2.1. In this context, no basis risk means $r = 0$, which makes the random variables $Loss$ and $Index$ identical: $Loss = Index$.[7]

Table 2.1 — Joint Probability of Income Loss and Index Payout

Joint probability		*Index*	
		0	*L*
Loss	*0*	*1—p—r*	*r*
	L	*r*	*p—r*

[6] We make the two random variables identically distributed for simplicity and to save on notation, but this condition can be easily removed.

[7] Note that if $r = p\ (1 - p)$, the random variables $Loss$ and $Index$ become identically and independently distributed. Therefore we require $r < p\ (1 - p)$ to have an index product that provides at least some insurance services.

The price of one unit of the index product is m times the actuarially fair price: pL. The farmer has the option to choose the quantity α of the index product she wishes to purchase. Her total spending is then αmpL and her net income $Y(S) = W - \alpha mpL - Loss(S) + \alpha\ Index(S)$, depending on one of the four potential states S. We summarize the income for each state with and without insurance in the real line in figure 1. With insurance, the expected income is $E(Y) = W - \alpha mpL - (1-\alpha)pL$ and the variance of income is $Var(Y) = [p(1-p)(1-\alpha)^2 + 2\alpha r]L^2$. The price multiple ($m$) negatively affects the mean but not the variance. Basis risk (r) does not directly impact the expected income, but it has a positive direct impact on income variance.

2.1.1 Maximization Problem

Given the joint probability distribution of $Loss$ and $Index$ and the price of the index product, the farmer chooses the quantity α of the insurance product that maximizes her expected indirect utility:

$$\max_{\boldsymbol{\alpha}} E\left[V\left(Y(S)\right)\right] \tag{2.1}$$

$$Y(S) = W - \alpha mpL - Loss(S) + \alpha\ Index(S) \tag{2.2}$$

The first-order condition is as follows:

$$\mathrm{E}\left[\mathrm{V'(Y(S))}\ \frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right] = 0 \tag{2.3}$$

In the absence of basis risk ($r = 0$) and actuarially fair price ($m = 1$), full insurance ($\alpha^* = 1$) is the optimal solution:

$$\mathrm{p\ V'(W - \alpha pL - L + \alpha L)\ (1-p)L - (1-p)\ V'(W - \alpha pL)\ pL = 0} \tag{2.4}$$

Under this scenario, income becomes $Y(S) = W - pL$ for all states with positive probability, and income variance is equal to zero.

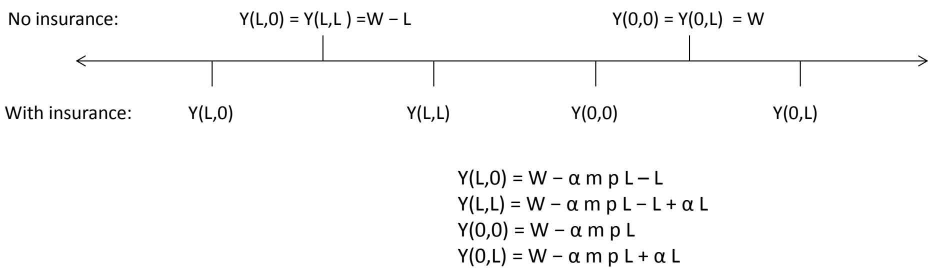


Figure 2.1 — Income with and without insurance

2.1.2 Basis Risk

Clarke (2016) shows that in the presence of basis risk ($0 < r < p(1-p)$) optimal insurance demand α^* is decreasing in basis risk r**.** Here, we discuss the economic intuition of that result.

After we introduce basis risk in the model, full insurance is no longer optimal. Basis risk entails a positive probability of getting into a very adverse situation in which income is subject to a loss L and there is no index payout despite having spent αmpL on the insurance product; under our notation income in this state would be $Y(L,0)$. This is even worse than experiencing a loss without any insurance. On the flip side, there is the very lucky situation in which, despite not having a loss, the index pays out; this state would be $Y(0,L)$.

As the farmer purchases more insurance (higher α), two effects appear on the dispersion of income: Income at *middle* states in figure 1, $Y(L,L)$ and $Y(0,0)$, becomes less disperse; and income at *extreme* states in figure 1, $Y(L,0)$ and $Y(0,L)$, becomes more disperse. When basis risk is low, the first effect tends to dominate, and for many values of $V''(.)$ it is optimal to purchase some insurance to decrease overall net income dispersion. When basis risk is high (high r), the second effect becomes relatively more important, as extreme incomes are more likely to occur.

While the proper notion of income dispersion is not entirely reflected by income variance (as the third moment is also relevant—especially in the case of highly risk-averse agents), how the income variance is affected by basis risk is still informative. In the absence of basis risk, more insurance helps decrease income variance until $\alpha^* = 1$, but in the presence of basis risk, variance is higher and more insurance helps decrease the variance only until $\alpha^* = 1 - [r/p(1-p)]$**,** which is less than 1, see figure 2. When basis risk is too high ($r = p(1-p)$), a positive amount of insurance ($\alpha > 0$) only increases the variance and the optimal choice becomes not to purchase insurance at all.

An alternative path to Clarke's (2016) formal proof is to use standard comparative static analysis. We differentiate the first-order condition with respect to α and r:

$$E\left[V''(Y(S))\left[\frac{\partial Y(S)}{\partial \alpha}\right]^2\right]d\alpha + \left[\sum_S\left(\frac{\partial P(S)}{\partial r}V'(Y(S))\frac{\partial Y(S)}{\partial \alpha}\right)\right]dr = 0 \qquad (2.5)$$

The second term is equal to $(1 - mp)\,L\left[V'(Y(0,L)) - V'(Y(L,L))\right] + mpL\left[V'(Y(0,0)) - V'(Y(L,0))\right]$. By concavity of the utility function, both terms in brackets are negative and therefore the whole term is negative.[8] Hence, in Equation 2.5 the first and second terms are both negative, and then $\frac{d\alpha^*}{dr} < 0$.[9]

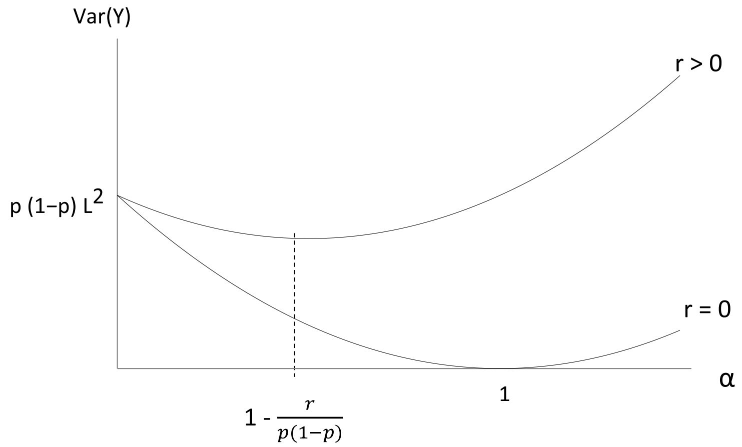


Figure 2.2 — Optimal insurance under basis risk

2.1.3 Price

We now turn to looking at the relationship between price and demand. We show through standard comparative static analysis Clarke's (2016) result that, in the case of a constant absolute risk aversion (CARA) utility function, the optimal demand α^* is monotonically not increasing in price mpL. Since we want to keep the probability p and the loss L constant, we consider variations in prices that come through variations in the multiple m. First, note that for a given insurance coverage α, a higher multiple m reduces income in all states S by the same magnitude. In this case, the reduction in the lowest income $Y(L, 0)$ is the most hurtful to the farmer's welfare and therefore induces a decrease in α^* to limit that particular welfare loss (by reducing total spending in insurance αmpL)**.** However, a lower α has a negative impact on the income of states (L, L) and $(0, L)$, and this might induce the farmer to increase α under certain conditions. Here, we show that this is not the case under a CARA utility function.

We first differentiate the first-order condition with respect to α and m:

$$E\left[\mathrm{V}''\big(\mathrm{Y(S)}\big)\left[\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right]^2\right]\mathrm{d}\alpha + E\left[\mathrm{V}''\big(\mathrm{Y(S)}\big)(-\alpha\mathrm{pL})\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right]\mathrm{dm} + E\big[\mathrm{V}'\big(\mathrm{Y(S)}\big)(-\mathrm{pL})\big]\mathrm{dm} = 0 \tag{2.5}$$

---

[8] As shown below, 1 – mp > 0; otherwise, the demand for insurance is zero.

[9] Note that by the first order conditions, $E\left(\mathrm{V}'\big(\mathrm{Y(S)}\big)\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right) = \sum_S \left(\mathrm{P(S)V}'\big(\mathrm{Y(S)}\big)\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right) = 0$. Multiply and divide the second term by P(S) to show that this is a negative term.

Dividing and multiplying the second term by $\mathrm{V}'\big(\mathrm{Y(S)}\big)$ we get

$$E\left[\mathrm{V}''\big(\mathrm{Y(S)}\big)\left[\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right]^2\right]\mathrm{d}\alpha \; + \alpha \mathrm{pL}\,\gamma E\left[\mathrm{V}'\big(\mathrm{Y(S)}\big)\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right]\mathrm{dm} \\ + E\big[\mathrm{V}'\big(\mathrm{Y(S)}\big)(-\mathrm{pL})\big]\mathrm{dm} \; = 0 \tag{2.6}$$

where γ is the coefficient of risk aversion. The first term is negative; the second term is equal to the first-order condition, multiplied by a constant, and thus equal to zero; and the third term is negative. Therefore $\frac{\mathrm{d}\alpha}{\mathrm{dm}} < 0$.

2.1.4 Demand Price Elasticity and Basis Risk

We claim that the responsiveness of the demand to variations in price is a function of the degree of basis risk r. Our conjecture is that when basis risk is low, the elasticity of demand is higher than when basis risk is high. Although we don't provide a formal proof of such relationship, our intuition is based on the fact that when there is no basis risk, the elasticity of demand is high; and when basis risk is extremely high, the demand for insurance is zero and unresponsive for any price above the actuarially fair price or $\mathrm{m} \geq 1$.

We have already established that in the absence of basis risk $(\mathrm{r} = 0)$ and actuarially fair price $(\mathrm{m} = 1)$, full insurance $(\alpha^* = 1)$ is the optimal solution and therefore income is constant across states such that $\mathrm{Y(S)} = \mathrm{Y}$ for any relevant state S. In this particular case we can estimate the responsiveness of the demand to a change in the multiple m (and therefore to a change in price while keeping constant p and L):

$$\left.\frac{\mathrm{d}\alpha}{\mathrm{dm}}\right|_{\mathrm{m}=1,\mathrm{r}=0} = -\frac{\mathrm{E}\big[\mathrm{V}'\big(\mathrm{Y(S)}\big)(-\mathrm{pL})\big]}{\mathrm{E}\left[\mathrm{V}''\big(\mathrm{Y(S)}\big)\left[\frac{\partial \mathrm{Y(S)}}{\partial \alpha}\right]^2\right]} = \frac{\mathrm{V}'(\mathrm{Y})}{\mathrm{L}(1-\mathrm{p})\mathrm{V}''(\mathrm{Y})} < 0 \tag{2.7}$$

which in the case of a CARA utility function becomes $-\frac{1}{\mathrm{L}(1-\mathrm{p})\gamma} < 0$. And more generally from the previous section we know that $\left.\frac{\mathrm{d}\alpha}{\mathrm{dm}}\right|_{\mathrm{r}=0} < 0$. Now we look at the responsiveness of the demand when basis risk is extremely high: $\mathrm{r} \geq \mathrm{p}\,(1-\mathrm{p})$. We show that for any price above the actuarially fair price ($\mathrm{m} \geq 1$), the demand is unresponsive and equal to zero (see Appendix).

2.1.5 Risk Aversion

Clarke (2016) shows that in the presence of basis risk and when premiums are above actuarially fair, a hump-shaped demand with respect to risk aversion is expected. Since the model presented here is a particular case of Clarke's model, the same conclusion applies here. The intuition is as follows: A risk-neutral agent won't buy an actuarially unfair product $(\mathrm{m} > 1)$, because the expected income is all that she cares about and expected income is decreasing in insurance coverage (α) when $\mathrm{m} > 1$. As income in the worst-case scenario is decreasing in insurance coverage (α), extremely risk-averse agents would also be unwilling to buy insurance as they would not be willing to sacrifice an income reduction in the worst-case scenario $Y(L, 0)$ (income in this state is decreasing in insurance coverage, α), despite the reduction of income dispersion in middle states that insurance coverage affords. Moderately risk-averse agents would choose to have some insurance coverage $(\alpha > 0)$ as they would be willing to trade some income loss in the worst-case scenario against less income dispersion in middle states (L, L) and $(0,0)$ and an income gain in state $(0, L)$.

2.2 Empirical Design

We worked with the insurance company HDFC ERGO to identify suitable villages to be included in our study. Suitable villages were defined as those that were 15 km or less from a weather station, in districts that were not notified for provision of subsidized insurance, and in villages in which HDFC had a marketing presence. Additionally, it was important to select villages that were neither too small nor too large for surveying and marketing activities. First, administrative data on the number of households within a village was used to exclude villages of fewer than 100 households and more than 500 households. This resulted in a list of about 120 villages in three districts. Second, 45 villages each in Dewas and Bhopal and 20 villages in Ujjain, 110 villages in total, were randomly selected for inclusion in this study.

The 110 sampled villages were randomly allocated into treatment and control villages: 72 treatment villages were selected for insurance to be offered in these villages. HDFC agreed that no insurance would be offered in the remaining control villages. More treatment villages than control villages were selected given the multiple treatment arms in this study. Villages in Bhopal and Dewas were allocated to treatment and control categories using a random draw with no stratification or blocking. Ujjain villages were allocated to treatment and control categories separately, on account of the later inclusion of this district. As such, stratification of the villages occurred along district lines. For this reason we include district dummies in our regression analyses.

Six index insurance products, simple weather hedges, were sold in each district covering deficit and excess rain at the beginning, middle, and end of the season. For every peril identified in each of the covered periods, two types of coverage were available: one for a fixed payout in case of a lower probability event and the other in case of a higher probability event. The product was designed in such a way that the period, payouts and perils covered were the same in all the districts but each district had a different set of triggers (index levels corresponding to a payout). Details of the products offered are provided in the Appendix Table 1. Farmers were free to choose the number and combination of policies (index insurance products) that they wished to purchase. The policies were priced at actuarially fair prices plus administration costs. There were only minor differences in pricing between the policies offered in 2011 and 2012.

The marketing of these weather hedges was carried out by HDFC ERGO General Insurance Company in three different phases, prior to the start of each cover period.[10] In 2011, sales began partway through the season, with farmers in two districts (Dewas and Bhopal) being offered policies for the middle and end of the season, and farmers in one district (Ujjain) being offered policies for only the end of the season. Given that randomization in Ujjain was independent from that in the other districts, this does not affect the estimation of treatment effects for 2011. In 2012, all farmers were offered all policies for the three coverage periods of the season. Moreover in 2012, a door-to-door marketing exercise was again conducted by sales agents to ensure that all households in the sample were again offered insurance.

2.2.1 Basis Risk

Some exogenous variation in the degree of basis risk associated with the weather hedges was introduced by installing three new randomly located weather stations that would trigger payouts. Only villages in Dewas and Bhopal were eligible for this treatment, meaning that the 13 Ujjain villages that received insurance used pre-existing reference stations. The new weather stations were installed in locations selected according to the following process:

1. We randomly selected one village in which to place a new weather station. All villages very close (10 km or less) to this one were then excluded from further selection.
2. Out of the remaining villages, we randomly selected a second location. All villages very close (10 km or less) to this one were then excluded from further selection.
3. Out of the remaining villages, we randomly selected a third location.

The three villages selected using this process were Polayjagir and Talod in Dewas, and Intkhedi Sadak in Bhopal. Villages within 10 km of these sites were eligible to be serviced by these new stations. Pre-existing weather stations in each of the three districts were used as a reference station for those not selected to be serviced by the new station. The full list of weather stations used in this study is listed in Table 2.2.

Table 2.3 presents comparisons of key village characteristics between treatment villages served by a new weather station and those served by pre-existing weather stations. The results from these tests of balance indicate that villages in these two treatment groups do not differ systematically on observable characteristics. We also compared villages closer to a weather station with those farther away from it. Tests of balance between

[10] Van campaigns, pamphlet distributions and mass SMS messages were undertaken in all the treatment villages. In addition, meetings of farmer groups were organized in the first two phases for marketing at the village level. After this, door-to-door visits took place for final marketing and selling of insurance policies.

villages below and above the median distance to the assigned weather station show no significant differences on observable characteristics.

Table 2.2 — Weather Station Assignment

Trigger weather station	Weather station with historical data used for product design	Number of villages covered
Dewas—Sonkatch (NCSML)	Indore (IMD)	9
Dewas—Polayjagir—New	Indore (IMD)	10
Dewas—Talod—New	Indore (IMD)	10
Bhopal (IMD)	Bhopal (IMD)	18
Bhopal—Intkhedi Sadak—New	Bhopal (IMD)	12
Ujjain-Khachrod (IMD)	Ujjain(IMD)	13

Note: NCSML refers to weather stations installed and operated by National Collateral Management Services Limited. IMD refers to weather stations run by the Indian Meteorological Department.

2.2.2 Price

Exogenous variation in the price of the weather hedges was introduced by randomly allocating price discount vouchers among treatment households. If a household then chose to purchase the index insurance product, it could exercise the voucher at the moment of purchase. Based on group discussions and given the level of education of targeted farmers, we concluded that absolute numbers would be easier to understand by farmers than percentage discounts. Four levels of absolute discounts were selected, broadly equivalent to 15 percent, 30 percent, 45 percent, and 60 percent of the cheapest policy, in order to have enough price variation along a hypothetical demand curve. The level of discount received by a household held across all available insurance products for that season. For example, if a household received a Rs.45 discount voucher, the household was entitled to write off Rs.45 from the price of all index insurance products it chose to purchase during the entire season.

There were important differences in price discounts between 2011 and 2012. In 2011, surveyed households received a discount voucher through a household-level random draw in two districts (Dewas and Bhopal) and through a village-level random draw in Ujjain. In 2012, discounts were randomized at the village level in all districts. The decision to provide the same level of price discount to all farmers (sampled and non-sampled households) in a village was made in response to the insurance company's concerns about a discouragement effect arising from not receiving a voucher, where farmers discriminated by the price-discount distribution process could develop a negative attitude toward the product. In addition, a village-wide price discount was cheaper to implement.

Table 2.3 — Tests of Balance between Villages Assigned to New and Old Weather Stations

	Mean across new weather station villages	Mean across old weather station villages	T-test of difference
Variables from households' listing			
Number of households	218.93	209.35	0.34
Proportion of type 0 households	0.39	0.42	–0.68
Proportion of type 1 households	0.21	0.2	0.53
Proportion of type 2 households	0.24	0.22	0.84
Proportion of type 3 households	0.05	0.05	–0.14
Proportion of type 4 households	0.11	0.11	–0.02
Proportion of female-headed households	0.04	0.04	1.12
Average years of education of the household head	4.43	4.38	0.16
Proportion of SC/ST/OBC	0.83	0.82	0.16
Average land owned (in acres)	3.64	3.42	0.56
Variables from baseline survey			
Distance to weather station (in kms.)	5.02	10.1	–5.89
Distance to market (in minutes)	46.15	47.35	–0.24
Average cultivated acreage	6.93	6.72	0.36
Proportion of land that is irrigated	0.74	0.78	–0.88
Average cultivated acreage in the *Rabi*	5.72	5.9	–0.37
Proportion of land on which wheat is grown	0.64	0.66	–0.46
Proportion of land on which chickpeas are grown	0.25	0.26	–0.21
Average cultivated area in the *kharif*	6.33	6.26	0.14
Proportion of land on which soybeans are grown	0.88	0.92	–1.69
Prop. of households reporting drought in last 10 years	0.11	0.14	–0.71
Prop. of households that had some agric. insurance	0.38	0.29	1.52
Prop. of households that had some insurance	0.64	0.56	1.52
Prop. of households with access to agricultural loans	0.94	0.93	0.69

Note: SC/ST/OBC refers to scheduled caste, scheduled tribe and other backward castes. Households are classified in types according to whether they own more (+) or less (-) than six acres of land, and whether the household decision maker has had more (+) or less (-) than five years of schooling. Type 0 households own no land. Type 1: (-)Land/(-)Schooling, Type 2: (-)Land / (+)Schooling, Type 3: (+)Land/(-)Schooling, Type 4: (+)Land/(+)Schooling. Source: Listing and household survey.

2.2.3 Understanding

Across all treated villages, decision makers in sampled households were invited to attend two hours of basic insurance literacy training in the first marketing season (2011). In these basic training sessions—which were also open to any other observers from the village—farmers were introduced to potential weather-related risks they might face and encouraged to discuss their current coping mechanisms. After this introduction, the majority of the remaining training focused on a general discussion of weather index insurance, the way in which it had been tailored to their circumstances, and the characteristics of the product. Interactive games were played with the farmers, with the games intending to illustrate the costs and benefits of purchasing these hedges. A final iteration of games helped farmers understand that the ability of the insurance company to pay their claims was not dependent on the weather outcome of other farmers. This was intended to build trust among the farmers toward the insurance company.

Additionally, 37 of the treated villages were randomly selected to receive an additional two-hour training. Households in our sample were again actively encouraged to attend the meeting, and all villagers were allowed to participate. In this additional training session the basic concepts were reiterated, and any questions and concerns that the farmers had were addressed.

Tests of balance between villages with basic training and villages with basic plus intensive training are presented in Table 2.4 and show that the two groups are balanced across common household characteristics. Some form of training was provided to all households in order to ensure a wide understanding of the product being offered. Because of this, we are able to analyse the impact of receiving intensive training with respect to receiving only basic training; we cannot estimate the impact of receiving some training against no training.

2.3 Data

A summary of the timeline of activities is provided in figure 3. An initial listing exercise was conducted in all selected villages during January and February 2011, before survey work, training, and insurance sales began. The listing exercise collected basic information on household characteristics such as age, gender, and education of the household head; caste; housing structure; landownership; and main crop of production in the *rabi* season.

Data Collection		Insurance Activities				Coverage Period					
		First Season—2011									
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Listing and baseline survey		New weather stations installed		Literacy training (Basix)	Discounts provided and insurance sales						
		Second Season—2012									
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Follow-up survey					Discounts provided and insurance sales						

Figure 2.3 — Timeline of activities

The listing survey provides two important contributions to the study. Because much of the randomization in our design is conducted at the village level, aggregation of household data from the listing survey allows us to ensure balance across village-level statistics. Additionally, given that purchase rates of index insurance are generally found to be quite low in similar studies, it was important for us to focus our energies on households that would be relatively more inclined to purchase insurance. The information from the listing questionnaire allows us to identify these individuals and oversample them.

Studies of insurance demand in India suggest that those who purchase weather index insurance have larger landholdings and higher education levels than those who do not (Giné, Townsend, and Vickery 2008; Cole et al. 2013). Data on education of the household decision maker and landholding of the household collected in the listing survey were used to classify households into five categories. Households in the first category, type 0 households, are those that do not own any land. As these households were not allowed to purchase weather insurance, they were not included in the survey sample or in the training and marketing activities. Households that own land were further categorized into four types:

- Type 1 households own six acres or less of land and have a household decision maker with less than five years of schooling
- Type 2 households own six acres or less of land and have a household decision maker with five years or more of schooling
- Type 3 households own more than six acres of land and have a household decision maker with less than five years of schooling
- Type 4 households own more than six acres of land and have a household decision maker with five years or more of schooling.

We sampled 30 households from each village. On average, the proportion of type 4 households was found to be much lower than the proportion of type 1, 2, or 3 households. However, type 4 households are, a priori, the ones most likely to buy insurance. As a result, these were oversampled in our sampling strategy, such that half of the 30 sampled households would belong to this category. The remainder of the randomly selected sample in each village consisted of 3 type 1 households, 6 type 2 households, and 6 type 3 households.

A baseline survey was conducted among all sampled households in January–February 2011, immediately after the listing exercise had been completed in a village. Sampled household characteristics are displayed in Tables 2.3 and 2.4. A follow-up survey was conducted in January–February 2012 among the same households. Attrition in the follow-up survey was only 2.16 percent and not significantly different between treatment groups.

On one hand, we use data collected in the follow-up survey on understanding of and attitudes toward insurance to assess whether the interventions had the expected effect. On the other hand, to assess the impact of the interventions on realized demand for insurance, we use administrative data on insurance sales. The advantage is that these data are available for both 2011 and 2012 and are expected to be more accurate than self-reported survey data on purchases.

Table 2.4 — Tests of Balance between Villages Offered Intensive and Basic Insurance Literacy Training

	Mean in intensive training villages	Mean in basic training villages	T-test of difference
Variables from households' listing			
Number of households	212.68	215.09	–0.1
Proportion of type 0 households	0.39	0.37	0.35
Proportion of type 1 households	0.23	0.22	0.49
Proportion of type 2 households	0.21	0.23	–0.7
Proportion of type 3 households	0.06	0.06	–0.52
Proportion of type 4 households	0.12	0.12	–0.48
Proportion of female headed households	0.04	0.04	0.43
Average years of education of the household head	4.17	4.29	–0.47
Proportion of SC/ST/OBC	0.81	0.82	–0.11
Average land owned (in acres)	3.77	3.82	–0.13
Variables from baseline survey			
Distance to weather station (in kms.)	8.02	8.91	–0.67
Distance to market (in minutes)	47.34	48.03	–0.15
Average cultivated acreage	6.85	6.8	0.09
Proportion of land that is irrigated	0.71	0.74	–0.82
Average cultivated acreage in the *rabi*	5.65	5.65	0
Proportion of land on which wheat is grown	0.63	0.6	0.8
Proportion of land on which chickpeas are grown	0.28	0.29	–0.27
Average cultivated area in the *kharif*	6.4	6.22	0.38
Proportion of land on which soybean is grown	0.9	0.93	–1.13
Prop. of households reporting drought in last 10 years	0.16	0.13	0.72
Prop. of households that had some agric. insurance	0.3	0.38	–1.59
Prop. of households that had some insurance	0.58	0.57	0.19
Prop. of households with access to agricultural loans	0.93	0.92	0.37

Note: SC/ST/OBC refers to scheduled caste, scheduled tribe and other backward castes. Households are classified in types according to whether they own more (+) or less (-) than six acres of land, and whether the household decision maker has had more (+) or less (-) than five years of schooling. Type 0 households own no land. Type 1: (-)Land/(-)Schooling, Type 2: (-)Land / (+)Schooling, Type 3: (+)Land/(-)Schooling, Type 4: (+)Land/(+)Schooling. Source: Listing and household survey.

In 2011, HDFC insurance agents sold a total of 308 contracts in treatment villages.[11] Take-up was 6.8 percent among sampled households. A majority of these sales, 263, correspond to the third insurance period. The remaining contracts covered the second period. The first period had no contracts because HDFC insurance agents did not get to market for it. In Bhopal and Dewas, the majority of transactions (more than 95 percent) were for the more comprehensive, more expensive contract; but in Ujjain, more than 90 percent of the purchased contracts were for the less comprehensive, cheaper policy. In

[11] This includes insurance purchases from both, sampled and non-sampled households within treatment villages.

2012, the program was less successful overall, with only 185 contracts sold. Take-up was 4.0 percent among sampled households. Although the majority of contracts purchased were for the less comprehensive, cheaper policy (96 percent), there was heterogeneity regarding the covered crop phases between districts. In Bhopal, almost two-thirds of the contracts covered risks in the third phase (around harvest), with the rest mainly concentrated in the first phase (sowing). In Dewas, purchases were relatively stable across the three periods. In contrast, in Ujjain, 90 percent of the purchased contracts corresponded to the first phase.

Table 2.5 — Summaries of Insurance Purchases by District and Sample

	Treated Villages			**Household Sample**		
	Number of sales	Acres insured	Acres insured per sale	Number of households insured	Acres insured	Acres insured per purchasing household
2011 sales						
Ujjain	115	59.5	0.5	10	8.5	0.9
Dewas	48	70	1.5	16	43	2.7
Bhopal	145	49	0.3	123	43.75	0.4
Total	**308**	**178.5**	**0.6**	**149**	**95.25**	**0.6**
2012 sales						
Ujjain	83	114	1.4	10	66	6.6
Dewas	75	150	2.0	44	85	1.9
Bhopal	27	60	2.2	33	31	0.9
Total	**185**	**324**	**1.75**	**87**	**182**	**2.1**

Transaction data are summarized in Table 2.5. In both years, more than half (173 in 2011 and 182 in 2012) of the total contracts were purchased by households sampled in the baseline survey and thus invited to attend training sessions. A number of households bought multiple contracts, as indicated by the total number of households insured in Table 2.5. In 2011, although relatively few households purchased insurance in Dewas and Ujjain, on average they insured more land. This is also true of 2012 sales. The remaining contracts were bought by 125 households (in 2011) and 75 households (in 2012) that were not in the baseline sample and did not attend the training sessions. In our analysis, we will focus on uptake and demand aggregated across the entire season.

2.4 Analysis

The random allocation of price discounts, placement of weather stations, and delivery of additional training allows us to estimate the intent-to-treat (ITT) effect of these interventions on demand for weather hedges through a direct comparison of purchases across different treatment arms.

For the most part, our dependent variable is a dummy variable taking the value of 1 if a household purchased insurance in a given season. To account for the dichotomous nature of our dependent variable, we estimate a logit model. For ease of interpretation, all tables report the average marginal effects across the sample. Since the distance of a farmer's plot to a pre-existing weather station is potentially endogenous, when we consider the impact of distance to the weather station on demand, we need to instrument this variable. We do so by following the approach in Smith and Blundell (1986), including the estimated residuals from the instrumenting regression in the main regression. We use as the instrument whether the insurance policy for a given farmer was referenced to an existing or to a new weather station. We bootstrap the standard errors of the main regression to account for the presence of the estimated residuals from the instrumenting regression.

We also present some results for the quantity of insurance purchased, defined as the number of acres of land that the household insured, taking a value of zero if the household did not purchase any insurance. We estimate these specifications by ordinary least squares and by two-stages least squares when endogenous explanatory variables are included.

2.4.1 The Initial Impact of Marketing Interventions

In Tables 2.6 and 2.7 we present results on the impact of our three interventions—price discounts, investment in weather stations, and intensive training—on demand for weather hedges. Table 2.6 captures demand with the take-up dummy discussed above. Table 2.7 uses the natural logarithm of units purchased as a dependent variable. The price variable is defined as the natural logarithm of the price (after discount) of the cheapest policy. Since the ratio of the price between the policies for the low- and high-probability events is constant across all districts, the natural logarithm of the discounted price of the cheapest policy is a good measure of the discount value for all policies.

First, we look at the impact of being offered a price discount. Table 2.6 shows that receiving a price discount has a substantial effect in terms of encouraging a household to purchase insurance. A 10 percent decrease in price seems to lead to a 1.3 percentage point increase in take-up, which, given the low levels of average take-up, corresponds to a 19 percent increase in demand from the average. Similarly, we find a significant price sensitivity in Table 2.7. Here, the dependent variable is the natural logarithm of the units purchased, and as such, the coefficient on the price variable is a measure of the elasticity of demand. We find a considerable price elasticity of 0.58, not significantly different from the price elasticity of 0.66 to 0.88 estimated by Cole et al. (2013) for weather index insurance in other states in India.[12]

[12] All of the above results are robust to including household-level covariates in the estimations. This is also true for the following results on weather station investments and intensive insurance training. This is expected, given that the tests of balance indicate no significant differences across household characteristics between treatment groups.

Table 2.6 — Take-Up Among Sampled Households, 2011

	(1) Logit	(2) Logit	(3) IV Logit	(4) Logit	(5) IV Logit	(6) Logit
Log (price)	–0.135***	–0.207***	–0.133***	–0.134***	–0.127***	–0.126***
	(0.024)	(0.032)	(0.026)	(0.023)	(0.031)	(0.038)
Log (distance to	–0.018***	–0.010*	–0.042	–0.023***	–0.056	
weather station)	(0.007)	(0.006)	(0.031)	(0.009)	(0.035)	
Intensive training	0.050*	0.020	0.049			
	(0.026)	(0.041)	(0.032)			
Log (distance) x log (price)				(+)**	(+)**	
Station is close						0.071*
						(0.041)
Station is close x log (price)						(–)**
Sample	Full	New station	Full	Full	Full	Far and close
Observations	2,183	848	2,183	2,183	2,183	932

Notes: Average marginal effects are reported. IV specifications instrument log of distance with assignment to a new weather station. In columns (4), (5), and (6), as marginal effects are not defined for the interaction term, we show the sign and significance for the parameter in the logit equation. Standard errors for the marginal effects, clustered at the village level, are in parentheses. Standard errors are bootstrapped in IV specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.7 — Log of Units Bought, 2011

	(1) OLS	(2) OLS	(3) IV OLS
Log (price)	–0.582***	–0.579***	–0.573***
	(0.133)	(0.134)	(0.134)
Log (distance)		–0.081**	–0.142
		(0.040)	(0.113)
Intensive		0.176*	0.177*
		(0.091)	(0.091)
Observations	2,183	2,183	2,183
R-squared	0.100	0.114	0.111

Notes: IV specifications instrument log of distance with assignment to a new weather station. Standard errors, clustered at the village level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Second, we consider the impact of weather station investment. To estimate the impact of basis risk on demand, we use the distance of a household to the reference weather station in the insurance policy they were offered. Using global positioning system (GPS) coordinates of households and GPS coordinates of the weather stations (new and old), we calculate the straight-line distance between each sample household and the reference weather station. However, this distance cannot be assumed as exogenous for households that were not assigned to a new weather station. Therefore, in column 2 we present results only for those that were assigned to a new weather station, and in column 3, for all sample households but instrumenting distance with an indicator variable taking the value of 1 if the reference weather station was a new, randomly assigned station and zero otherwise. Results from the first stage of this regression (Appendix Table 2) show a significantly negative relationship between the instrument and distance. On average, households assigned to new reference weather stations were 6 km closer than those assigned to preexisting ones. Both the reduced-sample logit and the full-sample IV estimates suggest that increasing the distance from the weather station reduces uptake, though the IV specification yields a non-significant estimate. Doubling the distance to the weather station from the average reduces take-up by roughly 1.25 percentage points. An overall similar pattern holds when we consider the natural logarithm of quantity purchased as the dependent variable (Table 2.7).

Third, we assess the impact on insurance demand of the intensity of training provided. Table 2.6 shows that households that were offered intensive training modules had significantly higher insurance demand. Take-up among those that were offered intensive training was about 5 percentage points higher than among those that were offered basic training only, but the effect is only weakly significant. The amount of insurance purchased (Table 2.7) was also weakly significantly higher among those that were offered the intensive training.[13]

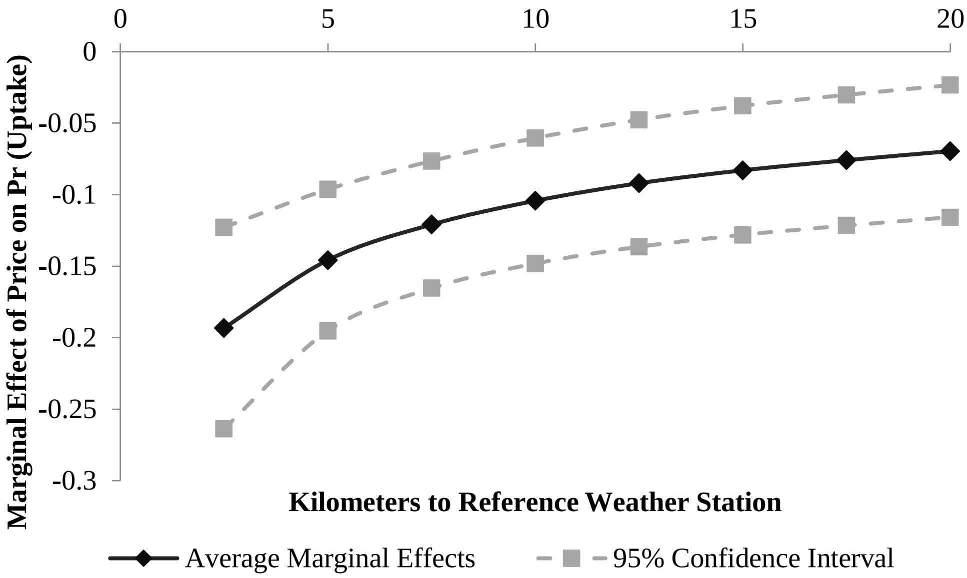


Figure 2.4 — Price sensitivity of demand as distance to the weather station increases, 2011

[13] For both (basic and intensive) training sessions the participation rates were considerably high. Around 70% of invited households attended, 20% attended through a representative, and the remaining 10% received training via later household visits. We find no differential results in take-up related to type of attendance. Also type of attendance is not related to household level covariates.

Thus far we have considered the impact of the treatments in isolation. However, we are likely to observe a different price elasticity for households offered a *good* insurance product, that is, with less basis risk, than for households offered a *bad* insurance product, one with high levels of basis risk. We test this assumption in columns 4 to 6 of Table 2.6 by interacting price and distance to the reference weather station. We indeed find this to be the case: The sensitivity of demand to price increases the closer a household is to the product's reference weather station.[14] We further show this divergence in figure 4, which plots the average marginal effect of the logarithm of price on the probability of take-up for households located at different distances from the reference weather station. As an alternative exercise, column 6 restricts the sample to only those households that are located less than 5 km or more than 12 km from their reference weather station. We then use an indicator variable that takes the value 1 if a household belongs to the first group. Households located less than 5 km from a weather station have a sensitivity to price almost 10 times higher (marginal effect of –0.20) than that of those located more than 12 km from a weather station (marginal effect of –0.02). Overall, this suggests that subsidies are more effective in encouraging demand when complementary investments are made aimed at reducing basis risk.

Table 2.8 — Impact of Offering Insurance on Insurance Knowledge and Attitudes

	(1) Knowledge about insurance	(2) Rainfall at weather station is similar	(3) Trust government insurance to pay	(4) Trust private insurance to pay	(5) Would buy if previously good year and no payout	(6) Would buy if previously bad year and no payout
Offered insurance	-0.001	0.038	0.010	0.065**	0.006	0.058
	(0.001)	(0.036)	(0.017)	(0.031)	(0.031)	(0.041)
Mean in control	3.02	0.14	0.93	0.19	0.76	0.27
Observations	3,237	3,209	3,234	3,231	3,235	3,232

As the knowledge variable takes values from 0 to 5, column (1) is estimated through an ordered logit. The rest of the columns are estimated through logit specifications. Average marginal effects are reported. Lagged dependent variable (from baseline survey) and district dummies are included but not shown. Standard errors, clustered at the village level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Household survey data.

2.4.2 Impact on Insurance Knowledge and Attitudes

Using follow-up survey data collected in January 2012, we can explore different hypotheses for the mechanisms behind the treatment effects. In particular, we examine (1) whether comprehension of the insurance product was higher for those that were offered the intensive training and (2) whether households with insurance linked to rainfall at a

[14] While the marginal effect of the distance variable is not significant in the IV specification, both the distance and the interaction parameters are strongly significant in the estimated logit equation (p-values of 0.007 and 0.008, respectively).

new, closer weather station believed it to better resemble the actual rainfall on their land. In Table 2.8 we assess the effect of being offered insurance on different self-reported measures of knowledge and attitudes, relative to households in villages where no insurance was offered. This is the combined average effect of all activities in the villages in which insurance was offered: the general marketing insurance activities and basic training—which took place in all villages—plus the average effect of intensive training, discounts, and the placement of new weather stations in selected villages, together with the potential effects of actually having purchased insurance.

We see that, on average, households in villages that were offered insurance did not have a better understanding of insurance by the time of the follow-up survey (approximately 7 to 8 months after training). This does not necessarily mean, however, that training was ineffective or that this knowledge was short-lived, as both control and treatment villages significantly increased their levels of knowledge about insurance.[15] This could be due to a number of different reasons, such as spillover effects from treatment to control villages or other external shocks that affected both types of villages, such as the roll out of government insurance programs. The table does show though that households in treatment villages seem to trust private insurance companies more. This is somewhat expected since payouts stemming from the rainfall realizations of the previous season were disbursed promptly and without any problem.

Table 2.9 — Impact of Training, Discounts, and Weather Stations on Insurance Knowledge and Attitudes

	(1) Knowledge about insurance	(2) Rainfall at weather station is similar	(3) Trust government insurance to pay	(4) Trust private insurance to pay	(5) Would buy if previously good year and no payout	(6) Would buy if previously bad year and no payout
Log (distance)	0.133*	–0.088**	–0.023	0.020	–0.089*	–0.053
	(0.076)	(0.042)	(0.023)	(0.043)	(0.047)	(0.058)
Intensive	–0.064	–0.035	0.005	–0.021	–0.015	–0.048
training	(0.063)	(0.032)	(0.019)	(0.036)	(0.038)	(0.053)
Log (price)	–0.106**	–0.008	–0.003	0.032	–0.009	0.045
	(0.043)	(0.016)	(0.011)	(0.023)	(0.021)	(0.031)
Observations	2,130	2,111	2,127	2,126	2,128	2,125

IV specifications instrument log of distance with assignment to a new weather station. Lagged dependent variable (from baseline survey) and district dummies included but not shown. Standard errors, clustered at the village level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Household survey data.

Table 2.9 disentangles some of these effects. We find that having received intensive training has no effect on comprehension of the insurance product or the other attitude-related questions that households were asked. One plausible explanation is that insurance

[15] These results are unreported.

training was indeed effective (as shown by its significant positive effect over first year's demand), and control villages simply caught up in terms of insurance knowledge by the time of the follow-up survey. Alternatively, it could have been the case that the additional training served more of a marketing purpose rather than provided additional knowledge about insurance itself, and insurance knowledge per se increased equally on average in all villages due to other unidentified factors, external to our intervention. Unfortunately we do not have data to distinguish between these hypotheses.

The table does show, however, that households that were offered higher price discounts knew more about the insurance product being sold (perhaps due to an encouragement effect or because they were more likely to actually buy insurance and thus to engage in learning by doing). Moreover, and as expected, households that were offered insurance referenced to a new, closer weather station were more likely to believe it was a good approximation of actual rainfall on their land.

Table 2.10 — Take-Up Among Sampled Households, 2012

	(1) Logit	(2) Logit	(3) IV Logit	(4) Logit	(5) IV Logit	(6) Logit
Log (price)	–0.026**	–0.053***	–0.027*	–0.024*	–0.026	-0.018
	(0.012)	(0.017)	(0.015)	(0.013)	(0.018)	(0.024)
Log (distance to weather station)	–0.008**	–0.013***	–0.001	–0.006*	–0.000	
	(0.004)	(0.002)	(0.015)	(0.003)	(0.017)	
Intensive training	0.005	0.022	0.005			
	(0.012)	(0.016)	(0.012)			
Log (distance) x log (price)				(+)**	(+)	
Station is close						0.011
						(0.018)
Station is close x log (price)						(–)**
Sample	Full	New station	Full	Full	Full	Far and close
Observations	2,183	848	2,183	2,183	2,183	932

Notes: Average marginal effects are reported. IV specifications instrument log of distance with assignment to a new weather station. In columns (4), (5), and (6), as marginal effects are not defined for the interaction term, we show the sign and significance for the parameter in the logit equation. Standard errors for the marginal effects, clustered at the village level, are in parentheses. Standard errors are bootstrapped in IV specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Administrative sales data.

2.4.3 The Longer-Run Impact of Marketing Interventions

We now turn to the effect of the marketing interventions on take-up in 2012, the second season of sales for the index insurance product. As described in Section 2.2, in 2012 we again offered price discounts, randomized at the village level, but left the other treatments unchanged: No new weather stations were installed and no additional training sessions were conducted.

Table 2.10 presents specifications parallel to those in Table 2.6 but for the second season of index insurance sales. The price of the insurance policy was again strongly significant in predicting demand. The distance to the weather station was also strongly associated with purchases both in the full sample and in the subsample of villages served by the new stations, although the instrumental variables results on the full sample no longer hold. The interaction between price and basis risk also replicates the results already discussed for 2011.

Table 2.11 — Take-Up Among Households, 2012, Including Price in 2011

	(1) Logit	(2) IV Logit	(3) Logit	(4) IV Logit	(5) Logit
Log (price, 2012)	–0.026**	–0.027*	–0.024**	–0.026	–0.018
	(0.012)	(0.015)	(0.013)	(0.018)	(0.025)
Log (price, 2011)	–0.002	–0.002	–0.003	–0.004	0.001
	(0.010)	(0.010)	(0.010)	(0.010)	(0.017)
Intensive	0.005	0.005			
	(0.012)	(0.013)			
Log (distance)	–0.008**	–0.001	–0.006*	–0.000	
	(0.004)	(0.015)	(0.003)	(0.016)	
Log (distance) x Log (price, 2012)			(+)**	(+)	
Station is close					0.011
					(0.018)
Station is close x Log (price, 2012)					(–)**
Sample	Full	Full	Full	Full	Far and close
Observations	2,183	2,183	2,183	2,183	932

Notes: Average marginal effects are reported. IV specifications instrument log of distance with assignment to a new weather station. In columns (3), (4), and (5), as marginal effects are not defined for the interaction term, we show the sign and significance for the parameter in the logit equation. Standard errors for the marginal effects, clustered at the village level, are in parentheses. Standard errors are bootstrapped in IV specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Administrative sales data.

Now, while during the second year both price discounts and the distance to the weather station still seem to significantly (though less strongly) affect the demand for insurance, the effect of intensive training completely fades out: Although receiving intensive training considerably increased demand in the season immediately following training, it had no significant effect on demand a year later. Again, as discussed in Section 2.3.2, this is consistent with both control and treatment villages showing similar increases in their levels of insurance knowledge, together with the fact that any pure marketing effects stemming from being offered intensive training during the first season may have worn off a year later. The results, however, are indicative of a learning by doing effect, where being offered a higher discount or being located closer to the weather station could encourage a household to get better informed about the insurance product.

Table 2.11 includes the (discounted) price faced in 2011 as an additional control but finds no effect on 2012 insurance demand. Recall from Table 2.9 that households who had received a subsidy in 2011 showed a much better understanding of the insurance product. Again, this seems to run against the idea that lack of knowledge about insurance stood behind the weak demand observed. In addition, this result does not seem to support the existence of a discouragement (encouragement) or a price anchoring effect of having received a higher (lower) discount in 2011 than in 2012.[16] In sum, the results suggest that insurance literacy training and subsidies have an immediate, but not sustained, effect on demand.

In Table 2.12 we test the correlation between an individual's experience with weather insurance in 2011 and demand in 2012. We find that prior experience with insurance is not, by itself, a strong predictor of demand. However, while purchasing insurance does not, on its own, have a substantial impact on demand, purchasing insurance *and* receiving a payout is strongly positively correlated with the decision to purchase insurance in the subsequent season.[17] Moreover, observing other households in the village receiving a payout has no significant effect on demand.[18] This pattern closely resembles that found by Stein (2011), and is in contrast to the work of Cole, Stein, and Tobacman (2014) –who find that experiencing a payout in the village is the only relevant predictor of future demand, regardless of individually receiving a payout or not--, and Karlan et al. (2012), who find individual as well as social network spillover effects. In any case, as we lack data on each household's social network, we cannot directly compare our results with the latter.

Finally, we present pooled results for the take-up in both seasons in Table 2.13. The resulting impacts simply reflect the average of the treatment effects over the first and second seasons. It is worth noting, however, that the interventions are still significant in the pooled results.

2.4.4 Is Demand Hump-Shaped in Risk-Aversion?

We now turn to testing the theoretical predictions outlined in Section 2.1 regarding the relationship between risk aversion and insurance demand. Specifically, we examine whether a hump-shaped relationship between demand and risk aversion exists for products priced with a multiple above 1 (that is, above the actuarially fair price) and a downward-sloping relationship for products with a multiple below 1. In our experiment, insurance

[16] This hypothesis is also rejected in alternative (not reported) specifications, such as using the difference in discount between both seasons or an indicator variable for whether the discount in 2012 was lower than in 2011.
[17] Around 10 percent (14 households) of the 2011 purchasers received a payout from the insurance product.
[18] These results are not shown.

was in most cases priced above the actuarially fair price, and so we would expect to observe demand, on average, initially increasing with risk aversion and then falling. We show our estimates for the relationship between risk aversion (measured through a hypothetical Binswanger lottery survey question) and demand for insurance in Table 2.14.[19] Although we do not find a significant difference in demand across different levels of risk aversion, when we graph the point estimates, we do observe the predicted hump-shaped demand for index insurance (as depicted in figure 5).[20]

Table 2.12 — 2012 Uptake Among Sampled Households, Including 2011 Uptake and Payouts

	(1)	(2)	(3)	(4)	(5)
	Logit	IV Logit	Logit	IV Logit	Logit
Log (price)	–0.024**	–0.025	–0.022*	–0.023	-0.007
	(0.012)	(0.016)	(0.013)	(0.019)	(0.024)
Bought insurance in 2011	0.016	0.019	0.015	0.017	0.048**
	(0.013)	(0.038)	(0.013)	(0.056)	(0.022)
Had a payout in 2011	0.073***	0.071**	0.071***	0.070	0.062*
	(0.018)	(0.036)	(0.019)	(0.050)	(0.033)
Intensive	0.005	0.005			
	(0.012)	(0.013)			
Log (distance)	–0.007**	0.000	–0.005	0.000	
	(0.004)	(0.016)	(0.003)	(0.017)	
Log (distance) x Log (price)			(+)*	(+)	
Station is close					0.007
					(0.018)
Station is close x Log (price)					(–)**
					(0.048)
Sample	Full	Full	Full	Full	Far and close
Observations	2,183	2,183	2,183	2,183	932

Notes: Average marginal effects are reported. IV specifications instrument log of distance with assignment to a new weather station. In columns (3), (4), and (5), as marginal effects are not defined for the interaction term, we show the sign and significance for the parameter in the logit equation. Standard errors for the marginal effects, clustered at the village level, are in parentheses. Standard errors are bootstrapped in IV specifications. *** p < 0.01, ** p < 0.05, * p < 0.1. Source: Administrative sales data.

[19] We calculate a range for the implicit risk aversion coefficients given a respondent's answer assuming a CRRA function. In this way, the upper (lower) bound of the range is the relative risk aversion parameter that makes the selected option equivalent (in a utility sense) to the immediate less (more) riskier one. We then use the geometric average between these two bounds in the graphs. In the case of the riskiest choice (number 5), we use the coefficient from the lower bound.

[20] In all figures, we plot a spline-smoothed curve across the implicit coefficients of relative risk aversion, calculated as explained above.

Table 2.13 — Pooled Results

	(1) Logit	(2) IV Logit
Log (price)	–0.073***	–0.073***
	(0.015)	(0.017)
Log (distance to weather station)	–0.010***	–0.020
	(0.003)	(0.017)
Intensive	0.029*	0.029*
	(0.015)	(0.016)
2012 season	–0.073***	–0.073***
	(0.015)	(0.017)
Observations	4,366	4,366

Notes: Average marginal effects are reported. IV specifications instrument log of distance with assignment to a new weather station. Standard errors for the marginal effects, clustered at the village level, are in parentheses. Standard errors are bootstrapped in IV specifications. *** p < 0.01, ** p < 0.05, * p < 0.1. Source: Administrative sales data.

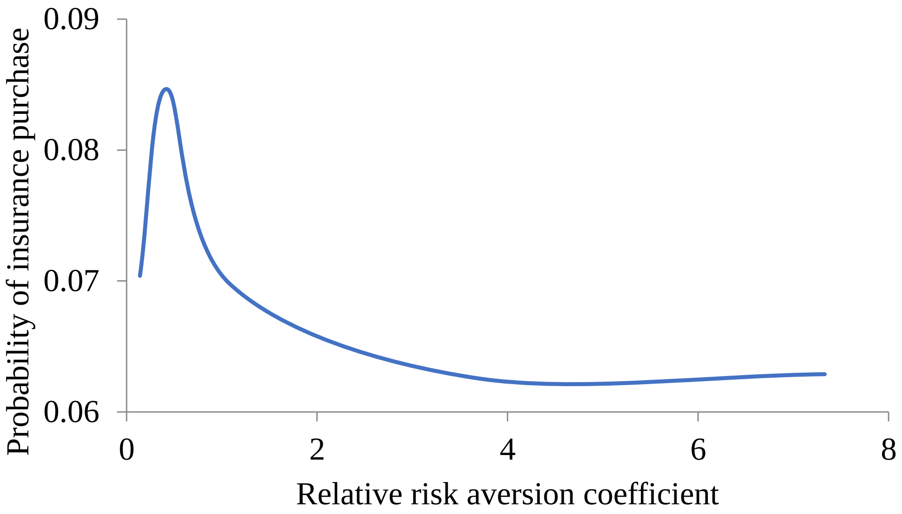


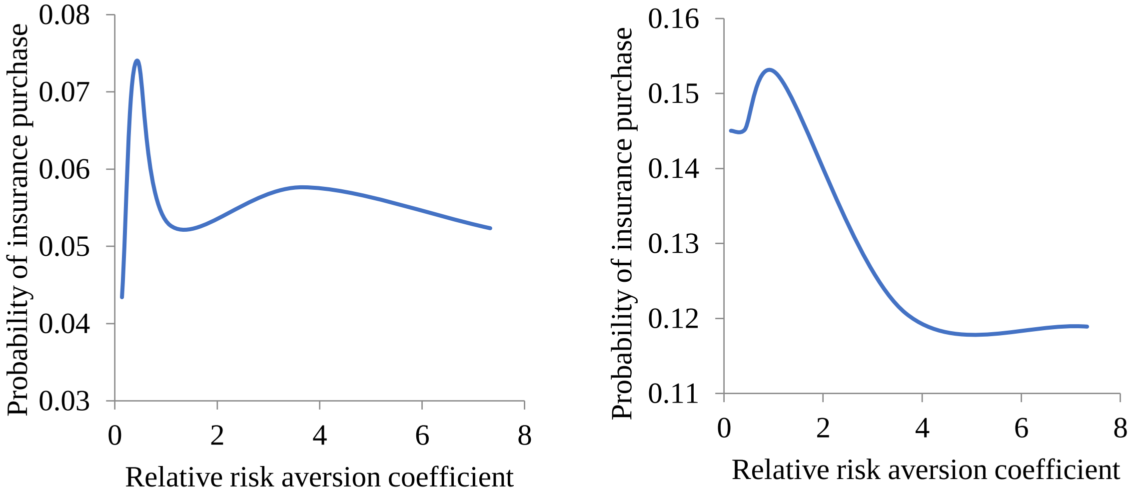
Figure 2.5 — Probability of purchase against coefficients of relative risk aversion,

Using historical weather data, we estimate the actuarially fair price of the insurance contracts that were offered. In Ujjain and Dewas, households receiving the two highest discount values faced an insurance contract that was actuarially favorable (with a multiple less than 1). This was also the case for those households in Bhopal who received the highest discount voucher. We then separately estimate the relationship between demand and risk aversion for the subgroups of households that faced a favorable (multiple below 1) or an unfavorable (multiple above 1) insurance contract. Results are presented in columns 2 and 3 of Table 2.14. Again, we see no significant difference in demand across risk aversion, but graphing the results is still instructive (figure 6). We observe the predicted hump-shaped demand for households facing an insurance contract priced at a multiple higher than 1 (left panel) and the predicted downward-sloping demand curve for households facing an insurance contract priced below the actuarially fair level (right panel). However, there is a kink in the demand curve for households with low levels of risk aversion that is not predicted by the theory.

Table 2.14 — Testing the Relationship between Risk Aversion, Price, and Demand

Dependent variable is 2011 uptake	(1)	(2)	(3)	(4)
Risk choice 1	0.008	–0.009	0.026	
(least risk averse)	(0.021)	(0.022)	(0.039)	
Risk choice 2	0.022	0.022	0.026	
	(0.018)	(0.023)	(0.037)	
Risk choice 3	0.007	0.000	0.034	
	(0.017)	(0.020)	(0.036)	
Risk choice 4	–0.000	0.005	0.002	
	(0.014)	(0.017)	(0.039)	
Less risk averse				0.010
				(0.010)
Log (price)				–0.136***
				(0.025)
Less risk averse x Log (price)				(–)
Sample	Full	Multiple above 1	Multiple below 1	Full
Observations	2,180	1,354	551	2,183

Notes: Average marginal effects are reported. In column (4), as marginal effects are not defined for the interaction term, we show the sign and significance for the parameter in the logit equation. Standard errors for the marginal effects, clustered at the village level, are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Source: Administrative sales data and household survey data.



Panel A. Multiple higher than 1 Panel B. Multiple lower than 1

Figure 2.6 — Probability of purchase against coefficients of relative risk aversion

Overall, results appear consistent with theoretical predictions; although, arguably because of limited power, no significant trend across levels of risk aversion is found. If these theoretical predictions were to hold true, we would also expect the price elasticity of demand for insurance to be higher among those who are less risk averse than among those who are more risk averse. This is potentially a test with higher power. We show this exercise in column 4 of Table 2.14 and do not find a significant difference between the price elasticity of demand for those who are more and less risk averse.

2.5 Conclusions

This chapter presents causal evidence on three factors that affect take-up of rainfall-based weather hedges in India: price, distance to the reference weather station (a proxy for basis risk), and insurance literacy.

We link our empirical analysis to an expected-utility theoretical framework of the demand for insurance under the presence of basis risk. In line with the predictions of the model, demand for the insurance products offered to farmers (a series of weather hedges) is decreasing in price and basis risk. These effects are robust and significant. In addition, we explicitly test a theoretical prediction from Clarke (2016) for insurance products with basis risk, where demand increases with aversion at low levels of risk aversion, while it decreases at higher levels. Empirically, we find some evidence for this relationship, though the results are not strong, arguably due to a lack of power.

We also find that demand increases as product comprehension increases. This is an important finding given that these weather hedges are being offered to farmers who have limited experience with formal financial products, and certainly not with products as complex as a weather hedge. However, while both price and investment in new weather stations (as a means to reduce the extent of basis risk) are fairly effective in encouraging future demand for the product, insurance literary training seems to be of a more transient nature, with no significant impact on understanding or demand after the first year of its implementation. Price discounts had a much stronger effect on understanding, consistent with a model of learning by doing. We also find that a prior positive experience with the product—as captured by having purchased insurance *and* having received a payout during the first season—significantly encourages uptake in subsequent seasons. This could also be explained by low levels of trust in the product or the insurance company. This is an interesting avenue for future research.

The results of this study suggest that the price and basis risk of these products are key drivers of demand and that weather hedges will prove to be a useful tool for farmers only if the price and basis risk associated with the products are substantially reduced. However, it is important to consider the size of investment needed to allow well-priced, low-basis risk products to be available. Although we cannot truly compute a measure of the cost-benefit of price discounts and weather station infrastructure until we assess the (largely unknown) benefits of buying insurance and any other spillovers these interventions may bring about, we can compare the cost of each in increasing uptake by 10 percentage points.

In 2011, the cost of distributing price discounts was around US$2.97 per capita. From the estimations in Table 2.6, a discount of about Rs.135 (about $2.7) is needed to increase average take-up rates by 10 percentage points. This amounts to a total of, roughly, $5 per capita to obtain the same response in demand that can be obtained through intensive training for $20. In 2012 discounts were implemented at the village level, which basically eliminated discount distribution costs, although the price effect was weaker with a Rs.180 ($3.6) discount needed to encourage an increase in uptake of 10 percentage points. This compares favorably with the cost of insurance literacy training, which increased demand by 5 percentage points in 2011 at a per capita cost of $10.40.

In our sample, installing a new weather station increased take-up by almost 5 percentage points as a result of the increased proximity to the trigger station afforded to households in nearby villages. Installing two new stations would increase take-up rates by 10 percentage points. The cost of installing two new stations was $13.34 per household serviced, but in reality this cost can be spread across more households and across multiple years, given installing a new weather station not only encourages take-up in the year of installation but also that in subsequent years. Just spreading the installation investment of a new weather station over five years (and assuming a 20 percent interest rate) reduces the per-person annual cost to $2.23. Moreover, additional welfare benefits may stem from an insurance product with less basis risk. This is an important topic to explore in future research.

CHAPTER 3

Estimating Spatial Basis Risk in Rainfall Index Insurance: Methodology and Application to Excess Rainfall Insurance in Uruguay

Weather index insurance has gained considerable attention during the past decade as a promising instrument to increase the resilience of rural agricultural households and unleash productivity potentials in smallholder farming.[21] By allowing insurance payouts to be determined by an objective and independently measured index, this instrument allows to overcome problems of asymmetric information and high loss verification costs, argued to have limited the expansion of traditional indemnity insurance into rural areas of developing countries.[22] Despite the general enthusiasm, however, the development of index insurance markets has not been without its drawbacks. A particularly problematic element in this regard has been the overall low level of demand among farmers, which has resulted in most existing studies redirecting efforts towards analyzing demand determinants and away from other important questions such as the extent of the benefits over the production process stemming from reduced agricultural risk.[23]

Perhaps the most cited obstacle for the adoption of index insurance has been *basis risk*, or the mismatch between crop losses (a farmer's true variable of interest) and insurance payouts. Basis risk is an inherent limitation of index insurance products, arising from the inability of an index to perfectly replicate an individual farmer's losses. By reducing the effectiveness of insurance, basis risk is not only expected to negatively impact demand, but also any potential positive effects on production and welfare. While a number of

[21] Hazell et al. (2010) provide a good overview of the topic.

[22] See Hazell, Pomareda, and Valdés (1986) for a comprehensive treatment.

[23] Cole et al. (2012) discuss the lack of demand for insurance across a number of pilot projects. A number of studies analyze the determinants of demand, such as Chapter 2 of the present dissertation; Cole et al. (2013); or Giné, Townsend, and Vickery (2008). For a few studies analyzing the impacts of index insurance on agricultural production see Cai (2016); Cole, Giné, and Vickery (2013); and Mobarak and Rosenzweig (2012).

innovative approaches have been suggested to minimize its consequences, basis risk still constitutes one of the most discussed issues around the general index insurance debate.[24]

Surprisingly, the ubiquity of basis risk in the index insurance literature and debates has not been matched with a concerted effort to understand and analyze its full extent and characteristics, and studies directly tackling the subject are scarce. Most of the attention has been centered on spatial (or geographic) basis risk—the fraction of basis risk arising from measuring an insurance's index at a physical location other than the farmer's plot.[25] While, to our knowledge, there are no available estimations of the relative importance of the different components of basis risk, spatial variation in weather is perhaps the most salient negative feature of an index insurance product from a farmer's perspective. For instance, a small number of studies in developing countries have shown that smallholder farmers seem to strongly react to the distance at which the insurance's index is measured, suggesting a good understanding of the properties and downsides of this component.[26] Nevertheless, distance to the weather station can only serve as a rough proxy for spatial basis risk and almost no efforts have been conducted to properly quantify the latter.[27]

A standard exercise commonly found in the literature has been to estimate rainfall decorrelation functions.[28] A typical application would take the following form: (1) calculate Pearson correlation coefficients between daily rainfall amounts at every pair of available weather stations; (2) link these with geographic distance through the estimation of a (parametric or non-parametric) smoothing function. While this type of exercise can provide a rough approximation of the overall spatial dependence in rainfall, it suffers from a number of limitations with regard to the estimation of spatial basis risk. First, correlation can be used as an appropriate measure to measure dependence only if the underlying bivariate distribution of two sites is elliptical in nature (see, for instance, Embrechts, McNeil, and Straumann, 2002). Unfortunately, such an assumption seems not to be appropriate for daily rainfall (see Sections 3.1 and 3.4.2 below). Second, given that, by definition, insurance products relate to the occurrence of extreme risk events, the relevant

[24] See Barrett et al. (2007), for example. Other more recent innovations are the so-called double-trigger insurance (Elabed et al., 2013) or gap insurance (Dercon et al., 2014), among others.

[25] A full characterization of the different components of basis risk is offered in Section 3.2.

[26] See Chapter 2, Hill, Hoddinott, and Kumar (2013), and Mobarak and Rosenzweig (2012). Jensen, Mude, and Barrett (2014) explain demand through a more comprehensive measure of basis risk that relies on the availability of longitudinal household loss data.

[27] One exception seems to be Kost et al. (2012), who estimate basis risk relying on both a crop model and a weather simulator. Contrary to our methodology, however, they do not characterize the statistical relationship between weather at different sites nor contemplate the regular joint occurrence of extreme weather events. Moreover, they do not provide any sense of the variability of their estimates to the underlying uncertainty of the (crop and weather) models they rely on for their analysis.

[28] See, for example, Wang and Zhang (2003), Ciach and Krajewski (2006), and Odening, Musshoff, and Xu (2007).

dependence between sites is that at the *tails* of the distributional support.[29] Third, insurance indices are usually *functions* of daily rainfall, sometimes defined over the course of an entire month or season. An appropriate correlation measure should thus be obtained for the indices themselves and not for daily observations, as dependence varies with aggregation.[30] However, the limited availability of years from which to observe annual index realizations can greatly affect the precision of these estimates.

In contrast, this chapter develops a novel methodology to quantify the extent and analyze the properties of spatial basis risk for any given rainfall index insurance product that can account for the above issues. The methodology is applied to an index insurance product covering against excess rainfall in Uruguay. We describe the results in terms of two main sources of spatial variation: distance and direction to the reference weather station. The latter, while highly context-dependent, has been largely ignored by the existing index insurance literature. We compare the model's results to a theoretical upper bound for basis risk derived by Clarke (2016) and to farmers' perceptions on geographic variation in rainfall, elicited during the insurance project's baseline survey.

The proposed methodology builds upon a multisite stochastic rainfall generator model standard in the hydrological literature, pioneered by Todorovic and Woolhiser (1975) and Katz (1977). In particular, the bivariate version of this model (Wilks, 1998) regards rainfall at two different sites as arising from two separate components: rainfall *occurrence* and rainfall *amount*. Rainfall *occurrence* is modeled as two correlated Markov chain processes (one per site); while rainfall *amount*—conditional on rainfall occurring—is regarded as arising from two correlated continuous univariate distributions.

We extend this model through the use of a copula, an increasingly popular instrument that allows us to model non-traditional dependence structures between the marginal rainfall amount processes at both sites.[31] This is particularly important, as rainfall amounts at two nearby sites generally exhibit so-called upper tail dependence, or a tendency of *extreme* rainfall realizations to jointly occur at both sites.[32] Such a characteristic is a consequence of occasional large-scale precipitation systems, which can cover large geographical areas with abundant precipitation, and we indeed find considerable evidence for it in our historical rainfall data. The ability to capture this type of dependence is of crucial importance for determining spatial basis risk of insurance products, since these instruments link payouts to the occurrence of extreme weather events, precisely those at

[29] Interestingly, the dependence at the tails of the support of a bivariate normal distribution converges to zero as one moves further into the tails, regardless of the degree of overall correlation.
[30] See Ciach and Krajewski (2006).
[31] Other studies that have explore the coupling of stochastic rainfall simulation models with copulas are Serinaldi (2009b) and Serinaldi and Kilsby (2014).
[32] Kost et al. (2012) comment the following on the same topic: "*When extreme events occur, such as the high rainfall totals in 2010 or a drought, farmers perceive that all locations are affected similarly.*" We provide further indicative evidence based on farmers' perceptions in Section 4.1 below.

the tail of their distributions. While these considerations have long been recognized by some strands of the insurance literature,[33] concerned about the sustainability of insurance supply and the necessary capital reserves required to face the potential of rare though extreme events, a large amount of studies have failed to incorporate this important issue.[34]

In order to calibrate the rainfall model we rely on 30 years of historical daily rainfall data from the national meteorological network. This dataset, though geographically sparse, is crucial to estimate the long-run statistical properties of rainfall at a single site. To obtain reasonable estimates for the degree of spatial dependence between sites, we augment the former with data from a unique, dense network of 34 rainfall gauges uniformly distributed around the study area and from a smaller network of gauges used as reference for the insurance product. With all the components in place we are now able to obtain a fully-specified bivariate rainfall distribution for an arbitrary plot location and its corresponding reference gauge, achieved by interpolating the calibrated model's parameters. Finally, spatial basis risk measures can be obtained from the model through Monte-Carlo simulations.

Of course, this same measure could be calculated directly from historical data. However, there are a number of drawbacks related to this. As the above measures must be calculated from complete index realizations and indices for insurance products are generally defined over the course of a season or a year, this implies working with between 20 and 50 observations (years) under most data availability scenarios. This is independent of the frequency at which the underlying data is available (hourly, daily, etc.). Moreover, as index insurance triggers are typically defined at very large (or low) extreme quantiles, and basis risk is only relevant around these triggers, the available number of observations to calculate these measures reduces even further. For instance, calculating basis risk for an insurance product that pays when the realized index is above its 95th percentile would imply observing a loss/payment once every 20 years. With 50 years of data, this would result in only around 2-3 years where to assess the degree of basis risk of the product. In contrast, thanks to (and at the cost of) imposing a specific distributional and dependence structure to the occurrence of rainfall at multiple sites, our methodology allows estimating these probabilities much more precisely than would otherwise be possible through simply simulating from the calibrated bivariate model an arbitrary number of times.

---

[33] See, for example, Okhrin, Odening, and Xu (2013) or Goodwin and Hungerford (2015) for two applications close to index insurance.

[34] For instance, Wang and Zhang (2003) consider the important topic of risk pooling in relation to sustainability in the U.S. agricultural insurance context. Their analysis, though, is based on Pearson correlations, a measure which is appropriate only when the underlying joint distribution is bivariate normal. Moreover, numerous studies from the finance literature still rely on value-at-risk measures based on normal distributions, even though the incorporation of non-classical association patterns through copulas has been much more widely incorporated (see, for example, Junker and May, 2005). In this regard, the methodology considered in this chapter can also be extended to account for risk in these other realms.

The results of this exercise are as follows. Spatial basis risk for the index insurance product marketed in Uruguay is not negligible. In particular, and depending on the farmer's location, basis risk is such that the insurance product would fail to pay between 1 to 5 times out of 10 in which a farmer were to experience critical crop losses. It is worth noting, though, that estimated basis risk for all farmers still lies in a range within which the theoretical model would predict positive demand from sufficiently risk-averse individuals. Variation of basis risk between farmers is mostly determined by the distance at which a farmer is located from the insurance reference weather station, this relationship being positive and concave. The latter property implies that, for instance, while basis risk doubles at short distances to the weather station (from around 10% at 1 kms. to around 20% at 10 kms.), it takes a much longer difference in distance to double again (around 40% at 40 kms.). Interestingly, however, farmers perceive similarities in rainfall pattterms to decrease much more rapidly. Based on farmer's survey answers, an average farmer considers rainfall patterns in two sites 10 kms. apart as being *similar*, while rainfall patterns at 40 kms. or more are considered to be *very different*. Finally, the results point to the importance of taking into consideration geographic variation in precipitation patterns—even within relatively small regions—when designing an index insurance product. This element is shown to considerably increase (or decrease) the degree of spatial basis risk, depending on the exact location of a farmer's plot and its insurance reference weather station.

The chapter contributes to the literature in a number of ways. First, by extending the model of Wilks (1998) through the use of copulas, our methodology describes a weather generator model through which to characterize the spatial properties of *extreme* rainfall in a given region. Second, we provide a framework to estimate the degree of spatial basis risk for an arbitrary rainfall index insurance product, which should serve to encourage better ex-ante assessments of future products. Third, our application enriches the broader index insurance debate by presenting the first direct empirical exploration of spatial basis risk, relying on an appropriate operational definition (beyond correlation in rainfall) and pointing to the importance of a directional element, generally disregarded as a relevant factor. Finally, the study contributes additional evidence on behavioral frictions in the insurance market by indicating a relative gap between the real and perceived extents of spatial basis risk. Given the low observed demand for index insurance products, closing this gap through appropriately targeted information seems an important consideration for the future development of the market.

The chapter proceeds as follows. Section 3.1 describes the general structure of the stochastic rainfall generator model and its extension to account for flexible dependence patterns through the use of copulas. Section 3.2 discusses in more detail the concept of basis risk, in addition to its empirical measurement and its theoretical upper bound. Sections 3.3 and 3.4 discuss, respectively, the general context in which we apply the methodology and the empirical calibration of the rainfall model. Section 3.5 presents the results and Section 3.6 concludes.

3.1 Precipitation Model and Extension through Copulas

Stochastic precipitation generator models have a long tradition in the hydrological literature. On one hand, they can be regarded as statistical representations of rainfall at a particular geographic area, which can be used to understand and analyze spatial patterns of local or regional precipitation systems. On the other hand, they can simulate long synthetic sequences of precipitation at various points in space, which can be later used as inputs for various applications related to risk and reliability assessment of agricultural and water resource systems (see Serinaldi and Kilsby, 2014; Ailliot et al., 2015; and references therein).

While a number of alternative stochastic precipitation generator models are available—each exhibiting different characteristics and better suited at tackling different types of questions or climatic environments—we will rely on a widely-used version: the chain-dependent stochastic model (Todorovic and Woolhiser, 1975; Katz, 1977; Wilks, 1998), also known as Wilks model. One of the advantages of this model is that it consists of an intuitive representation of daily rainfall, with one model component defining the binary occurrence process (rain or no rain) and a separate component driving the rainfall amount process, conditional on rainfall occurring. Another advantage of this model is that it can be calibrated using daily rainfall series at multiple sites, in contrast to other models that require more disaggregated rainfall observations (both temporally and spatially) or more complex climatological specifications. Finally, the model allows for interpolation of distributional and dependence parameters across space, which will be crucial to characterize basis risk at arbitrary locations where rainfall gauges are not available. As our application is targeted at insurance and other risk practitioners, we believe that the relative simplicity and flexibility of the chain-dependent stochastic model is attractive, while the estimation of more sophisticated or niche models seems unwarranted.

It has been shown that the synthetic values arising from such a model are able to reproduce the most relevant statistical properties of daily rainfall fairly accurately (Wilks, 1998).[35] Moreover, Mehrotra, Srikanthan, and Sharma (2006) compare three popular stochastic precipitation generators and find that the chain-dependent stochastic precipitation model offers the best overall performance, with the added advantage of its simple structure. This model has been used in a variety of applications, including

[35] However, it is interesting to notice that this model may not accomodate all situations equally. For example, the literature has noted that this model cannot reproduce the statistical properties of length of dry spells as accurately as other aspects of the distribution (Buishand, 1978; Racsko, Szeidl, and Semenov, 1991; Lettenmaier, 1995). These cases can be handled by allowing for a higher-order markov chain driving the occurrence process or by alternative stochastic weather models. Nevertheless, the current methodology would generally be adaptable to these alternative specifications. In addition, the model has been shown to be less accurate at preserving statistical characteristics at higher time scales than the one used for its calibration (i.e. monthly and annual time scales for a daily model). See, for instance, Srikanthan and Pegram (2007) and Pegram (2009), and references therein.

applications to finance and insurance (see, for example, Cao, Li, and Wei, 2004; Lopez Cabrera, Odening, and Ritter, 2013; and Ritter, Musshoff, and Odening, 2014).

One potential shortcoming with the application of this model is known as *low variability bias* (see Dubrovský, Buchtele, and Žalud, 2004; Odening, Musshoff, and Xu, 2007), or the model's inaccuracy at reproducing statistical properties of rainfall at lower temporal frequencies, such as monthly or annual.[36] However, since the object of our analysis is an insurance product linked to cumulative rainfall over 10 consecutive days only, we believe low variability bias not to be a first-order concern in relation to the numerous issues that would arise from working with an alternative model. Moreover, basis risk—the primary element of interest—is arguably related more to the modelling of the spatial dependence pattern of rainfall than to the accurate characterization of rainfall at a single site.

Below we describe the basic structure of the chain-dependent stochastic precipitation model together with the proposed extension in the dependence structure of the amount process and provide a brief introductory description of the concept of copulas.

3.1.1 Chain-dependent stochastic precipitation model

In its most simple bivariate form, the model decomposes precipitation at each of two sites into an *occurrence* component and an *amount* component.

The (binary) *occurrence* of rainfall at each site is regarded as arising from a first-order Markov chain. Spatial correlation between rainfall *occurrence* at both sites is modeled by allowing the random realizations driving each Markov Chain to arise jointly from an underlying bivariate normal distribution with a certain correlation coefficient.

Conditional on rainfall occurring at a site, the precipitation *amount* process (i.e. number of millimeters over a certain time period, for instance a day) is modeled as arising from a given univariate distribution function, most commonly gamma or double-exponential. Correlation between rainfall amounts is introduced through a second, underlying bivariate normal process.

All parameters—Markov chain probabilities, parameters of the amount distributions, and correlation coefficients for the underlying normal distributions driving the dependence between amount and occurrence processes—are then calibrated from existing data. Section 3.4 below provides more details on each of the components of the model, clearly illustrating how they all fit together. Once calibrated, the model is able to simulate a series of correlated precipitation data at both sites through Monte Carlo simulations.

[36] In some strands of the literature this feature is also known as *overdispersion*.

3.1.2 Extension

Now, a limitation of the model is the assumed normality of the joint process driving the association between rainfall amounts at the two sites, which depends on just one correlation coefficient to describe the entire dependence structure. It has long been recognized that linear correlation is not appropriate to model joint dependence of real-world variables (Blyth, 1996; Shaw, 1997). For instance, this measure cannot capture underlying non-linear relationships between variables and it is not invariant under monotonic transformations. Embrechts, McNeil, and Straumann (2002) provide a lucid discussion about the risks of using linear correlation to summarize dependence of risks beyond the particular case of elliptical distributions.

An interesting aspect of the association between two random variables is the degree of dependence in the occurrence of extreme values, known as upper or lower tail dependence in accordance to whether this dependence takes place in the upper or lower tail of the distribution, respectively. This aspect of joint distributions is particularly relevant in the case of insurance products. For example, large scale natural disasters or large terrorist attacks such as 9/11 may affect multiple exposures of an insurance company at the same time (Kousky and Cooke, 2009). It is important to note that tail dependence in a joint distribution is independent of the existence of fat tails in the marginal distributions. Moreover, it does not share a direct relationship with linear correlation, as indicated by the fact that a bivariate normal distribution can accommodate any level of (negative or positive) correlation but at the same time always exhibits zero tail dependence.

As geographic basis risk relates to the probability of an index being above or below a certain (extreme) trigger at two separate locations, tail dependence seems a natural property to consider in this context. In particular, as previously discussed, we wish to allow for positive upper tail dependence in the dependence process for rainfall amounts at different locations. We thus propose to extend the precipitation model above by introducing a bivariate copula as the driver of the spatial association between the *wet* part of the rainfall distribution at two sites. A natural way to achieve this is through the use of a copula, a flexible instrument that allows for a number of alternative dependence structures beyond the one implicitly imposed by the bivariate normal.

Next, we briefly review the concept of copulas and describe alternative copula family candidates that may accommodate the association patterns most commonly found in the case of rainfall realizations.

3.1.3 Copulas

The word copula was first employed in a mathematical or statistical sense by Sklar (1959) in the theorem (which now bears his name) describing the functions that “join together” one-dimensional distribution functions to form multivariate distribution functions (Nelsen, 2006). Sklar’s theorem states that *any* multivariate distribution can be represented by an appropriate copula function with arguments consisting of all the marginal cumulative distribution functions.[37] In particular, a 2-dimensional copula is a function C that maps values in the unit hypercube $[0, 1]^2$ to values in the unit interval [0, 1] such that:

$$F(X_1 = x_1, X_2 = x_2) = C(F_1(x_1), F_2(x_2)) \tag{3.1}$$

Where $F(.)$ is the bivariate cumulative distribution function (CDF) for random variables X_1 and X_2, and $F_1(.)$ and $F_2(.)$ are the univariate CDFs for X_1 and X_2, respectively. Alternatively, a copula may be regarded as a multivariate distribution function with standard uniform marginal distributions. For a formal treatment of the topic, see Joe (1997) and Nelsen (2006). Danaher and Smith (2011) provide an accessible introduction with an application to marketing.

During the past 20 years, copulas have been increasingly used in a number of contexts where the joint distribution of two or more variables is a central topic of interest. Apart from the statistical literature, where copulas have existed for several decades, one of the main areas where copulas have been practically applied has been in finance, as an instrument to jointly model asset prices and other financial variables of interest (see, for instance, Kharoubi-Rakotomalala and Maurer, 2013; Patton, 2012; and references therein). Another common area of application has been the marketing literature (e.g. Danaher and Smith, 2011). Finally, copulas have been widely applied in hidrology and biostatistics (Bárdossy, 2006; Genest and Favre, 2007). Of particular interest to this chapter, a few studies have approached joint modeling of weather outcomes through copulas (Serinaldi, 2009a; Serinaldi, 2009b; Bárdossy and Pegram, 2009; Serinaldi and Kilsby, 2014).

Modelling the dependence between two random variables through a copula has a number of advantages. First, copulas allow to model the dependence between two variables with different underlying marginal distributions. For instance, by using a copula it is possible to jointly model a discrete and a continuous variable (such scenarios are common in the marketing literature, see Danaher and Smith, 2011). Second, copulas can allow for a very flexible dependence structure between two variables, which can fit nonlinearities in association and different degrees of dependence along different regions of the support of the distributions. One important such application is the modelling of tail dependence.

As discussed above, tail dependence relates to the degree of association between two random variables at the (upper or lower) tail of their joint support. This is in principle independent of the overall degree of association (and, in particular, of the level of linear correlation) between them.[38] In particular, let X_1 and X_2 be random variables with

[37] Of course, a number of technical requirements need to be fulfilled for a function to be considered a copula. See Joe (1997) and Nelsen (2006) for further details.

[38] In the case of parametric families of bivariate distributions, however, tail dependence and different measures of correlation will be related to each other.

distribution functions F_1 and F_2, respectively. Upper tail dependence (λ_U) is defined as the probability of X_1 being greater than its t-th percentile, given that X_2 is greater than its t-th percentile, as t approaches 100 (Nelsen, 2006):[39]

$$\lambda_U = \lim_{t \to 100} Prob\left(X_1 > F_1^{-1}\left(\frac{t}{100}\right) \middle| X_2 > F_2^{-1}\left(\frac{t}{100}\right)\right) \tag{3.2}$$

In the case of copulas, upper and lower tail dependence are inherent properties of any copula family. Thus, copulas can be classified by their degree of tail dependence. For instance, as mentioned above, the Gaussian copula exhibits zero (upper or lower) tail dependence (asymptotic independence).

In this chapter we consider the Gumbel copula (Gumbel, 1960), a well-known and widely used copula family belonging to the Archimedean class. It takes the following form:

$$C(u, v; \theta) = exp\left(-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{1/\theta}\right) \tag{3.3}$$

This copula exhibits weak lower tail dependence but strong upper tail dependence. These properties make the Gumbel copula an appropriate modelling choice when two outcomes are likely to simultaneously realize upper tail values (Trivedi and Zimmer, 2007). As we show below, this is exactly the type of dependence commonly observed between rainfall amounts at two nearby sites (conditional on both sites experiencing rainfall).

3.2 Basis Risk: Definition, Measure, and Theoretical Upper Bound

This section defines basis risk in the context of our application and describes the measure chosen for the analysis. The advantages of this measure are both its intuitive appeal and its connection to a theoretical model of a farmer's demand for index insurance.

In the context of weather index insurance products targeted at smallholder farmers, basis risk represents a situation where the insurance's payout does not perfectly correspond to the insured farmer's crop losses. The sources of this dissociation, however, can be manifold. For instance, a farmer may experience a loss due to an inadequate handling of the plot, or due to an unexpected pest or disease. A broad characterization of the different components of basis risk can be thought of as follows: (i) *spatial*, related to geographic variation of the weather index between a farmer's plot and the insurance's reference weather station; (ii) *design*, related to the design of the insurance product, including the choice of weather variable, the specific index for the chosen weather variable, and the functional form of payouts; (iii) *other covariate risks*, related to other common factors affecting farmers in a specific area (e.g. pests), and (iv) *idiosyncratic risk*, related to

[39] Lower tail dependence (λ_L) is similarly defined as the probability of X_1 being lower than its t-th percentile, given that X_2 is lower than its t-th percentile, as t approaches 0.

specific farming practices, and other plot and farmer's characteristics (e.g., soil quality, input use, etc.).[40]

Components (ii) to (iv) depend heavily on features of the crop production function of a farmer, which can only be accurately estimated through sufficient longitudinal farm-level production data. As we do not count with such data, we are only able to focus on the *spatial* component of basis risk, implicitly assuming that experiencing a loss can be accurately captured by the index at the farmer's plot being above the insurance's trigger. As such, this measure will typically reflect a lower bound for a product's overall degree of basis risk. To be fair, though, an appropriate benchmark for basis risk should consist of an indemnity insurance product that insures against the *specific risk* that the index insurance product is designed to address, in our context excess of rainfall. In this case, then, only components (i) and (ii) would represent relevant aspects of the imperfect insurance provided by index products. It is interesting to note that data from the baseline survey seems to provide some support for the higher importance of the geographic component of basis risk: out of the 335 farmers who expressed an initial lack of interest in the rainfall index insurance product offered to them (about 48% of the total surveyed farmers), only 69 raised some aspect of basis risk as a reason,[41] with most of these (58 farmers, or 84%) mentioning distance to the reference weather station as their main concern (compared to only 11 raising other aspects of basis risk as being important).

Clarke (2016) defines basis risk as the unconditional probability of *both* experiencing a critical loss and not receiving an insurance payout (downside basis risk) or as the unconditional probability of *both* not experiencing a critical loss and receiving an insurance payout (upside basis risk).[42] Clarke (2016) calls this measure r and derives a theoretical upper bound beyond which an individual's demand for insurance would be expected to be zero (regardless of their underlying utility function), as follows: $r < p(1-q)$; where p and q represent, respectively, the farmer's critical loss probability and the probability of receiving an insurance payout. Intuitively, this upper limit reflects the fact that an insurance product needs to give a sufficiently strong signal about the farmer's losses so as to provide some insurance benefit (see Clarke, 2016; and Chapter 2 for further discussion).

Throughout the rest of the chapter we focus on downside basis risk, given that it represents the main object of interest for a farmer buying an insurance product. Moreover,

[40] Other characterizations can be found in the literature, as this changes according to the specific objectives of the study and features of the insurance product at hand, see, for example, Elabed et al. (2013), or Jensen, Mude, and Barrett (2014).

[41] The remaining fraction of farmers expressed other concerns, such as (in order of importance): excess rainfall not being a relevant risk for their crops (39%), lack of understanding of the product (13%), and a range of other minor issues unrelated to basis risk (27%).

[42] His model incorporates both measures as being symmetric, though in reality these do not need be.

for our estimates of downside basis risk, we rely on a more intuitive measure related to Clarke's that is commonly used in the literature (see, for instance, Kellner and Gatzert, 2013). The measure represents the probability of mismatch between losses and payouts (as above), *conditional* on the farmer experiencing a critical loss.[43] In other words, it can be interpreted as the number of times a given insurance product fails to pay, given that a farmer has indeed experienced a critical crop loss.

3.3 Study Context and Data

On November 2013 an innovative weather index insurance product was launched in the department of Canelones, Uruguay, as part of a project led by the International Food Policy Research Institute (IFPRI). The aim of the product was to cover horticultural farmers against excess rainfall around harvest, which is generally associated to severe losses due to the rotting of crops and the increased difficulty to access the plots. Given the heterogeneity in crops and time of planting, the product was designed as a portfolio of independent insurance 'units'. Each unit promised to pay a fixed amount when the maximum accumulated rainfall over any 10 consecutive-day period within a calendar month (the *index*) exceeded a certain pre-determined trigger. Different products were offered for each of the months between January and April (the main horticultural harvest season), and for two severities of excess rainfall, with triggers corresponding to the 85th and 95th percentiles of the historic distribution of the index. This resulted in a total of eight insurance units, which could be freely combined by each farmer according to her own crop portfolio, timing of planting, and overall risk profile.

The department of Canelones represents the main horticultural producing region in Uruguay. It is relatively small, covering a total area of 4,536 km^2, and topographically flat to a large extent. It has a humid temperate climate with lower levels of precipitation relative to the rest of the country (up to 2,000 mm yearly, on average). It is important to note that, as rainfall is highly dependent on topography and other geographical characteristics, studies conducted in more complex settings should take this into consideration much more carefully in the modelling.

3.3.1 Survey and Farmers' Perceptions

A baseline survey covering 700 horticultural farmers was conducted in Canelones between September and December of 2013. One of the objectives of the survey was to assess farmers' perceptions about the different elements of the weather index insurance product on offer.

Table 3.1 shows survey responses to questions on spatial variability in rainfall patterns. On one hand, and as expected, farmers do recognize that the pattern of rainfall between two given sites is more dissimilar the farther away they are from each other (panel

[43] Using Clarke's notation this measure would be r/p.

A). In particular, when asked about the degree of similarity in rainfall patterns between their plot and the reference gauge closest to them, farmers located farther away from the reference gauge do indicate that precipitation patterns are more dissimilar than farmers located closer by. It is however interesting to note that there is a large heterogeneity in farmers' perception of rainfall similarity. Overall, these observations are in line with findings in the index insurance literature about geographic basis risk being inversely related to insurance demand (Hill, Hoddinott, and Kumar, 2013; Chapter 2; Mobarak and Rosenzweig, 2012).

In addition, panel B in Table 3.1 shows farmers' beliefs on the maximum distance at which a plot and a reference rain gauge can be apart from each other in order to consider their rainfall patterns as being similar. Overall, about two thirds of the respondents believe that 5 kilometers (kms.) or less is the extent to which rainfall patterns can be considered similar between two locations, while 90 percent of the sample believes that this distance should not exceed 10 kms.

Moreover, farmers' perception seem to be in line with the upper tail dependence property of the Gumbel copula discussed above. Figure 3.1 shows estimated kernel probability density functions on the perceived maximum distance at which either (i) overall rainfall patterns, and (ii) excess rainfall patterns can be considered similar between two sites. These data come from an endline survey conducted with 582 horticultural farmers during July and August 2015. The figure shows that farmers do seem to perceive *excess* rainfall patterns as being similar across broader geographic regions than overall rainfall. This supports the need for modelling spatial dependence for rainfall through more flexible models than the elliptical assumption implicit behind the popular Pearson correlation coefficient.

Finally, it is interesting to note that further unreported analyses show no evidence of perceptions about rainfall similarity depending on the direction at which the reference gauge lies with respect to the farmer's plot. This observation supports the isotropy assumption made in the following section.

Table 3.1 — Farmers' Perception of Geographic Variability in Rainfall Patterns

Panel A. Rainfall patterns between plot and reference gauge

Rainfall pattern between plot and reference gauge is	**All farmers**	**Distance to reference gauge**			
		Less than 5 kms.	**Between 5 and 10 kms.**	**Between 10 and 15 kms.**	**More than 15 kms.**
Very similar	2.9%	6.4%	1.9%	3.8%	1.2%
Similar	40.7%	53.2%	49.0%	35.9%	25.5%
Slightly similar	14.3%	13.8%	11.3%	14.7%	19.3%
Different	29.6%	19.1%	29.2%	35.3%	29.8%
Very different	9.0%	4.3%	5.8%	7.1%	18.6%
Non-response	*3.6%*	*3.2%*	*2.7%*	*3.3%*	*5.6%*

Panel B. Maximum distance for rainfall patterns to be similar

Maximum distance	**No. Of farmers**	**%**	**Cum. %**
1 km. or less	121	17.3%	17.3%
Between 2 and 4 kms.	179	25.6%	42.9%
5 kms.	163	23.3%	66.1%
Between 5 and 9 kms.	60	8.6%	74.7%
10 kms.	110	15.7%	90.4%
More than 10 kms.	67	9.6%	100.0%

Note: This table shows survey responses on farmers' perceptions about spatial variability of rainfall. Panel A shows whether rainfall patterns between the farmer's plot and the closest reference gauge are considered to be similar or different, for all farmers in the sample (column 2) and by distance categories (columns 3 through 6). Panel B shows the maximum distance at which two sites can be apart for their rainfall patterns to be considered as similar. The data come from a baseline survey conducted with 700 horticultural farmers in Canelones, Uruguay, during September and December 2013.

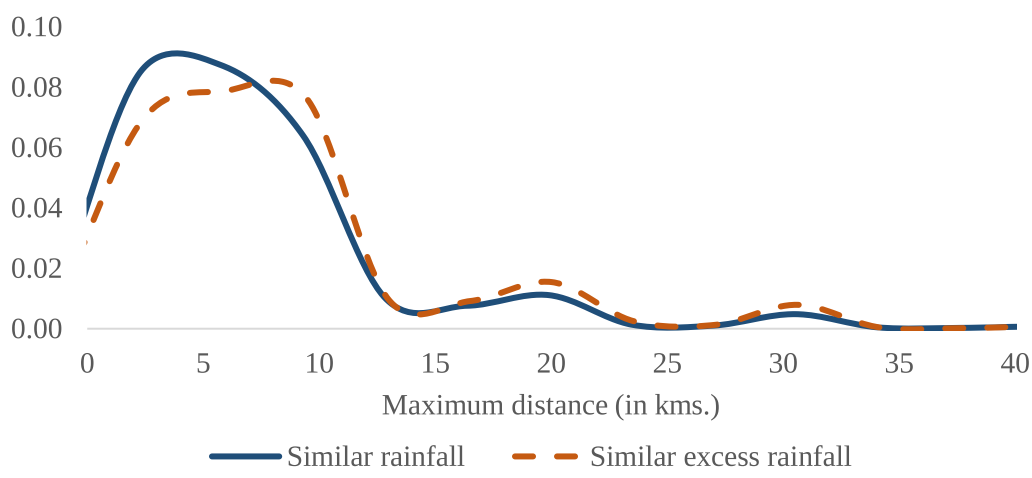


Figure 3.1 — Perception of similarity in rainfall and excess rainfall patterns

Note: This figure shows estimated probability distribution functions for the maximum distance separating two sites at which patterns of rainfall (solid line) and excess rainfall (dashed line) are considered to be similar by farmers. The curves correspond to estimated nonparametric densities using a Gaussian kernel. The data come from an endline survey conducted with 582 horticultural farmers in Canelones, Uruguay, during July and August 2015.

3.3.2 Data

For the application of our methodology we will focus on the summer agricultural season, which corresponds to the months of December, January, and February. This season is characterized by high levels of precipitation and is the main harvest season for a number of important horticultural crops such as onion, sweet potato, and tomato, among others. We work with the entire season (as opposed to individual calendar months or other agricultural time periods) for two main reasons. First, the Uruguayan meteorological institute (Instituto Uruguayo de Meteorología, INUMET) considers these three months as climatologically similar, grouping them into the same single season for meterological purposes. Second, working with a three-month period allows us to estimate distributional and dependence parameters using three times the amount of data than would be possible with monthly periods, thus benefiting from greater precision. In the sections that follow, then, all parameter estimates (and the subsequent simulations stemming from these) will be based on daily data corresponding to the full summer season.

We use a number of data sources for the analysis. First, we count with historical daily data from the entire rain gauge network of the Uruguayan meteorological institute (Instituto Uruguayo de Meteorología, INUMET). We refer to these gauges as *pre-existing*. While these data span over 30 years in some cases, allowing us to estimate representative long-term distributional parameters, the network is geographically sparse in relative terms. We focus on 11 gauges located on or in the vicinity of the department of Canelones, Uruguay, since this is the area where the index insurance product is being offered. Second, we use daily data from five automatic weather stations located in Canelones, installed as part of insurance project activities during December, 2013. As these are used as reference gauges for the insurance product, we refer to them as *insurance reference gauges*. Finally, we use data from a unique network of 34 weather stations, installed during October and December of 2013, throughout the entire area of Canelones, which we refer to as *monitoring gauges*. These last two sources, though short in time span, provide us with geographically dense data that allow us to explore the dependence pattern of rainfall at short distances, thus complementing the long-term but geographically-sparse data from the *pre-existing gauges*. Combining these data sources allows us to tackle a problem that would otherwise be very difficult to solve with a unique data source. Figure 3.2 shows the location of all the rainfall gauges used in the analyses. Appendix Table 3 shows summary statistics for these gauges, including start and end dates, the observed proportion of rainy days, and the average daily rainfall amount, conditional on rain occurring, over the summer months.

3.4 Estimation

This section first discusses in more detail the calibration of the chain-dependent stochastic rainfall model at each available rain gauge. Next, the section discusses the estimation of different correlation measures of rainfall in space and the calibration of the copula parameters used to model the dependence between pairs of sites. Each subsection ends with a description of the interpolation method implemented to predict the relevant distributional parameters at arbitrary points in space. The section concludes by providing an overview of all the components in the bivariate spatial rainfall generator.

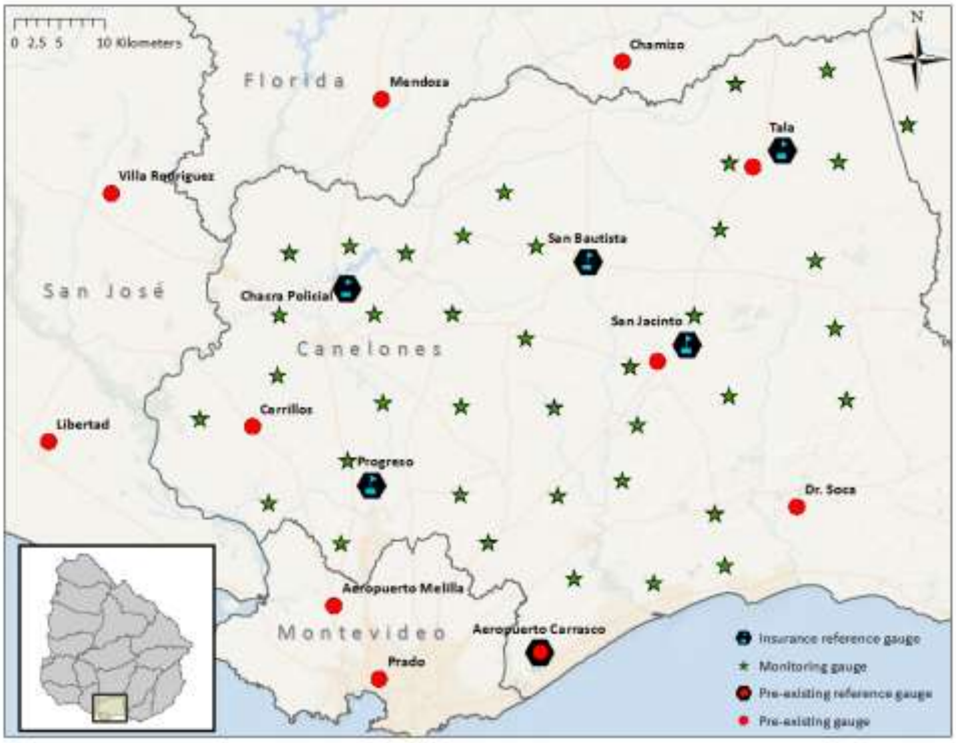


Figure 3.2 — Location of Rainfall Gauges around Canelones, Uruguay

It is important to note at this point that the model calibration requires a number of ex-ante choices on functional forms. As our aim lies on illustrating the features of the model and its application for basis risk estimation, we make the most reasonable choices based on standard assumptions in the literature and practical considerations. Wherever possible, we try to provide rationale and formal goodness-of-fit tests to assert the appropriateness of a certain assumption, though purposely avoid full-fledged discussions on particular modelling choices. Were this model to be used for insurance pricing or other product design considerations affecting real-world outcomes these assumptions should certainly be approached with much more caution.

In order to describe the model calibration, it is worth recalling that the model separates rainfall into an occurrence component and an amount component. The first component relates to the binary event of occurrence or non-occurrence of rainfall (also referred to as *wet* or *dry* days, respectively). The second component defines the number of millimeters of rainfall, conditional on the day being wet, over a given time span (24 hours in our case).

3.4.1 Rainfall occurrence process

In our model, the process for the occurrence of rainfall is completely determined by the two parameters of a first-order Markov chain: (i) p_{01}, or the probability of rain occurring in day t conditional on rain not having occurred in $t-1$, and (ii) p_{11}, or the probability of rain occurring in t conditional on rain not having occurred in $t-1$.[44] In this way, if the previous day were a wet day, the probability of raining on the current day would be p_{11}, and the probability of not raining $1-p_{11}$. Similarly, the probability of rain or no rain, conditional on the previous day being dry, would be determined by p_{01} and $1-p_{01}$, respectively.

We estimate these first-order Markov chain probabilities using the historical rainfall occurrence series at every gauge. We focus only on *pre-existing* gauges since these are the ones with a long enough time-span to appropriately capture climatological long-term occurrence probabilities for rain. Using only one to two years of data from the *insurance reference* and *monitoring* gauges could yield biased probability estimates due to the large inter-annual variability of rainfall. Table 3.2 shows the estimated probabilities for the eleven *pre-existing* rain gauges in our sample.

Spatial dependence

Following Wilks (1998), the occurrence of rainfall at one particular site can be determined at the simulation stage by taking a random draw from a standard uniform distribution and comparing it to either p_{01} or p_{11} (depending on whether the previous day was dry or wet): if lower, the day is considered to be wet, otherwise it is considered to be dry. This procedure can be extended to allow for dependence between multiple sites by drawing jointly from a (latent) bivariate normal distribution with correlation parameter ρ, calculating both (standard normal) cumulative probabilities, and comparing these to each site's occurrence probability for day t (given day $t-1$). Since the latent normal process is not observable from the data, estimation of ρ is achieved by iteration until the correlation between simulated occurrence series at both sites is close enough to the one observed in the data.

Unlike the estimation of the Markov-chain parameters, we include here all available rainfall gauges. The rationale behind this is that joint precipitation occurrences at the daily level between sites is still informative even in the presence of considerable inter-annual rainfall variability. Figure 3.3 shows a scatterplot of the estimated correlation parameters for the occurrence process between all pairs of sites. It is interesting to note that correlation declines relatively slowly with distance, such that two sites at a considerable distance (for

---

[44] While a first-order Markov chain is the overwhelming choice in the literature, the model can in principle accommodate a higher-order process. This could be relevant in the present case since we are mostly concerned about cumulative rainfall throughout several days. However, for all sites analyzed a second-order Markov chain does not improve the fit over a first-order one according to the Bayesian information criterion (BIC). We prefer the BIC statistic since it is considered as the preferred measure by the literature, with the Akaike information criterion (AIC) being known to overspecify model order for large sample sizes (Wilks, 1998)..

instance, farther than 50 kms. from each other) still experience a very similar pattern of wet and dry days.

Table 3.2 — Estimated First-Order Markov Chain Probabilities and Model Selection

Rainfall gauge	p_{01}	p_{11}	BIC for 1st order MC	BIC for 2nd order MC	Number of Obs.
Aeropuerto Carrasco	0.215	0.414	3,162.1	3,178.1	2,797
Aeropuerto Melilla	0.222	0.419	3,209.1	3,225.1	2,797
Cerrillos	0.137	0.288	2,423.5	2,439.5	2,797
Chamizo	0.148	0.296	2,538.4	2,554.3	2,797
Dr. Soca	0.119	0.292	2,239.7	2,255.7	2,797
Libertad	0.153	0.321	2,609.1	2,625.0	2,797
Mendoza	0.157	0.312	2,632.8	2,648.7	2,797
Prado	0.204	0.430	3,105.7	3,121.7	2,797
San Jacinto	0.127	0.288	2,324.5	2,340.5	2,797
Tala	0.122	0.277	2,267.4	2,283.4	2,797
Villa Rodriguez	0.158	0.300	2,629.3	2,645.3	2,797

Note: This table shows estimated first-order Markov chain probabilities for rain occurrence for each of the eleven pre-existing rainfall gauges in the study area. p_{01} indicates the probability of rain occurring on day t conditional on no rain having occurred on day $t-1$, while p_{11} indicates the probability of rain occurring on day t conditional on rain having occurred on day $t-1$. The third and fourth columns show the value of the Bayesian Information Criterion (BIC) for, respectively, a 1st order Markov chain (MC) and a 2nd order Markov-chain. A larger value indicates that a model is preferred according to the criterion.

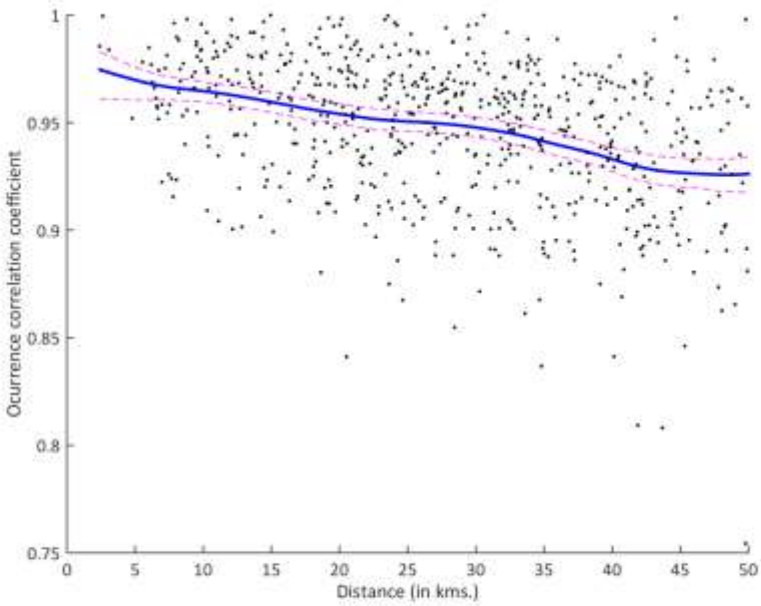


Figure 3.3 — Dependence Structure of Precipitation Occurrence Process

3.4.2 Rainfall amount process

Traditionally, chain-dependent stochastic models represent the process determining the amount of rainfall at two different sites (conditional on rainfall occurring) through two univariate distributions (one per site), which are linked together through an underlying latent bivariate normal distribution with a given correlation parameter ρ (calibrated in a similar fashion to the one in the latent bivariate normal process that links the occurrence processes explained in the previous subsection). As described above, in this chapter we extend this model by the use of copulas in order to account for much more flexible dependence structures between the two site-specific univariate distributions. The following subsections describe the estimation of each site's univariate distribution and that of the proposed dependence between two sites.

Univariate distributions

The most appropriate parametric distribution for representing the amount of rainfall at the daily scale has long been an issue of discussion in the literature. Several parametric families have been put forward as suitable candidates, some examples being gamma, Weibull, double exponential, and lognormal, among others. While the literature has not reached a consensus on which family can best represent daily rainfall amounts, the gamma family is a frequent choice that has been shown to adapt well to the skewness and other properties of daily rainfall processes (Ison, Feyerherm, and Bark, 1971; Woolhiser, 1992). In addition, the gamma family is a suitable choice to represent precipitation data because of its flexibility in distribution shapes making use of only two parameters: shape and scale (Wilks, 1990).

Table 3.3 — Estimated Gamma Parameters

Rainfall gauge	**Shape**	**Scale**	**Number of Obs.**
Aeropuerto Carrasco	0.598	18.291	753
Aeropuerto Melilla	0.590	19.007	775
Cerrillos	1.008	16.024	451
Chamizo	1.129	16.710	486
Dr. Soca	1.064	16.729	401
Libertad	1.150	15.009	514
Mendoza	1.157	15.301	519
Prado	0.565	19.286	738
San Jacinto	1.065	17.103	423
Tala	1.003	19.397	405
Villa Rodriguez	1.125	15.305	514

Note: This table shows estimated scale and shape parameters for a gamma distribution for each of the eleven pre-existing rainfall gauges in the study area. The estimated distribution is used to represent rainfall amount on a certain day, conditional on rainfall occurring on that day.

In our analysis, we assume that daily rainfall amount (conditional on a day being wet) follows a gamma distribution with unknown shape and scale parameters, which we estimate from available data for each *pre-existing* rain gauge.[45] The rationale for narrowing the sample only to the *pre-existing* gauges is again—as discussed above for the case of the Markov chain parameters—to count with a long-enough history to observe sufficient realizations from the underlying distribution. It is worth noting that assuming a different suitable distribution family such as Weibull does not qualitatively change our results or conclusions. Table 3.3 shows the estimated gamma parameters for the eleven *pre-existing* rain gauges in our sample.

Spatial dependence

We have discussed above that any bivariate distribution can be uniquely represented through an appropriate copula function evaluated at the corresponding marginal distributions' CDF. Having selected the marginal distribution for each site we now need to select an appropriate copula model to represent dependence between rainfall amounts at any two sites.

As argued above, we will work with the Gumbel copula, as it suits well processes that simultaneously realize upper tail values. To motivate this choice and illustrate the presence of upper tail dependence in our data, Figure 3.4 shows a scatterplot of the joint rainfall realizations (expressed as cumulative probabilities of the underlying estimated distribution) between *Prado* and *Aeropuerto Melilla* rain gauges (panel A) and *San Jacinto* and *Mendoza* gauges (panel B).[46] It can be seen that for the pair of gauges relatively close to each other, there is an overall high degree of dependence between the two series, but this dependence seems to increase at higher regions of the distributional support (upper-right corner of the scatterplot). The joint occurrence of values at the upper tail is evident even for the pair of gauges that are more than 40 kms. apart from each other, even though the overall degree of dependence between the two series is substantially lower.

It is well known that determining the right copula family through goodness of fit (GoF) tests can be problematic. While the bootstrap procedures suggested in Genest, Rémillard, and Beaudoin (2009) represent the state-of-the-art techniques, these are very expensive computationally and, based on Monte Carlo simulations, the authors conclude that no one single test is appropriate for all situations. Moreover, as the power of any test increases with sample size, it is rare for empirical data to properly fit any one parametric family,

[45] The choice of a single distribution for all rain gauges may not be the most appropriate, but it is a key assumption that allows to spatially interpolate the parameters to be used in the simulations. A fully non-parametric distribution for each site would be ideal. This, however, seems problematic in our context, since for the calculation of basis risk we need to be able to simulate at arbitrary points in space, calling for some sort of spatial interpolation of distributions. The work by Mosthaf, Bárdossy, and Hörning (2015) on random mixing of spatial random fields could provide a potential solution to this problem.

[46] The results are not linked to the specific choice and estimation of the univariate distribution at each site. Scatterplots using cumulative probabilities from the empirical (non-parametric) CDF or the univariate rank for each observation yield equivalent characteristics.

which generally results in all candidate distributions being rejected.[47] Instead, we follow here an alternative specification check proposed by Serinaldi (2008). The idea behind this methodology is to compare the theoretical relationship between Kendall's τ rank correlation and the degree of tail dependence implicit in different candidate copula families, against the empirical (estimated) counterparts, for all pairs of series. For instance, a given pair of gauges has a certain estimated Kendall's τ and coefficient of tail dependence.[48] By taking all available pairs one can plot an empirical curve of one measure against the other, and then compare this empirical curve to different theoretical curves implicit in each copula family's definition. Figure 3.5 shows the result of this exercise. We compare the empirical curve to those from other families exhibiting positive upper tail dependence, such as Gumbel, Student t, and three other families from the Archimedean class. Even though none of the curves perfectly accommodates the statistics from the observed rainfall data, it can be seen that the Gumbel family is the one that best approximates the empirical curve, particularly for high levels of dependence. Altogether, the evidence, though heuristic in nature, indicates that the Gumbel copula can be considered as an acceptable representation of the dependence between rainfall amounts in our study area.

Figure 3.6 shows a scatterplot of the estimated Gumbel copula parameters for the amount process between all pairs of sites, including all available rainfall gauges (in a similar way as for the estimation of the dependence between rainfall occurrence processes). In contrast to the correlation in rainfall occurrence (Figure 3.3), the dependence between rainfall amount seems to decrease at a faster rate with distance to the reference gauge. Equivalent to the findings in decorrelation analyses, this feature is directly behind the common perception of basis risk being larger for rainfall products than for products based on other common weather variables, such as temperature. In the next section we will directly quantify the degree of basis risk in order to determine how the observed decrease in dependence translates into unattractiveness of an insurance product.

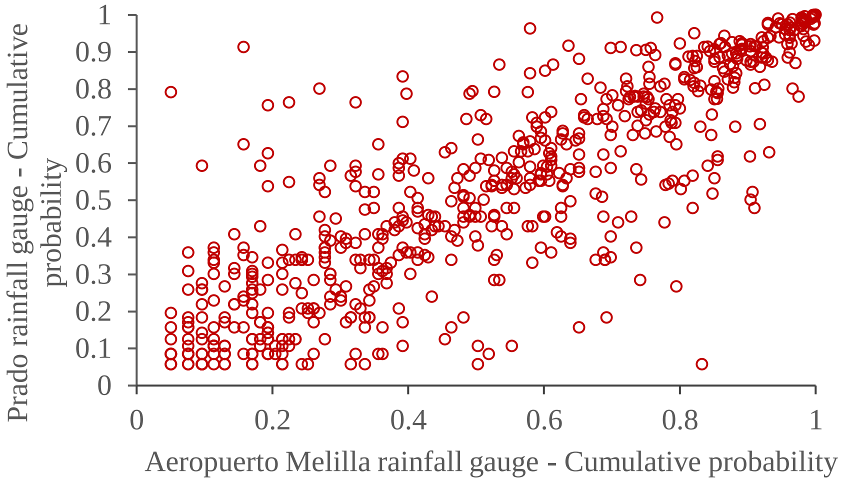
3.4.3 Spatial Interpolation of Parameters

In order to estimate the degree of basis risk across our study area, we need to arrive at a full description of the underlying bivariate precipitation distribution at two given locations. For this, we need to be able to interpolate the parameter estimates discussed in the preceding subsections at arbitrary points in space. Different methods of spatial interpolation of distributional parameters—estimated from site-specific precipitation

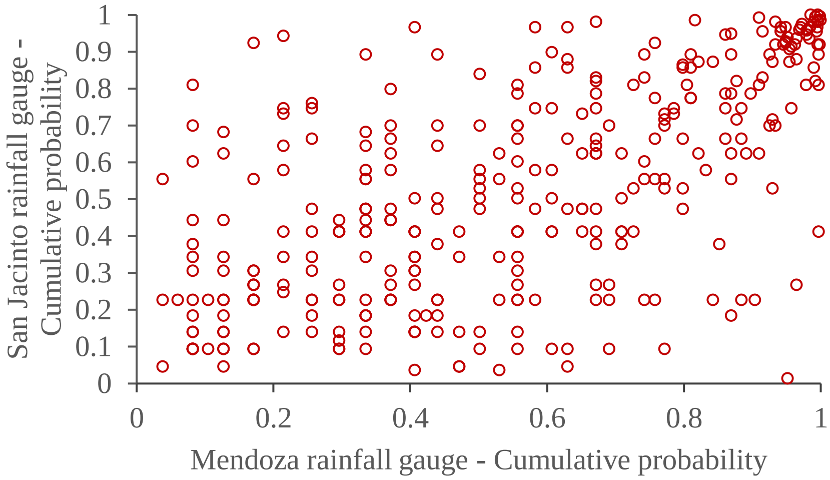
[47] The results from a (least computationally intensive) bootstrap test show that the Gumbel copula cannot be rejected (at the 5% level) in about 49.5 percent of the pairs. Given the large number of daily observations we count with, we take this as an additional piece of evidence towards the reasonably good fit of this copula family.

[48] We rely on a nonparametric estimator for the upper tail coefficient (λ_U) based on the work by Capéraà, Fougères, and Genest (1997). See Serinaldi (2008) for details and further discussion.

data—have been implemented in the literature (e.g. Baffault et al., 1996; Kleiber, Katz, and Rajagopalan, 2012; Serinaldi and Kilsby, 2014). This subsection describes the two methods we use in our context.[49]



Panel A. Rainfall amount at *Prado* and *Aeropuerto Melilla* gauges (9.6 kms. apart)



Panel B. Rainfall amount at *San Jacinto* and *Mendoza* gauges (42.3 kms. apart)

Figure 3.4 — Overall Correlation and Tail Dependence of Rainfall at Sample Sites

Location-specific parameters (pertaining to the assumed functional forms for either rainfall occurrence or amount at one specific site) require a method that takes into consideration the specific location at which these parameters were estimated. These

[49] It is important to note that the following methods implicitly assume spatial isotropy in precipitation dependence within the study area. This implies that the relationship between two given points does not depend on the direction at which they lie from each other; in contrast to potential directional precipitation patterns caused by, for instance, prevailing winds or geographic accidents. Simple directional tests cannot reject the null hypothesis of isotropy in precipitation, while survey responses described above indicate that farmers in the area are not systematically aware of directional patterns in rainfall. For these reasons we proceed with this assumption.

parameters are the two probabilities in the first-order Markov-chain—from the rainfall occurrence process—, and the scale and shape parameters of the gamma distribution—from the rainfall amount process.

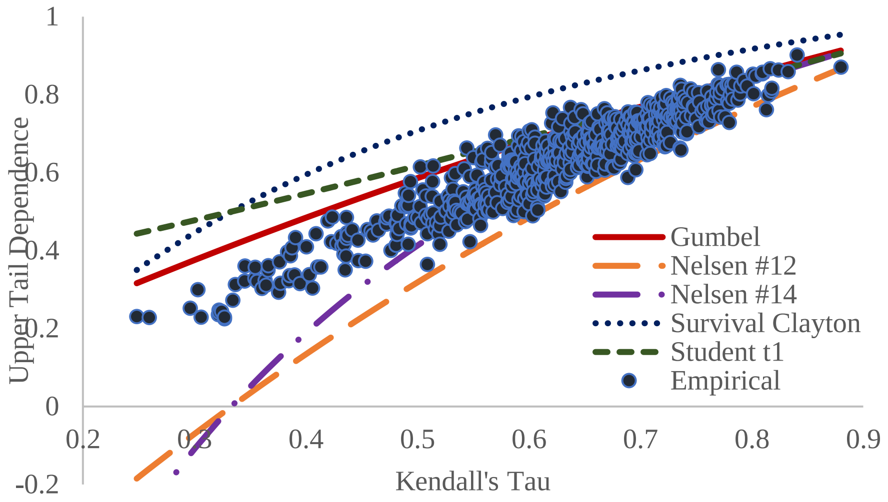


Figure 3.5 — Kendall τ Correlation and Upper Tail Dependence λ_U

In contrast, for the parameters that relate to the dependence (in either model component) between two sites we assume spatial homogeneity and thus only take into consideration the distance between the two locations in the interpolation. These parameters are the correlation coefficient ρ driving the dependence in rainfall occurrence and the Gumbel copula parameter θ behind the relationship between the amount of rainfall, conditional on rainfall occurring, at two different sites.

Location-specific parameters

We choose inverse distance weighting (IDW) as the interpolation method because, despite being one of the less sophisticated approaches available, it is simple enough and works well in regular geographies such as our study site.[50] IDW estimates the value at an arbitrary, target point in space as the average value of observed adjacent points, weighted by the inverse of the (n-powered) distance to them. In particular, the estimation for an arbitrary point r is as follows:

$$\alpha_r^* = \sum_{l=1}^{L} \frac{\alpha_l}{{D_{rl}}^n} \Big/ \sum_{l=1}^{L} \frac{1}{{D_{rl}}^n} \qquad (3.4)$$

[50] An arguably better, alternative method would be ordinary kriging, which makes use of an estimated spatial covariance structure between sites (covariogram) to obtain an interpolated point estimate at an unobserved location, together with its standard error. Unfortunately, for the estimation of the location-specific parameters we focus only on the subset of eleven pre-existing gauges, which provides us with too few pairs to obtain reliable estimates for the covariogram.

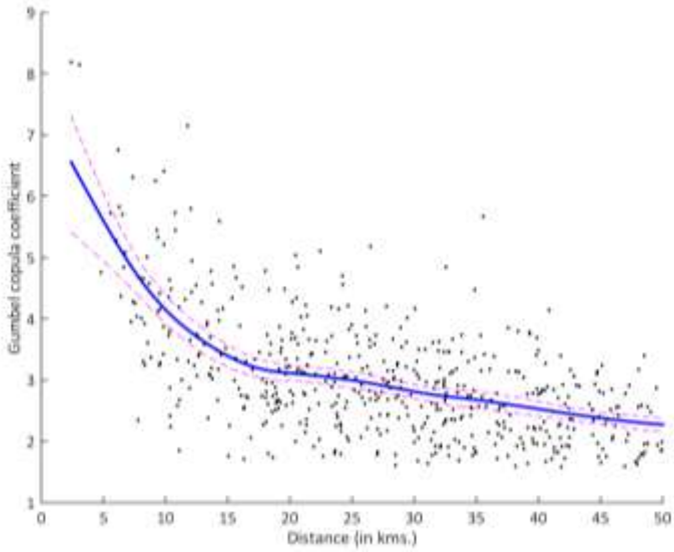
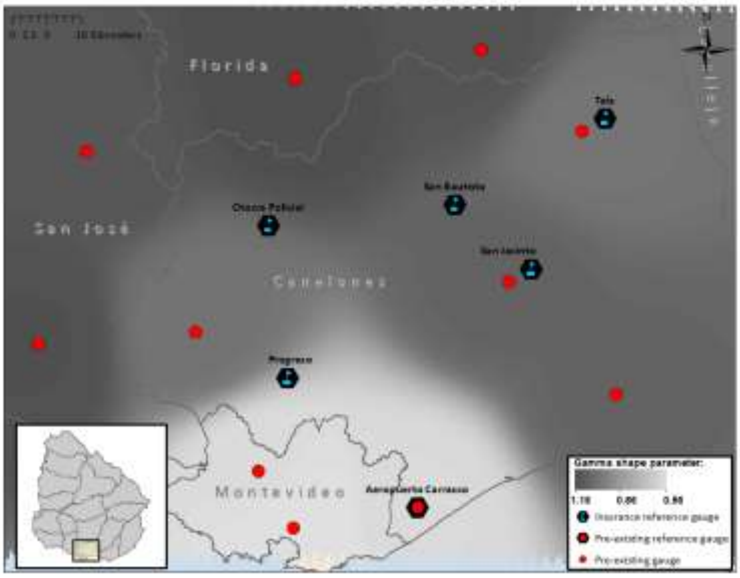


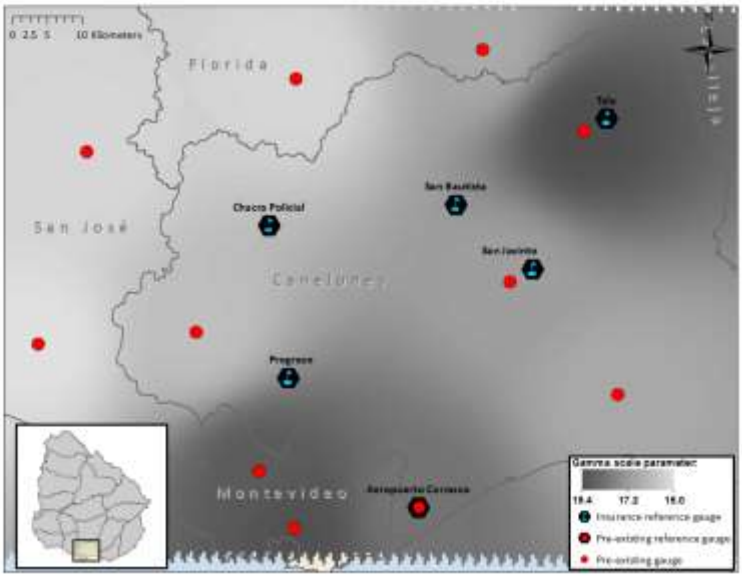
Figure 3.6 — Dependence Structure of Precipitation Amount Process

Where α_l is the observed value at each of the L locations and D_{rl} is the straight-line distance between location r and location l. Note that the power parameter n is a free parameter that controls the degree of smoothness of the interpolation. Very large values of n implicitly mean that the only relevant observations are those in the relative vicinity of the target point, while small values of n take into account observations from farther away locations. We choose the value for this parameter by cross-validation (leave-one-out method), that is, we choose the power n that minimizes the mean square error (MSE) between the estimated value (excluding that site) and the observed value, at all locations.

Figure 3.7 shows shaded maps with the interpolated values of the shape (panel A) and scale (panel B) parameters for the gamma distribution. In turn, Figure 3.8 shows similar maps with the interpolated values of the Markov-chain precipitation probabilities after a dry (panel A) and wet (panel B) day, respectively. It can be seen that, while the estimated parameters don’t fluctuate substantially among the department of Canelones (our main study area), they seem to be consistently different near the city of Montevideo. In particular, relative to the more rural areas in Canelones, rainfall around the city of Montevideo seems to occur more often, yet, conditional on occurring, accumulated rain seems to be lower (see also Appendix Table 3). This could be related to these gauges’ proximity to a large body of water or to the existence of an *urban effect*. This will be of importance for the analysis below since ‘Aeropuerto Carrasco’ –one of the reference gauges to which the insurance product is linked– is located in this area.



Panel A. Shape parameter



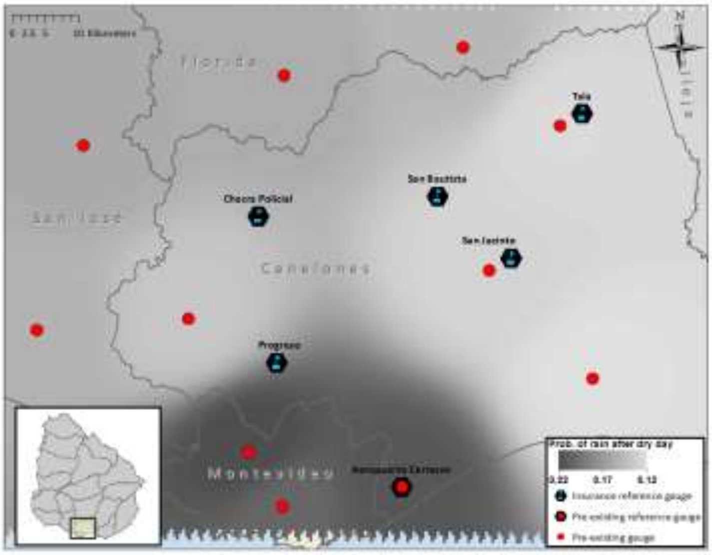
Panel B. Scale parameter

Figure 3.7 — Interpolated Gamma Distribution Parameters

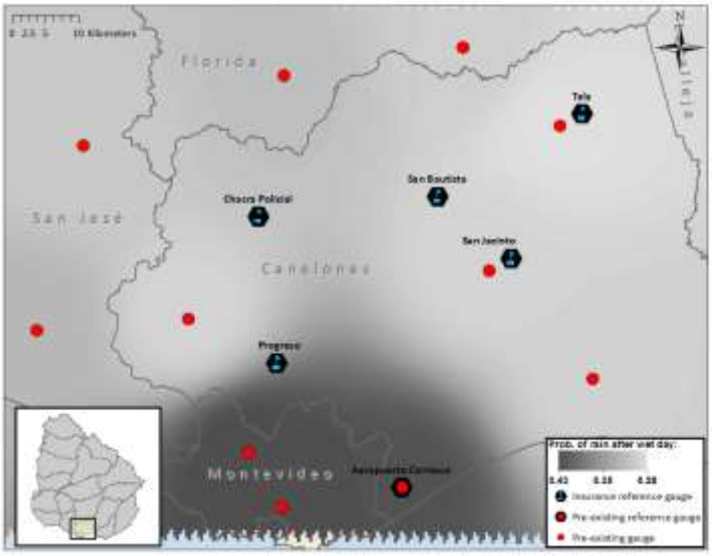
Dependence parameters

For the interpolation of the parameters relating to the dependence structure of rainfall occurrence and rainfall amount we only focus, as discussed above, on the distance separating any two pairs of points, after which we fit a cubic B-spline using information from all available pairs (Bjørnstad and Falck, 2001).

In addition to the scatterplots depicting the estimated dependence parameters pertaining to, respectively, the rainfall occurrence and amount processes, Figures 3.3 and 3.6 show the estimated splines together with 95% confidence intervals calculated according to Bjørnstad and Falck's (2001) method. In order to interpolate the dependence parameter between two arbitrary points in space, we simply use the distance separating them and obtain the corresponding value from the fitted spline curve.



Panel A. Probability of Rain Occurrence after Dry Day (p_{01})



Panel B. Probability of Rain Occurrence after Wet Day (p_{11})

Figure 3.8 — Interpolated Markov-Chain Probabilities for Rain Occurrence

3.5 Simulations and Results

This section describes the estimation of downside basis risk for the insurance product described above across a uniform 20 by 35 grid of hypothetical locations within the limits of our study area, for a total of 700 sites. These represent potential locations of farmers and allow us to characterize basis risk in a general sense.

We first identify the closest reference gauge for each location in the grid. Next, we interpolate all distributional and dependence parameters as described in the previous section, thus obtaining a complete characterization of the bivariate distribution for each pair of sites. Finally, we draw 450,000 realizations of daily rainfall by Monte Carlo simulation. This process is as follows:

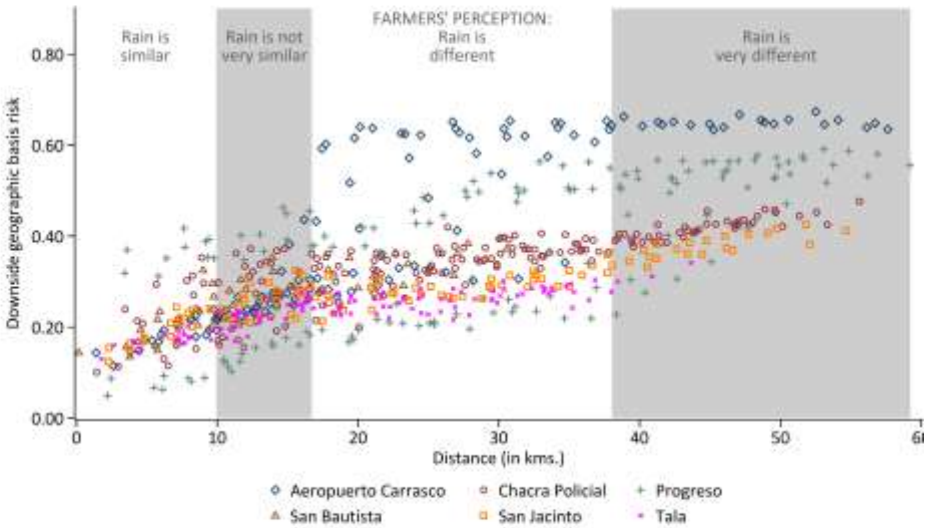
1. Simulate a bivariate rainfall *occurrence* process combining the Markov chain parameters for each site in the pair and a bivariate random process with the corresponding correlation parameter (given the distance between sites).
2. Simulate a bivariate rainfall *amount* process combining the univariate gamma distribution parameters for each site and 'linking' them through a Gumbel copula. In particular:
    a. Draw from a bivariate Gumbel copula with dependence parameter $\hat{\theta}$ (given the distance between sites).
    b. For each site, obtain the daily rainfall amount evaluating the inverse gamma distribution (with the corresponding parameters) at the cumulative probability drawn in (a) from the Gumbel copula.[51]
3. Finally, combine (1) and (2) to obtain the simulated time-series of rainfall at both sites.

After obtaining the daily time-series, we accumulate rainfall into overlapping 10-day periods and calculate the maximum value among these over a 90-day period (the equivalent of one year's summer season). This allows us to obtain a simulated time-series of the index at both sites along 5,000 seasons. With the entire simulated series of the index at a given pair we are then able to calculate alternative measures for basis risk.

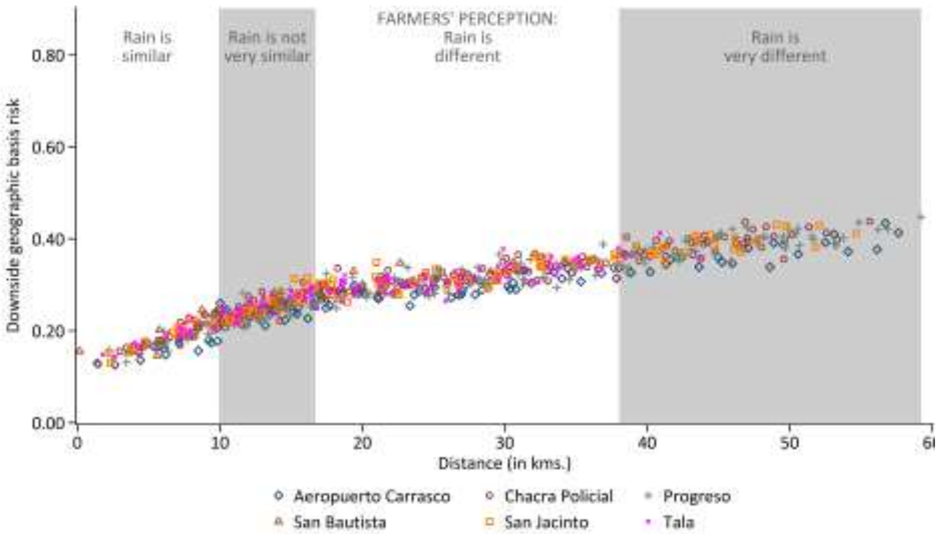
Figures 3.9 and 3.10 show estimated probabilities of downside basis risk for, respectively, the 85th-and 95th-percentile insurance products. Panel A in each figure plots the estimated conditional probability of not receiving an insurance payout when a loss occurs by relying on the interpolated distributional parameters. This represents our main measure of actual basis risk of the insurance product under consideration. Panel B, instead, shows a benchmark measure derived under the assumption that distributional parameters at the farmer's plot are identical to the ones at the insurance reference gauge. In this way, while panel B reflects the basis risk implicit in the spatial separation of the two locations (that is, arising only from the imperfect dependence in rainfall patterns due to the distance

[51] An important aspect to take into account in stochastic rainfall generator models is the so-called *spatial intermittence* issue as referred to by Bárdossy and Plate (1992). This is related to rainfall amounts at an arbitrary location being generally lower conditional on nearby stations being dry than conditional on nearby stations being wet. In other words, if one imagines a given storm as a continuous field, locations in the center of the field (thus surrounded by all wet locations) should receive more rainfall than those in the edges of the field (thus nearby some dry locations). Failure to address this problem leads to unrealistically sharp transitions between wet and dry portions of the spatial domain. Wilks' (1998) approach is to model the univariate rainfall amount distribution as mixed exponential and to later select the exponential in the mix with the lower mean whenever a day is simulated as being wet by a random draw close to the boundary implied by the Markov chain rainfall probability. In this chapter, instead, we take advantage of the *mixtures of powers* simulation method for the Gumbel copula described in Trivedi and Zimmer (2007) and, in a similar fashion to Wilks (1998), impose draws from the lower part of the distribution when one site in the pair is dry.

between them), panel A reflects also the *additional* basis risk arising from the different site-specific rainfall distributions.



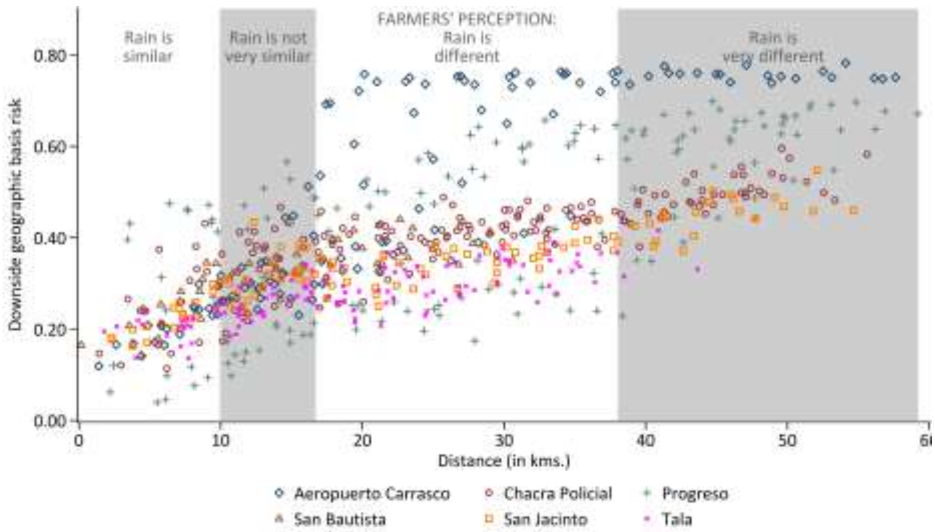
Panel A. Interpolated Univariate Rainfall Distributions



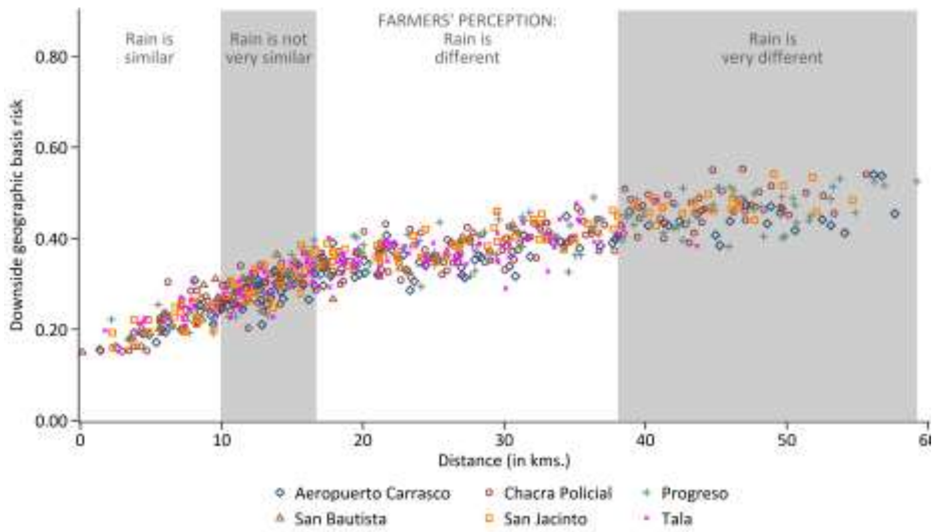
Panel B. Identical Univariate Rainfall Distributions

Figure 3.9 — Downside Basis Risk - 85th Percentile Product

A few interesting patterns arise from the figures. First, the level of basis risk for this particular insurance product is not negligible. At best, in the case of plots located very close to the reference rainfall gauge, the insurance product would fail to pay 1 out of 10 times that the farmer experienced a significant loss. In general, the insurance product would fail to pay farmers between 10 and 50 percent of the times they experienced a loss, depending on their specific location, with this rate going up to 80 percent in extreme cases. Even though the above figures look large, it is interesting to see how these levels of basis risk compare to the theoretical upper bound for basis risk derived by Clarke (2016). Table 3.4 shows summary statistics for the ratio of basis risk at each location and its theoretical upper bound, by reference weather station and for both insurance products under consideration. It can be seen that downside basis risk is well below the upper theoretical bound in all cases, at levels roughly between 10 and 60 percent of this upper bound. In other words, downside basis risk seems to be within the limit in which a sufficiently risk-averse farmer would benefit from (and thus demand) a positive level of insurance. However, the table also shows that this ratio shows considerable variation across reference gauges, a point to which we return below.



Panel A. Interpolated Univariate Rainfall Distributions



Panel B. Identical Univariate Rainfall Distributions

Figure 3.10 — Basis Risk - 95th Percentile Product

Second, and as expected, downside basis risk increases the farther away a plot is from the insurance reference gauge. Interestingly, however, the relationship between basis risk and distance seems to be concave. In unreported regression results, distance and distance squared variables are, respectively, positively and negatively (and statistically-significantly) related to our basis risk measure. This is in line with the results in Chapter 2, which finds demand for insurance products to be negatively related to the logarithm of the distance to the reference weather station. This feature of basis risk partly stems from the upper tail dependence implicit in the Gumbel copula; an alternative model assuming dependence under the traditional Gaussian assumption would thus fail to incorporate this property.

Third, basis risk does not seem to increase as pronouncedly with distance as it is generally perceived. For instance, based on the 85th-percentile product with identical distributional parameters (Figure 3.9, panel B), the probability of not receiving a payout when a loss occurs is around ten percent when a plot is one kilometer away from the reference gauge,[52] with this probability increasing to around 30 percent when the two sites are located 20 kilometers away and to around 40 percent when a plot is considerably far from the reference weather station (50-60 kms.). This is in contrast to the general notion that basis risk increases rapidly with distance when the underlying process is rainfall. This notion is captured by the shaded areas in the figures, which show estimated distance ranges at which an average farmer considers rainfall patterns to coincide or not.[53] For instance, on average farmers consider rainfall patterns at two sites located at a distance between 9.9 and 16.8 kilometers as *not very similar*, and rainfall patterns at distances above 38 kilometers as *very different*. However, in the latter range, an insurance product would still correctly pay around 3 out of 5 times in which a significant loss occurred at the farmer's plot. Overall, this indicates that, on one hand, farmers seem to overestimate the degree of geographical variation in rainfall and, on the other hand, that even though it may be true that the overall similarity between rainfall patterns at two sites decreases rapidly with distance, this may not be the case for the dependence between *extreme* rainfall patterns directly behind basis risk.

Fourth, downside basis risk is generally lower, all things being equal, for the 85th-percentile product than for the 95th-percentile one. Two effects come into play for this. On one hand, the probability mass around the 85th-percentile of the rainfall distribution is higher than around the 95th-percentile, which implies a larger number of realizations around the trigger and thus an increase in the unconditional mismatch probability. On the other hand, the probability of experiencing a loss (the denominator in our basis risk measure) is by definition higher in the case of the 85th-percentile product. Overall, the second effect dominates the first one and results in a lower basis risk for the 85th-percentile

[52] It is worth noting that basis risk estimates in the lower end of the distance spectrum are not precisely estimated, given the lack of weather station data located at minimal distances from each other.

[53] These ranges are derived from estimating an ordered logit model with perceived rainfall similarity (Categories: very similar, similar, not very similar, different, very different) as the dependent variable and distance as the independent one. The ranges are robust to including squared distance in the estimation, as well as to including dummies for the direction at which a reference weather station is located, the previous variables not being statistically significant in any specification.

product. This result is intuitive, in that an insurance product that covers against frequent risks should compensate often at rainfall levels above the trigger, and would only be at risk of incorrectly not compensating the farmer when rainfall is within a narrow band around the trigger.

Table 3.4 — Downside Basis Risk and Theoretical Upper Bound

Panel A. 85^{th} Percentile Product

Rainfall gauge	Median	Mean	Std. Dev.	Min.	Max.	Number of Obs.
All rainfall gauges	0.36	0.39	0.16	0.06	0.79	700
Aeropuerto Carrasco	0.38	0.48	0.23	0.13	0.79	122
Chacra Policial	0.42	0.40	0.09	0.12	0.56	171
Progreso	0.46	0.43	0.19	0.06	0.70	149
San Bautista	0.33	0.32	0.08	0.16	0.45	47
San Jacinto	0.33	0.33	0.08	0.15	0.50	100
Tala	0.29	0.28	0.05	0.15	0.40	111

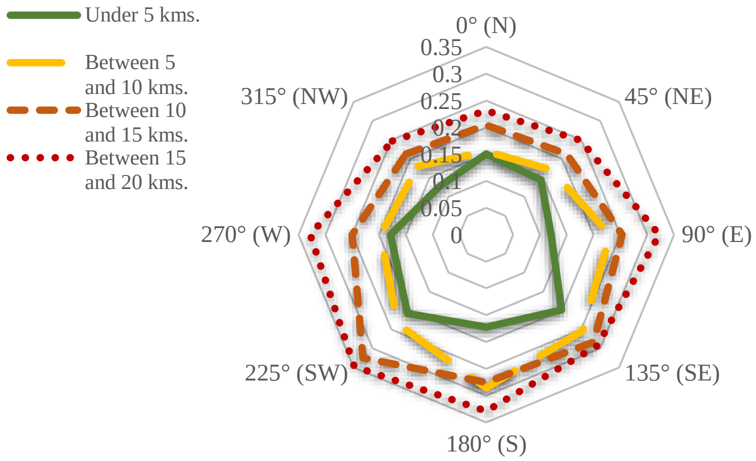
Panel B. 95^{th} Percentile Product

Rainfall gauge	Median	Mean	Std. Dev.	Min.	Max.	Number of Obs.
All rainfall gauges	0.32	0.35	0.14	0.05	0.71	700
Aeropuerto Carrasco	0.34	0.43	0.20	0.12	0.71	122
Chacra Policial	0.38	0.36	0.08	0.11	0.50	171
Progreso	0.41	0.39	0.17	0.05	0.62	149
San Bautista	0.30	0.28	0.07	0.14	0.40	47
San Jacinto	0.30	0.30	0.07	0.13	0.45	100
Tala	0.26	0.25	0.04	0.14	0.36	111

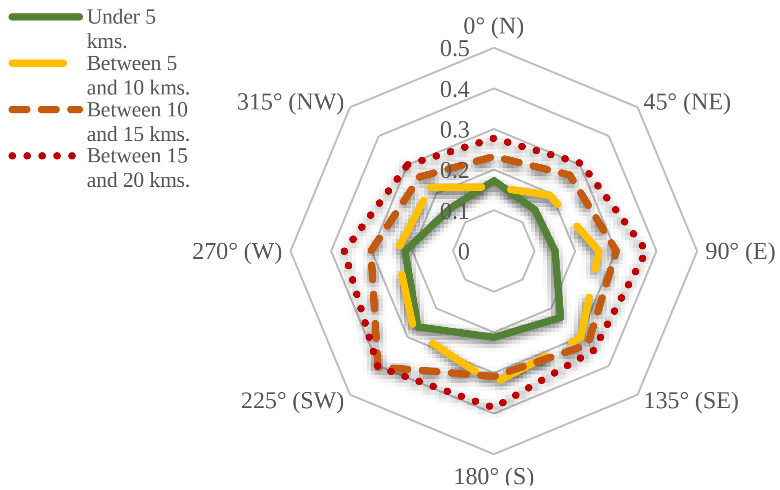
Finally, an important consideration that has been generally overlooked by the literature is the natural variation in precipitation amount between different geographic locations.[54] In our particular study area, though relatively small in extent, rainfall exhibits a subtle but non-negligible amount of variation between locations (e.g. near the coastal city of Montevideo vis à vis near the center of Canelones). This small variation in rainfall patterns can dramatically increase or reduce the degree of basis risk of an insurance product, depending on the exact locations of both the plot and the reference gauge. This can be clearly seen by comparing panels A and B in both figures (which differ in that panel B excludes any variability in the sites' precipitation distributions), where this feature almost doubles the level of downside basis risk for locations assigned to Aeropuerto Carrasco

[54] The literature has however widely acknowledged the limitations of index insurance products in regions with large topographic variation. See Kost et al. (2012), for example.

since, as discussed before, overall rainfall seems to be lower for this particular reference gauge relative to other weather stations (see Appendix Table 3).[55]



Panel A. 85th Percentile Product



Panel B. 95th Percentile Product

Figure 3.11 — Downside Basis Risk and Direction to Reference Weather Station

Figure 3.11 portrays this feature in more detail. In particular, the figure shows the average level of downside basis risk taking into consideration the direction at which the reference weather station lies from each location. This is shown along eight directions and for groups of locations at four different distance ranges from the weather station. It can be seen that, if the assigned reference gauge lies to the south-west of a particular location, this location is subject to higher levels of basis risk relative to other locations within a similar distance to the reference gauge. This aspect of basis risk demands for a much more careful

[55] This effect can be even stronger if rainfall dependence were anisotropic, that is, if dependence in precipitation varied according to the direction in which the two points lay from each other. In our specific case, however, as argued above, this property does not seem to hold.

study of regional precipitation patterns when designing a new index insurance product, ideally involving local meteorological experts.

3.6 Conclusions

We develop a novel methodology for estimating the extent of spatial basis risk for an arbitrary rainfall index insurance product. We rely on a stochastic rainfall generator model standard in the hydrological literature and extend it to accommodate non-traditional patterns of dependence between rainfall distributions at two nearby sites through the use of a bivariate copula. In particular, we intend to capture the general tendency of *extreme* precipitation amounts occurring jointly at nearby sites more often than amounts at other regions of the support of the rainfall distribution.

We apply the methodology to estimate the degree of spatial basis risk in an index insurance product against excess rainfall targeting horticultural farmers in Uruguay. After calibrating the model using unique historical precipitation data, we conduct Monte Carlo simulations using the bivariate stochastic rainfall generator which allow us to calculate our measure of downside basis risk for this product.

We find that even though the degree of basis risk is considerable, it remains well below the theoretical upper bound from a model of demand for index insurance developed by Clarke (2016), implying that the product under consideration provides valuable insurance properties. This suggests that basis risk is not large enough so as to fully explain the lack of demand generally found in index insurance pilots.

In addition—and as expected—basis risk increases concavely with distance, but the rate of increase is generally lower than what farmers perceive it to be. This points to the existence of important information asymmetries and indicates the need to complement the introduction of new index insurance products with extensive training on the spatial properties of rain, and basis risk in particular. Of course, low levels of education among farmers in developing countries represent a barrier to these type of technical trainings. The point is still crucial, though, since farmers are typically presented with a complex insurance product framed in terms of millimeters of rain and expected to make rational insurance decisions based on their prior knowledge and expectations on complex meteorological phenomena.

Finally, a central aspect to consider in future products is the natural variation in rainfall patterns within a given area, even in the case of regular terrains such as those in our study. Subtle differences in rainfall patterns between the plot location and its reference rainfall gauge may result in substantial disparities in basis risk between locations. For some locations in our study, this effect was responsible for almost half of the total levels of downside basis risk. In extreme cases, local geo-climatic features could prove more important for determining basis risk than the actual distance to the reference gauge. Surprisingly, this point has been completely ignored by the index insurance literature until now, beyond evident settings such as mountainous topographies.

In summary, the present study has two main policy implications for existing and future index insurance schemes. First, much more thorough training needs to be offered to targeted farmers, aided by historical data and a careful description of the spatial relationship between overall rainfall patterns and that around extreme rainfall events. Second, in order to minimize the extent of spatial basis risk a careful consideration of subtle regional variations in rainfall patterns is essential, in addition to the more commonly studied distance aspect. This should be ideally carried out at the initial design stages of an index insurance scheme, when selecting from existing or determining the placement of new weather stations to be used as a reference for the product.

In general terms, even though the methodology applies to products insuring against excess rainfall, it directly applies to any other products based on a rainfall index such as those covering against droughts, and it can be extended to products based on indices of other weather variables, such as temperature or humidity. In the case of temperature, for example, several multisite stochastic temperature generators exist, which could be used to calculate mismatch probabilities between temperature at a particular site and temperature at a reference weather station.

Despite its advantages, however, this methodology can only tackle the estimation of the spatial component of basis risk. As discussed above, there exist a number of different sources for the mismatch between a farmer’s losses and an index insurance product’s payout. An analysis incorporating these other sources would require longitudinal farm-level data with which we do not count. This is an important avenue for future research. Such an analysis can contribute to the design of innovative index insurance products that help mitigate the negative aspects of basis risk and reveal the potential for novel financial instruments to enhance farmers’ resilience. This is important as, for instance, if most basis risk were to arise from spatial variability in weather, the commonly proposed solution of increasing weather station density would be appropriate. Alternatively, however, if a large fraction of basis risk were to be explained by idiosyncratic differences in farmers’ abilities and technologies to cope with weather, policy recommendations would be entirely different.

CHAPTER 4

Demand Heterogeneity for Index-Based Insurance: The Case for Flexible Products*

There has been an ongoing debate over the past few decades around weather index insurance and its merits as a tool for smallholder farmers' agricultural risk management. Weather index insurance's attractiveness—as an alternative to traditional, indemnity insurance products—arises from its cost-minimizing features, by reducing loss verification and information assymetry problems. The flip side is that, by solely relying on an objective index, insurance payouts cannot fully capture the losses of an individual farmer, a mismatch known as *basis risk*. Weather index insurance implementations around the world (at both small and large scales) have met with mixed success, giving rise to a substantial literature analyzing different aspects of the ex-ante and ex-post risk-management benefits and the determinants behind the feeble demand for this type of instruments.

Notably, the vast majority of studies in this area have largely overlooked a subtle, yet important aspect in which the implementation of weather index insurance has also departed from traditional risk management instruments. While indemnity insurance provides a farmer with coverage against all adverse events that may affect his or her crops—thus naturally adapting to any specific risk profile—, the implicit rigidity in most existing weather index insurance products has made them largely inadequate in this realm, greatly undermining their effectiveness as risk management tools. To date, the design of these products has been characterized by a predetermined payout structure, generally calibrated considering the *standard* risk profile of a representative farmer. The reasons for this one-size-fits-all structure can be traced to long-established insurance habits, where an individual needs only to purchase one single indemnity-based product to cover her entire risk exposure. However, the relative simplicity of offering one standardized product comes at the cost of ignoring heterogeneity in agricultural risk profiles, considerably lowering the

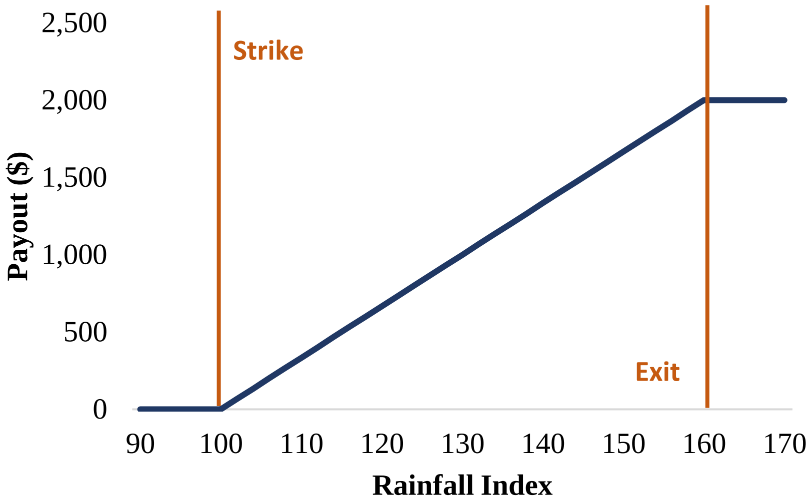
* This chapter is co-authored with Miguel Robles. The authors' contributions are as follows. MR conceptualized the project and obtained funding for it. FC and MR designed the insurance product, carried out fieldwork, and collected the data. FC conceptualized and conducted the analyses and wrote the paper. All authors read and approved the final manuscript.

product's worth as a risk management tool for the large proportion of farmers deviating from the standard profile.

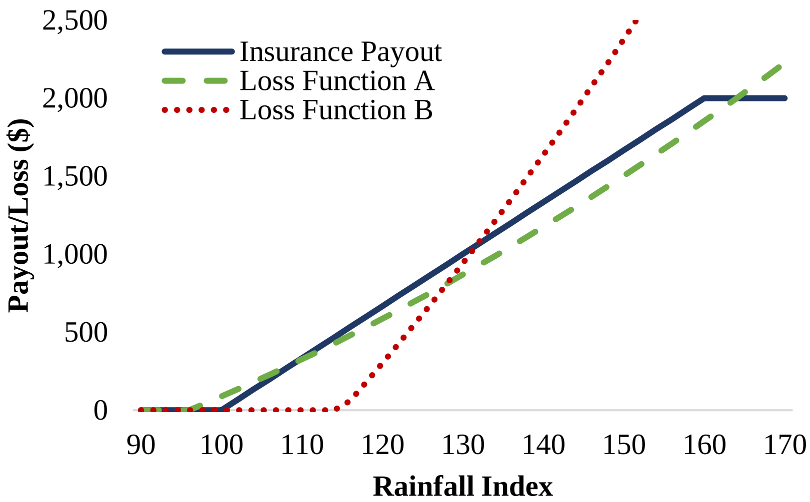
In reality, this issue is nothing else than a subcomponent of basis risk—known as *design* basis risk—, related to the elements in the design of an index product that contribute to the mismatch between payouts and losses. For instance, if important losses in a particular rainfed crop were to ocurr after 5 dry days, an insurance product that pays only after a dry-spell of 8 days would entail a high degree of design basis risk. Alternatively, if the risk of drought for this crop were not related only to lack of rain but also to high temperatures, not considering this latter weather variable would introduce additional design basis risk. In this case, however, the rigidity in payout structure goes beyong a context-specific, case-by-case design issue. Rather, failing to consider the entire distribution of agricultural risk profiles constitutes a fundamental instance of design basis risk which has been internalized and inherited throughout most existing index products to date.

To be specific, a typical index insurance product offers a payout that is linear in the index above/below a certain initial threshold (strike), up to a maximum/minimum level (exit) after which the payout is then constant. In the rest of the chapter, we refer to this particular type of index insurance as *composite* products. The different elements of an arbitrary composite product against excess rainfall are shown in Panel (a) of Figure 4.1. In order to illustrate the heterogeneity in agricultural risk profiles, Panel (b) shows hypothetical loss functions for the crops of two different farmers. The payout structure of the composite product follows closely loss function A; that is, when this farmer suffers small (large) losses around rainfall index values of 110 (150), the insurance compensates him approppriately with small (large) payouts. For the farmer with loss function B, however, the composite insurance unecessarily pays a small amount around index values of 110 (where the farmer suffers no crop losses), and notably under-compensates the farmer at values around 150, when his losses are very high.

The relevance of the above problem hinges on farmers effectively facing disparate loss functions, and we argue that this is indeed the case. One important aspect—key in our study area—which can greatly induce heterogeneity in farmers' loss exposure to weather realizations are differences in the mix and timing of crops. Even when this is not the case, such as in more homogeneous contexts where farmers generally plant the same crops at roughly the same points in time, there are other potential sources of heterogeneity. For instance, farmers may rely on alternative farming practices that affect the sensitivity of their crops to weather (climate-smart agriculture practices can improve a crop's resilience to certain natural hazards), they may have access to different types of mechanisms to cope with weather risks (specialized machinery, access to and type of irrigation), or they may use different seed varieties designed for managing particular risks (drought-tolerant or pest-resistant seeds). In addition, other, more structural characteristics, may greatly affect a farmer's exposure to risks, such as soil type, slope, proximity to certain geographic features that mitigate or exacerbate the effects of weather, among many others.



Panel A. Individual Composite Product



Panel B. Loss Functions and Composite Product

Figure 4.1 — Insurance Coverage under Composite Product Scheme

This chapter pursues two main objectives: (i) to describe the ways and extent in which farmer heterogeneity affects their demand for agricultural insurance, and (ii) to assess the value of providing a *flexible* index insurance product for hedging against agricultural risks.

For this, we analyze a set of index insurance products covering against excess rainfall recently marketed in Uruguay that were specifically designed to accommodate this heterogeneity. The idea behind these products—originally proposed by Hill and Robles (2011)—is to provide an array of so-called *insurance units* independently covering against specific, well-defined risks, which an agent can then freely combine into an optimal basket that directly applies to his particular risk profile. In order to illustrate this conceptually, Panel (a) in Figure 4.2 shows two such hypothetical units: insurance unit #1 pays $500 if the rainfall index value is at or above 120 and nothing otherwise; insurance unit #2 does so at value of the index above 150. Panel (b) shows how combining these units into different portfolios can allow farmers with risk profiles A and B (discussed above in Figure 4.1) to manage their exposure. By purchasing 1, 1, and 2 of, respectively, insurance units #1, #3, and #5, the farmer with loss function A can construct an insurance portfolio that roughly protects him against his potential crop losses. In turn, loss function B can be hedged by purchasing 1, 2, 1, and 1 of, respectively, insurance units #2, #3, #4, and #5.[56]

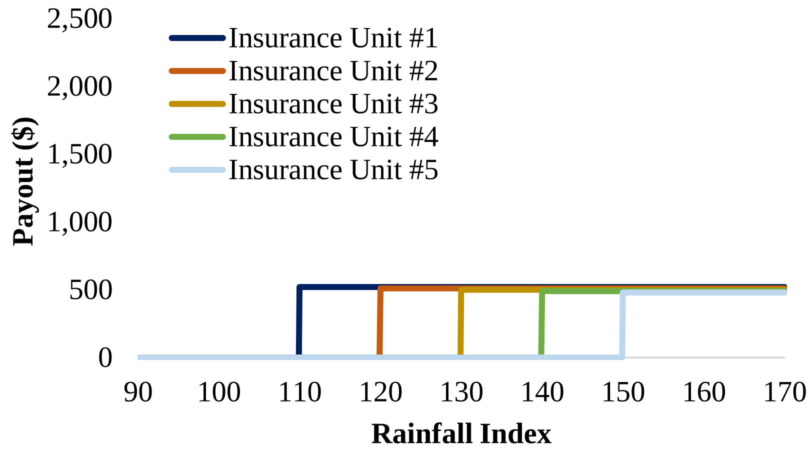
In our study context, these insurance units separately covered the risk of excess rainfall during different time periods (calendar months) and for different degrees of intensity (different levels of the tail of the rainfall distribution). During the marketing efforts of the product, farmers were thus purposely encouraged to purchase a *portfolio* of these insurance units that best suited their individual needs. The demand for the product was reasonable in light of other rollouts of index insurance products in developing countries and of existing agricultural insurance penetration rates in the region. In particular, 1,088 insurance units were purchased, covering around 15% of the total horticultural hectares in the region. Interestingly, we observe significant heterogeneity in farmers' insurance portfolios. This heterogeneity is evident even among farmers growing a similar set of crops, indicative of other important sources of heterogeneity in a farmer's demand for insurance.

Due to the nature of our purchase data, where a single farmer can purchase multiple types of products and multiple units of each of these types (a problem known as *multiple discreteness*), there are no existing econometric methods that can directly tackle the estimation of reduced-form demand equations.[57] Traditional discrete choice methods model the agent's choice of one out of two product types. Extensions such as multinomial or multivariate models can, respectively, handle single choices out of many mutually-exclusive types or multiple correlated binary choices, but there are no existing discrete choice methods that can deal with multiple discreteness. On the other hand, demand equation type models focus on modelling demand shares or, more generally, the choice of (correlated) continuous quantities over a number of different categories of products, also resulting inappropriate for multiple discreteness problems. Moreover, the approaches above rely on linearity-in-parameters and overall restrictive error-distribution assumptions—sometimes inadequate for handling the non-linearities found in risk management applications—, in addition to requiring the estimation of a large number of

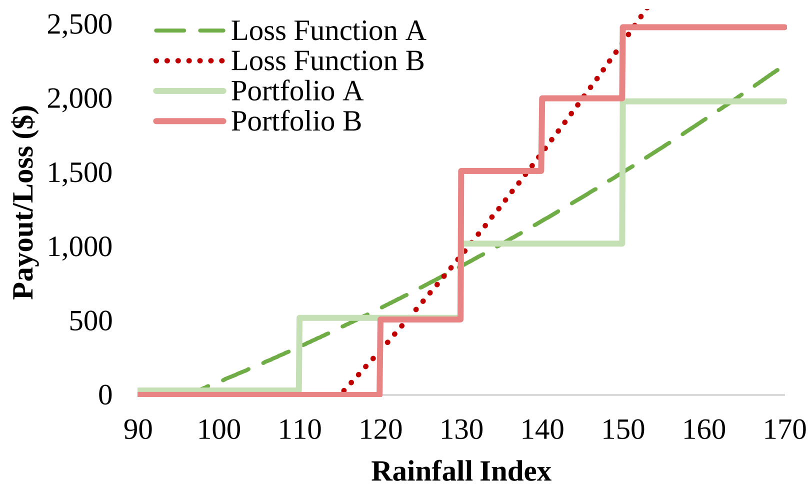
[56] Of course, the step-wise insurance coverage from these portfolios does not perfectly correspond with the hypothetical potential losses. In practice, however, this problem can be solved by allowing for additional flexibility through more types of insurance units or through smaller versions of composite products that can take into consideration the continuous nature of crop losses.

[57] Multiple discreteness can be naturally found in a large number of contexts. As summarized by Hendel (1999), *"Although pervasive, multiple-discrete choice problems have received little attention because of their complexity."*

parameters to allow for flexible specifications of the data. Finally, none of these approaches allow to assess the effect of counterfactual insurance schemes on farmers' welfare.



Panel A. Individual Insurance Units



Panel B. Portfolios of Insurance Units

Figure 4.2 — Insurance Coverage under Insurance Units Scheme

Closer to our context, there is a relatively small strand of the industrial organization literature that has dealt with this type of problems when analyzing multi-unit purchases of different varieties of consumer products. For instance, Kim, Allenby, and Rossi (2002) focus on households' choice of yogurt flavors and use a modelling strategy that explicitly relies on households' taste for variety, an approach that is however hardly adaptable to an insurance context. Dubé (2004) analyzes households' choices for different sizes and types of carbonated soft drinks, modelling the purchased basket as the result of anticipating an unobserved (and unknown) number of future consumption ocassions. His model is based on Hendel (1999) analysis of firms' demand for personal computers, who models firms' choices by considering the match between computer attributes and different unobserved potential tasks that the firm requires computers for.

Since the various attributes of our differentiated insurance products are directly related to their risk management characteristics, we choose to model the risk management problem faced by farmers explicitly through the estimation of a structural model. In particular, we contemplate a farmer's choice of insurance portfolio in the context of an expected utility theoretical framework, where a farmer chooses the set of insurance products that maximizes his next-period expected utility in the presence of uncertain future crop yields and rainfall realizations. To our knowledge, this chapter is the first one to provide theory-based empirical estimates of real-world farmers' risk management behavior.

With the model in place, we then introduce different sources of farmer heterogeneity by modifying various aspects of the farmer's decision-making problem. We seek out to explore the following sources of heterogeneity: (1) Crop composition; (2) Planting dates; (3) Soil type; (4) Product understanding; and (5) Distance to the weather station. We focus on these specific sources because they are particularly important in our context, with other potential aspects of heterogeneity (such as geographic features, coping mechanisms, or differences in equipment) either not having much relevance in this study area, not being available for the risk of excess rainfall, or not showing enough variability across farmers. We rely on data collected from farmer-level surveys conducted by the study team before and after the 2014-15 harvest season. Each of these sources is introduced together with one or more relevant free parameters mediating their effect. These parameters are then structurally estimated using GMM and their statistical significance assessed through regular inference tools. Since our main purpose is to explore how heterogeneity in farmers' characteristics and preferences affect insurance demand, we focus on the sample of farmers who actually purchased insurance.

The general process can be described as follows: the model provides a mapping between the parameters accompanying the different sources of heterogeneity and the optimal insurance choices for each farmer in our sample. With these optimal portfolios we can calculate the difference between the predicted and the observed purchased quantities of each of the 12 insurance units. Now, if the model were to perfectly explain the observed purchase decisions for all agents, these differences would be exactly equal to zero. More realistically, we construct empirical moments as the average difference between predicted and observed purchases across all farmers for each type of insurance unit, which should approach zero as the sample size grows if the model were to reflect true behavior. Finally, we solve for the optimal parameters by setting a given distance function of the moments closest to zero.

The above analysis allows us to tease out the significance of each source of heterogeneity in a farmer's demand for insurance. As a final step, the calibrated theoretical model for the demand of insurance provides a framework to conduct welfare analysis, with the aim of obtaining a relative measure of the benefits of offering a flexible insurance product—as opposed to a fixed-structure composite product resembling other index insurance instruments.

Identifying sources of heterogeneity in the demand for weather index insurance is relevant for a number of reasons. First, understanding the nature of heterogeneity can shed light on appropriate policies to address market imperfections. For instance, certain sources of heterogeneity can be regarded as information asymmetries arising in imperfect information contexts, such as misperceptions about the spatial variation in rainfall or about the characteristics of the insurance product being offered. The appropriate response toward these could entail a larger emphasis on educating farmers about these elements when introducing new weather index products. Second, other sources of heterogeneity can be of a more intrinsic nature, such as farmers' crop composition, planting dates, or soil type. Such underlying differences in farmers' risk profiles call for a more careful consideration at the design stage of new insurance products. In particular, an insurance scheme that can directly adapt to the differing needs of heterogeneous farmers seems to be a relevant alternative for a policymaker's toolkit. Finally, the introduction of index insurance has generally suffered from relatively low levels of demand. If farmer heterogeneity were to be a key aspect behind this lack of demand, the marketing of an array of more flexible index insurance products could greatly contribute to their increased use as risk management tools.

The chapter proceeds as follows. Section 4.1 describes the context for our study and the data used in the analysis. Section 4.2 outlines the base expected utility model and Section 4.3 discusses the different extensions to this model that will capture farmer heterogeneity. Section 4.4 explains the procedure and implicit assumptions for structurally estimating the extended model, the results of which are presented in Section 4.5. Finally, we discuss in detail the nature of the different sources of heterogeneity, together with the policy implications of our analysis, in Section 4.6.

4.1 Context and Data

These innovative insurance products were implemented in the context of a project led by the International Food Policy Research Institute (IFPRI), with support from the Uruguayan Ministry of Agriculture (MGAP), and were underwritten and marketed by Banco de Seguros del Estado (BSE), the federal insurance agency. The project's main objective was to cover horticultural farmers' harvest risk from excess rainfall, generally associated to severe losses due to the rotting of crops and the increased difficulty to access the plots. The products were offered commercially to all horticultural farmers in the department of Canelones, the main horticultural region in Uruguay, during the 2014-15 summer harvest season. The marketing of the product was carried out through the existing network of BSE insurance brokers, who held pre-existing relationships with farmers in the area.[58]

Twelve insurance units were offered in total: two degrees of intensity (*Strong* rainfall, equivalent to the 85th historical percentile of the index, and *Extreme* rainfall, equivalent to the 95th percentile) for each of the six months of the horticultural harvest season (November through April). The index was chosen as the maximum accumulated rainfall over any consecutive ten-day period during a given calendar month.[59] The triggers were selected considering the 85th and 95th percentiles of the historical index values by coverage month (see first two columns of Table 4.1 for trigger values). Each insurance unit promised to pay 500 U.S. dollars (about the cultivation costs of one-quarter of a hectare) if the realized rainfall index during a specific month was above its coverage month- and degree-specific trigger. The insurance units were priced at actuarially-fair cost plus a 40% administrative markup. Premiums were subsidized in 90% by MGAP, up to a total subsidy of 400 U.S. dollars per farmer after which farmers had to pay the fully unsubsidized amounts of $35.70 for each Extreme coverage unit and $107.15 for each Strong coverage unit (regardless of the coverage month).

Table 4.1 — Insurance Units Triggers and Purchases by Coverage Month and Degree – 2014-15 Season

	Trigger		Num. of purchased units			Num. of purchasing farmers		
Month \ Type	Strong	Extreme	Strong	Extreme	Total	Strong	Extreme	Total
November	130	161	28	81	109	7	11	15
December	106	114	44	269	313	17	34	47
January	113	224	63	98	161	19	22	38
February	122	144	74	197	271	24	35	55
March	178	269	42	99	141	14	21	34
April	137	207	11	78	89	6	12	18
Total	-	-	262	822	1,084	63	79	128

We count with administrative data on insurance purchases provided by BSE, which identifies the number of insurance units purchased by coverage month and degree for every farmer (see Table 4.1). These data represent our main focus of interest, which we will attempt to match using the predictions from the calibrated theoretical model. A total of 1,084 insurance units were purchased by 128 farmers to cover their horticultural crops

[58] These relationships exist since farmers in our sample generally purchase vehicle, property, life, and other types of insurance. Agricultural insurance, on the other hand, is rare, except for hail insurance which is relatively common.

[59] Since we did not count with sufficiently-dissagregated, historical yield data to identify the optimal index, we relied on substantial qualitative work including focus groups with farmers and semi-structured interviews with horticultural experts.

from excess rainfall during the 2014-15 summer season.[60] Overall, it can be seen that farmers lean towards the Extreme product, which pays out if the cumulative rainfall index is above its 95th historical percentile. In addition, December and February seem to be the two most demanded coverage months. In general, though, the table shows that the purchases of insurance units are well distributed across all products offered.

To get a sense of the level of portfolio heterogeneity at the farmer level, Figure 4.3 shows the array of individual portfolios purchased by farmers. In particular, each vertical line in the figure represents the purchases of one single farmer; with the shapes and their vertical position indicating, respectively, the coverage degree and coverage month purchased. The figure shows that farmers not only choose a variety of different insurance units to construct their portfolio, but also that these portfolios differ considerably between farmers. While the purpose of the empirical model below is to formalize the channels through which this heterogeneity in portfolios arises, this graphical evidence already supports the hypothesis that the flexibility provided by our insurance scheme is indeed relevant and welcomed by farmers.

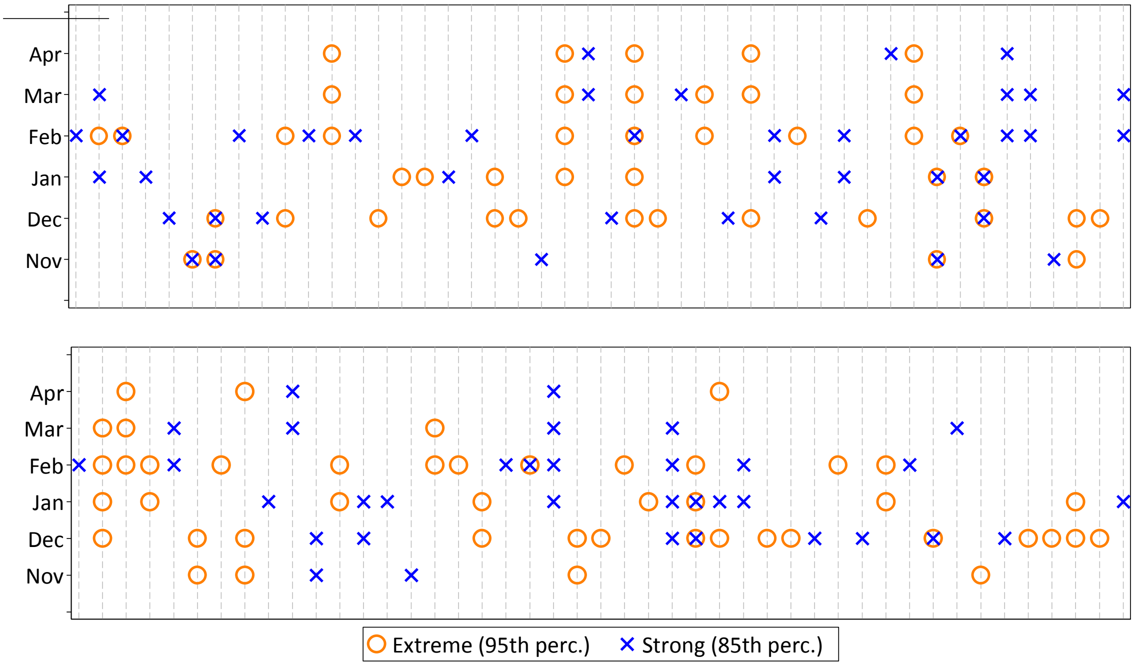


Figure 4.3 — Farmer-Level Portfolios of Insurance Units

Our second main source of data are two farmer-level surveys conducted by a local survey firm before and after the 2014-15 harvest season. These surveys provide us with detailed data on general farmer demographics, past agricultural inputs and outputs, area planted under each crop, and other plot and soil characteristics. In addition, we elicited

[60] Insurance units for October were also available. However, due to administrative delays the marketing of the product was not able to begin in time for farmers to consider the purchase of these units, since a rule was in place that forbidded sales of insurance units within two weeks of the start of the coverage month. As a result, only four units for October were sold, which we decide to exclude from the analysis.

data on farmers' risk aversion, attitudes towards insurance, and expectations about rainfall. Table 4.2 shows summary statistics of general farmer characteristics (including the variables that will be used in the empirical analysis) and horticultural profiles. The sample is comprised by the 91 farmers that purchased insurance units who will be used in the analysis.[61]

Table 4.2 — Summary Statistics of Farmer-Level Survey Data

Panel A. General Characteristics

Variable	Mean	St. Dev.	No. of Obs.
Age (in years)	46.8	9.9	91
Years of education	7.7	3.1	91
Farmer is female	0.18	0.38	91
Landholding size (in ha.)	18.3	17.2	91
Horticultural size (in ha.)	6.6	7.2	91
Soil drainage	0.44	0.38	91
Product understanding	1.6	0.6	91
Distance to weather station (in kms.)	10.6	5.2	91
Coefficient of risk aversion	0.78	1.20	91

Panel B. Horticultural Profile

Variable	Mean	St. Dev.	No. of Obs.
Number of horticultural crops	4.8	3.0	91
Hectares cultivated in hort. crops	5.7	6.5	91
Sweet potato	1.7	1.2	43
Onion	2.4	3.1	47
Cantaloupe	0.7	0.5	15
Sweet pepper	0.7	0.6	37
Roma tomato	0.5	0.5	26
Heirloom tomato	1.7	3.0	26
Carrot	2.0	1.7	13
Round zucchini	0.7	0.8	33
Squash	3.4	5.6	41

Additionally, we tap into secondary data sources to calibrate some of the elements of the model. First, we count with historical daily data from the entire rain gauge network of the Uruguayan Meteorological Institute (INUMET). These data span over 30 years in some cases, allowing us to estimate representative, long-term distributional parameters for rainfall. Second, we gathered data on horticultural crop yields from the Southern-Region Horticultural surveys, which collect annual self-reported crop yield information for a large array of horticultural crops from a representative sample of farmers in the southern region of the country (of which the department of Canelones comprises the largest and most important fraction). These data include average crop production, planted area, and yields

[61] This is lower than the total number of farmers who purchased insurance units since some farmers were not interviewed during the baseline survey (even though efforts were made to reach all horticultural farmers in Canelones).

for 32 different horticultural crops and spans 12 summer seasons, from 2001-02 through 2014-15 (survey results are not available for a few of these seasons). Finally, we use price data published by the main market for agricultural products (Mercado Modelo), located in the capital city of Montevideo, which contains daily data on wholesale prices of different varieties and qualities for an array of fruits and vegetables. The use of these data will be made explicit in Section 4.4 below.

4.2 Model

This section describes the theoretical framework characterizing the decision problem of farmers, which will be used as the basis for the empirical analysis of Section 4.4. Overall, the objective of this framework is to provide a simple yet reasonable specification for farmers' decision-making under risk, while avoiding to fully model related aspects of decision-making not regarded as central to the choice of an insurance units portfolio.

4.2.1 Farmer's Utility

A farmer f maximizes expected utility over his own next-period income, $E\left[U_f\left(W_f\right)\right]$, where $U_f(\cdot)$ is a well-behaved utility function with $U_f'(\cdot) > 0$ and $U_f''(\cdot) < 0$. For the sake of simplicity, we assume that farmers share a common utility function though potentially with different parameters α_f (e.g. risk aversion): $U_f\left(W_f\right) = U_f\left(W_f, \alpha_f\right)$.[62]

Next-period income W_f consists of the revenue from selling the harvest of his particular mix of J crops, plus the net income from his insurance purchases: insurance payouts, I_f, minus insurance premiums, c_f; as follows:

$$W_f = \sum_{j=1}^{J}\left(H_{fj}Y_{fj}P_j\right) + I_f - c_f \tag{4.1}$$

H_{fj} is the number of hectares cultivated with crop j, Y_{fj} is the yield of crop j in units per hectare, and P_j is the (expected) market price obtained from selling one unit of crop j.[63] In order to focus on the insurance demand decision we consider the area cultivated under each crop (H_{fj}) and the (expected) market price for each crop (P_j) as predetermined for the upcoming season.[64]

[62] For instance—and most importantly—parameters controlling their degree of risk aversion.

[63] We purposely ignore costs of cultivation, such as the costs of agricultural inputs, since these would not influence the portfolio choice decision that constitutes the main interest of our analysis.

[64] In doing this we abstract from important considerations that have been traditionally analyzed in the crop insurance literature, such as the impact of insurance on production decisions. However, while restrictive, this aspect of the model aligns well with the reality of our empirical application, where the insurance product was announced closely before or after planting decisions had already been made by farmers.

4.2.2 Insurance Units

The available insurance units protect against *excess* rainfall, which can damage crop yields around harvest. Each insurance unit is linked to the maximum cumulative rainfall observed over any 10-consecutive-day period in a particular coverage month (the *index*). These are available for a total of six months (November through April) and for two degrees of rainfall intensity: Strong (s) and Extreme (e). In particular, the Strong (Extreme) product for month m pays a fixed amount when the rainfall index in month m is above its historical 85th (95th) percentile:

$$I_{mv} = \begin{cases} 500 \; if \; \delta_m \geq \bar{\delta}_{mv} \\ 0 \quad otherwise \end{cases} \tag{4.2}$$

where δ_m is the realized rainfall index for month m and δ_{mv} is a predetermined trigger for month m and coverage type $v = (s, e)$.

The farmer chooses to purchase a number q_{fmv} for each of the 12 insurance units, at a predetermined price of c_v per policy, which varies with the coverage degree v but is the same across coverage months m ($c_{mv} = c_v \; \forall m$ and $c_{strong} > c_{extreme}$).[65,66] The total payout (I_f) and cost (c_f) of farmer's f insurance portfolio are then:

$$I_f = \sum_{m=1}^{6} \sum_{v=s,e} q_{fmv} \cdot I_{mv} \tag{4.3}$$

$$c_f = \sum_{m=1}^{6} \sum_{v=s,e} q_{fmv} \cdot c_v \tag{4.4}$$

4.2.3 Decision Problem

Farmer's f decision problem thus consists of maximizing expected utility over his own next-season income by choosing q_{fmv}:

$$\max_{q_{fmv}} E\left[U_f\left(W_f\right)\right] = E\left[U_f\left(\sum_{j=1}^{J}\left(H_{fj}Y_{fj}P_j\right) + I_f - c_f\right)\right] \tag{4.5}$$

The uncertainty in the model arises from two sources: rainfall indices and crop yield realizations. As described above, the insurance units depend on the realizations of six random variables (rainfall indices across six calendar months). Crop yields, in turn, depend on the realization of J additional random variables. These $J + 6$ random variables are

[65] Given that premiums do not vary across farmers, we are not able to analyze common topics in the literature such as the price elasticity of farmers' demand for insurance.

[66] In addition, farmers face a non-linear budget set which consists of a 90% premium subsidy up to a total subsidy amount of $400, and lack of subsidy thereafter. While this is included in the empirical model, we exclude the details from the equations for the sake of simplicity.

jointly modelled through a parametric multivariate distribution, the realizations of which represent the different *states of the world* under which the farmer forms the expectation for his next-period income. Importantly, note that a (negative) relationship between Y_{jf} and I_f should exist in the model for an agent to seek insurance; we discuss this relationship in the next subsection.

An analytical solution of this problem is very difficult to obtain when normality cannot be assumed for the distribution of crop yields and rainfall indices. In addition, even if normality were assumed, the non-linearity of the insurance units' payout would require to integrate the multivariate distribution over 3^6 regions (three support regions for insurance payouts within each month). To solve the model, thus, we rely on simulation: $J + 6$ random realizations are jointly drawn from this multivariate distribution K number of times, which determine the support of the expectation operator. Given the parameters in his utility function, the farmer then chooses the 12 q_{fmv} to construct a portfolio of insurance units that maximizes the expected utility of his next-period income. This maximization is done numerically. Since a farmer cannot purchase a fraction of an insurance unit, the optimal portfolio is discretized by direct evaluation of his expected utility function at every combination of the upper and lower integer bounds of the optimal portfolio. Under concavity of the utility function, the combination that maximizes this expected utility function is thus the optimal discrete portfolio.

The general argument for the farmer to choose positive amounts of insurance is well-known: under a concave utility function insurance smoothes a farmer's income across states of nature, thus a farmer will choose to give up a certain amount (insurance premium) in order to receive state-contingent income (insurance payouts) in those states with a high marginal utility of consumption (i.e. the states where loss of income occurs).

4.2.4 Calibration

In order to complete the formulation of the model we must determine the functional form and calibrate the two sources of uncertainty faced by the farmer: rainfall and crop yields.

First, we assume that the rainfall indices (that is, the maximum amount of rainfall accumulated over ten consecutive days in a given calendar month) for the six months, $\Delta = [\delta_1, \dots, \delta_6]$, can be represented by a multivariate lognormal distribution: $\Delta = e^{R \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$, with $\boldsymbol{\mu}$ a 6x1 vector of means and $\boldsymbol{\Sigma}$ a 6x6 variance-covariance matrix. While arbitrary, this family is chosen because of its relative simplicity and strictly positive support with a long upper tail (capturing the occurrence of excess rainfall events, which is the key of our analysis).[67] This distribution is calibrated using 30 years of historical rainfall data from the most representative weather station in our study area: Aeropuerto Carrasco. It is worth

[67] Goodness of fit tests cannot reject the null hypothesis for joint log-normality of the historical series.

noting that in doing so we are implicitly assuming that farmers form rational expectations over past information.

Second, we assume that future crop yields can be represented through a multivariate normal distribution. This is arguably a stronger assumption. Our justification for this choice is twofold: (1) Joint normality is found plausible by a strand of the empirical literature on crop yields;[68] (2) This assumption simplifies the empirical treatment when including a large number of crops. This distribution is in turn calibrated using historical yield data from the annual horticultural surveys conducted in our study area.[69]

Finally, in order for the insurance to be relevant, we must account for the dependence of yields on rainfall. Unfortunately, we do not count with sufficiently long and dissaggregated data on both yields and rainfall across our study area to empirically estimate said relationship. We sort these obstacles by assuming a certain dependence structure that relies on a number of free parameters which will be estimated later as described in Section 4.4. In particular, we make the following assumptions in the model about the nature of dependence between crop yields and the rainfall indices, where farmer f's yields for crop j, Y_{fj}, depend on the month m rainfall index, δ_m, as follows:

$$Y_{fj} = Y_j \cdot \left[1 - \sum_{m=1}^{6} \theta_{jm,f} \cdot max(0, \delta_m - h)\right] \tag{4.6}$$

The term in brackets in Equation 4.6 acts as a multiplier that reduces "potential" yields Y_j according to the realizations of rainfall. In particular, we allow yield to decrease linearly with rainfall when the rainfall index is above a certain threshold h.[70] When all the rainfall indices are below the threshold h, this term is equal to 1 and yield is thus optimal; but when one or more of the rainfall indices is above h, the term takes a value below 1 and the maximum potential yield is reduced by a certain fraction. The $\theta_{jm} > 0$ capture the linear

[68] See, for example, Just and Weninger (1999). Another strand challenges these findings, importantly Ramirez, Misra, and Field (2003). Overall, however, there does not seem to be a consensus on the most appropriate distributional form for crop yields.

[69] It is important to clarify that these data represent the *average* crop yield in the area, thus concealing the true amount of variation in individual farmer yields. For the purposes of our empirical model, we do not consider this to be a first order limitation. Rather, since single-farm outcomes could in principle be much worse than an entire area's outcomes, this assumption limits the true loss distribution of farms and would bias our model's predictions towards acquiring less insurance.

[70] We assume h to be fixed across coverage months, as the risk of excess rainfall in horticulture is generally associated to the rotting of crops from being underwater or to reduced accesibility to the field for harvesting, both problems that depend on the amount of rainfall and terrain characteristics but not on the specific crop planted. However, it is important to note that the effect of excess rainfall on yields is not constant, since the θ_{jm} can differ across crops, capturing the fact that some crops are more resilient than others to being exposed to excess water.

negative effect of excess rainfall (accumulated rainfall above the threshold h) on crop yields and are constructed as

$$\theta_{jm,f} = \theta_j \cdot exposure_{jm,f} \tag{4.7}$$

where $exposure_{jm,f}$ indicates the fraction of month m during which crop j is exposed to excess rainfall. Since the qualitative analysis indicates that the main risk of excess rainfall occurs around harvest, we further assume that a crop's exposure to large amounts of rain (i.e. high δ_m) takes place 60 days (two months) before the expected date of harvest, otherwise exposure to δ_m is null. Note that exposure varies according to a farmer's choice of planting a given crop j and to the farmer's choice of planting dates. Finally, the θ_j are J free parameters that are estimated as explained in Section 4.4.

4.3 Sources of Heterogeneity

In this section we discuss the effect of alternative sources of farmer heterogeneity on demand for insurance in the context of our model. We extend the model in Section 4.2 to accomodate for each source and, in Section 4.4, empirically estimate these parameters to assess whether they fit the insurance demand data.

4.3.1 Crop Composition

Naturally, this is an important aspect of heterogeneity and its inclusion in the model should explain a substantial amount of variation in the observed insurance portfolios. In order to assess whether the purchased insurance portfolios are related to farmers' crop composition we compare the benchmark model in Section 4.2 with a modified version which considers that all farmers face the same mix of crops. In order to maintain farmer size constant, we calculate the total number of hectares in every crop and determine shares for each of them, which we then apply to a farmer's total landholding size. If the model taking into consideration a farmer's actual crop composition results in a better overall fit of the observed insurance portfolio data, this would provide evidence in support of the relevance of this source of heterogeneity.

4.3.2 Planting Dates

Another direct source of heterogeneity in farmer's risk profile is related to differences in farmers' timing in the planting of crops. Differences in planting dates could stem from common strategies such as staggering sowing to reduce specific risks at later stages of crop growth or to take advantage of seasonal price variation (e.g. selling early or late in the market, when supply for that crop is low). Other important underlying causes behind differences in planting dates are preference for seed varieties with longer or shorter durations.

In the context of the model, this is taken into consideration by allowing the $exposure_{jm,f}$ variable described in Equation 4.7 to change according to the full set of planting dates for each crop reported (ex-post) by the farmer. In other words, for each reported planting date d, we calculate $exposure^{d}_{jm,f}$ as the fraction of month m during which crop j—planted in date d—is within two months of its harvest date (and thus exposed to excess rainfall). In order to obtain $exposure_{jm,f}$ for the overall relationship between crop j and the different months m, we simply calculate the mean of $exposure^{d}_{jm,f}$ across all planting dates d reported by farmer f.[71]

4.3.3 Soil Type

As mentioned above, the risk of excess rainfall is generally interdependent with terrain characteristics such as slope or poor drainage. Since our study area is homogeneously flat, slope does not seem to be a relevant factor in our context. Soil drainage, however, may be a relevant factor in a farmer's demand for insurance. In particular, soils with poor drainage are more prone to water accumulation after heavy rains, while soils with better drainage can withstand higher levels of rainfall before suffering the same consequences. In terms of insurance demand, a farmer with the former type of soil would thus be more prone to buying insurance that protects him against the occurrence of lower amounts of rainfall.

In the context of our model, we introduce this factor by modifying the threshold parameter h in Equation 4.6, which controls the amount of cumulative rainfall above which crop yields start to be negatively affected. In particular, we impose the following structure:

$$h = h + \beta_{soil} \cdot X_{soil} \tag{4.8}$$

where $X_{soil} \in [0,1]$ captures the proportion of farmer's f horticultural area comprised of heavy soils with poor drainage and β_{soil} is a parameter to be estimated. In order to assess the relevance of this characteristic on demand for insurance we can then test the hypothesis $H_0\colon \beta_{soil} = 0$.

4.3.4 Product Understanding

There is a large literature that identifies farmers' limited product understanding as an important factor in their reduced demand for index insurance (e.g. Cole et al., 2013; Sibiko, Veettil, and Qaim, 2016; among others). It thus seems appropriate to evaluate this aspect of farmer heterogeneity within the context of our model.

While a limited understanding of the insurance product could translate into different misconceptions within the farmer's decision making problem, we summarize its effect in the model as an additive term to the insurance's payout, I_f. In particular, we posit that limited understanding is reflected in the farmer's decision problem as a lower payout in those states of nature where the insurance product pays:

---

[71] Ideally, we would want to weight this average by the surface of crop j planted in date d . However, we do not count with detailed information on surface cultivated by planting date and thus must resort to this simplyfing assumption.

$$I'_f = I_f + \beta_{cons} + \beta_{und} \cdot X_{und} \tag{4.9}$$

where X_{und} is a dummy variable that takes the value of one if the farmer's response to the question *"Do you know how the insurance product works?"* was *"A little"* or *"No"*, and β_{und} a free parameter (capturing the effect of lack of understanding of the insurance product).

In addition, we allow for an additional parameter, β_{cons}, that affects all farmers' perception of the insurance payout (regardless of their level of understanding). This parameter can be interpreted as the distrust that is normally observed when introducing a completely brand new type of insurance, different from any other product previously experienced. This parameter also serves by way of a constant in a linear model, bringing the average level of insurance demand to its observed counterpart.

4.3.5 Distance to the Weather Station

As our weather index insurance product pays according to the rainfall recorded at the reference weather station (i.e. the weather station closest to the farmer's plots), insurance becomes less appealing for farmers farther from the weather station. This effect is known in the index insurance literature as spatial basis risk.[72]

In order to capture this element in the model, we introduce a second set of rainfall indices, $\Delta^{WS} = e^{R \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$, which captures rainfall at the weather station and thus determines insurance payouts. The original set of rainfall indices, Δ, is then interpreted as reflecting rainfall at the farmer's plot (and thus associated to the loss in crop yields under excess rainfall). The two sets of random variables are identically distributed and correlated, with the degree of correlation between them determined by the distance (in kilometers) of the farmer's plots to the weather station, X_{dist}, as follows:

$$Corr(\Delta, \Delta^{WS}) = (1 + \beta_{dist} \cdot X_{dist}) \tag{4.10}$$

with $\beta_{dist} < 0$ a free parameter. In other words, if the farmer's plots are located at the same place where rainfall is recorded ($X_{dist} = 0$), the correlation between the two sets of random variables is exactly 1 and thus realizations from them are equivalent. This would result in the rainfall indices affecting a farmer's crop yields being the same as those determining the payout of the insurance product, thus eliminating spatial basis risk. Now, at positive distances from the weather station, the correlation between these two sets of

[72] This is related to the general perception of spatial variation in rainfall being very high. While this is mostly true for average rainfall events—which tend to be short and localized—, extreme rainfall events (e.g. large storms) are of a more reginal nature and as such highly spatially correlated. Since excess-rainfall index insurance products pay around the upper tail of the rainfall distribution, spatial basis risk is mitigated. Nevertheless, farmers seem to consistently overestimate the negative effect of distance on correlation in rainfall, dampening their demand for insurance. See Chapter 3 for more details.

rainfall indices would fall linearly with distance, and the rainfall affecting crop yields would only be positively correlated—not identical—to the rainfall determining payouts.

4.4 Structural Estimation

This section describes the structural estimation of the model parameters using generalized method of moments (GMM). The set Γ of free parameters for the full model including all extensions is $\Gamma = [\theta_j, \beta_{cons}, \beta_{und}, \beta_{soil}, \beta_{soil}]$, with the J θ_j parameters governing the degree of dependence between excess rainfall and crop yields introduced in Section 4.2 and the β parameters associated with each of the sources of heterogeneity discussed in Section 4.3.

The aim of the estimation process is to find the parameters Γ^* that predict optimal insurance portfolios that are closest to the actual portfolios purchased by farmers. To do this, we use as moments the difference between the number of predicted and actual insurance units for each of the 12 coverage month (November through April) and degree (Strong and Extreme 10-day cumulative rainfall) combinations. We carry out the estimation by two-step GMM, using as instruments a constant and the fraction of land owned by the farmer (thus assuming that, after taking into account all aspects related to the farmer's risk profile in the model, the fraction of land owned by the farmer should be systematically unrelated to any remaining differences in purchases). In the second step, we calculate the efficient weighing matrix following Hansen (1982), which results in well-known asymptotic distributions for the estimates.

The model's structure provides a mapping from farmer characteristics and model parameters to farmers' optimal choices of q_{fmv}. For a given value of the parameters, we can identify the portfolio of insurance units that maximizes each farmer's expected utility. As it is impractical to obtain a closed-form analytical solution for the farmer's problem, we rely on simulation.[73] For this, we draw K joint random realizations for all the random variables in the model (rainfall indices plus crop yields), and calculate each farmer's utility at every one of these *states of the world*.[74] A farmer's expected utility is the simple average of his utility across all (simulated) states of the world. We can then identify, for each farmer f, the insurance portfolio (q_{fmv}^{opt}) that maximizes his expected utility given the parameters.

The GMM estimation proceeds as follows: given values for all farmer characteristics and initial (arbitrary) parameters $\Gamma^{(1)}$, a set of farmer choices $q_{fmv}^{(1)}$ are obtained for all F

[73] An analytical solution would entail solving truncated cumulative probabilities of 6 lognormal (rainfall indices) and 9 normal (crop yields) distributions over 3^6 regions of integration (since the payouts of the insurance product are non-continuous at the Strong and Extreme triggers for each of the 6 months).

[74] See Pakes and Pollard (1989) for a formal treatment of estimation under simulated moments. Following their suggestion, we keep the set of random draws from the stochastic elements of the model constant across GMM iterations (i.e. alternative parameter sets) in order to maintain the equicontinuity of the objective function. See also Low and Meghir (2017).

farmers. These choices are in turn compared to the true observed choices q_{fmv}^{T} by calculating the sample moments (as the average of each individual moment across all farmers) and the GMM objective function is computed for a given weighting matrix. The procedure continues in this fashion, with a new set of model parameters $\Gamma^{(s)}$ at each iteration s, until the objective function cannot be decreased any further (i.e. where the predicted optimal choices are found to be closest to the true observed choices, under a given distance measure and precision). The resulting parameter set Γ^{*} is thus the one that achieves the closest fit to the observed insurance purchase data.

We use $K = 300$ realizations to represent the different states of nature in the farmer's decision problem. As for the missing elements of the model, we assume a constant relative risk aversion (CRRA) utility function, $U(W_f) = \frac{W_f^{1-\gamma_f}}{1-\gamma_f}$, with relative risk aversion parameter γ_f calibrated from a (non-incentivized) modified Holt and Laury (2002) risk aversion elicitation game carried out during the baseline survey, following Balsa, Gandelman, and González (2015). We work with nine crops: sweet potato, onion, cantaloupe, sweet pepper, roma and heirloom tomato, carrot, round zucchini, and squash. Based on data from our baseline survey, these crops account for almost 90% of the area cultivated under horticultural crops and include more than 80% of the horticultural farmers in the department of Canelones. Moreover, this group of crops represents two thirds or more of (i) the total horticultural hectares and (ii) the number of horticultural crops cultivated for the large majority (87%) of farmers in our sample.

4.5 Results

We first discuss the results from the parameter estimation, discussing their statistical as well as economic significance. Then, we describe counterfactual policy experiments carried out with the estimated model that allow us to quantify the value of providing a flexible insurance scheme in the form of insurance units.

4.5.1 Model Estimates

Table 4.3 shows two-step GMM results for the full model using the *average* planting dates (as determined from qualitative local expert knowledge).

Overall, the results point to a significant effect of all sources of heterogeneity considered in the model on insurance demand. First, differences in farmers' understanding of the insurance product seem to be a strong predictor of their purchase decisions. Understanding little or nothing about how the product works is equivalent to reducing the payout of the product in around $200 (out of $500), thus shrinking the product's value by about 40%. Second, a farmer with good soil drainage in all of his plots is much more tolerant of higher rainfall amounts as indicated by his insurance purchases (thus leaning his purchases towards the Extreme product). In the context of the model, having good drainage in 100% of the plots is equivalent to a substantial higher threshold above which excess rainfall affects crop yields: 155 millimeters compared to 100 millimeters for farmers with bad soil drainage. Third, distance to the weather station matters, farmers located farther away from the product's reference weather station do indeed perceive rainfall realizations at their plot as different from rainfall realizations at the weather station. The coefficient in the table indicates that being one extra kilometer away from the weather station is equivalent to reducing the correlation between the rainfall distribution at the plot and at the weather station by around 0.069. To get a sense of the order of magnitude of this estimate, such an effect size would indicate that farmers' perceive a distance of around 14 kilometers as the maximum distance between two locations at which there can be any correlation between their rainfall patterns. Finally, the parameter capturing overall distrust in the insurance product (also referred to previously as the constant term) is indeed negative and statistically significant, diminishing the overall value of the product for all farmers in about a third. Note that the presence of this constant effect could in reality be indicative of the presence of other type of farmer characteristics and/or beliefs affecting the general level of insurance demand but unnacounted for in our model.

Table 4.3 — Model Parameter Estimates using Average Planting Dates

Panel A. Parameter Estimates

Parameter	Estimate	Standard Error	Z-Statistic
Constant: β_{cons}	-168.516	29.759	-5.663
Understanding: β_{und}	-200.663	33.150	-6.053
Soil drainage: β_{soil}	54.581	2.638	20.689
Spatial basis risk: β_{dist}	-0.069	0.002	-39.215
$\theta_{sweet\ potato}$	0.0023	0.0002	9.741
θ_{onion}	0.0015	0.0004	3.826
$\theta_{cantaloupe}$	0.0015	0.0004	3.767
$\theta_{sweet\ pepper}$	0.0015	0.0004	3.633
$\theta_{roma\ tomato}$	0.0015	0.0003	5.113
$\theta_{heirloom\ tomato}$	0.0015	0.0005	2.795
θ_{carrot}	0.0027	0.0004	6.732
$\theta_{round\ zucchini}$	0.0015	0.0004	3.827
θ_{squash}	0.0015	0.0004	4.324
Objective function	0.4480		
Number of obs.	91		

Panel B. Test for Over-Identifying Restrictions

	Actual value	5% value	p-value
J-statistic	40.7717	19.675	0.000

The final value of the objective function—derived from the sample moments and the optimal weighting matrix—is 0.448. Panel B shows the results for the test of overidentifying restrictions, with a test statistic equal to 40.772 (equivalent to the objective function times the number of observations). Under the null, this statistic is distributed chi-squared with 11 degrees of freedom (24 moment restrictions minus 13 parameters): $\chi^2_{0.05}(11) = 19.675$. Thus, the null hypothesis of the model being an accurate representation of reality is rejected.

Table 4.4 shows instead results for the model using the *actual* planting dates, as collected directly from farmers during the follow-up survey. The estimated parameters for farmers' understanding of the product, type of soil drainage, and sensitivity to the distance to the weather station are lower in magnitude than those estimated when using average planting dates, though they are still strongly statistically significant and economically important. The general distrust parameter is also lower (though still statistically significant), seemingly indicating that this alternative model is indeed capturing the variability in insurance purchases more accurately.

Interestingly, the model using actual planting dates in Table 4.4 results in a considerably lower objective function than that using average planting dates (Table 4.3).

4.5.2 The Value of Flexibility

The above results show that certain farmer characteristics do actually matter for their choice of insurance portfolio. The question remains, however, of whether farmers are better-off by being able to construct their portfolios as opposed to buying the pre-designed portfolios implicit in a one-size-fits-all, composite index insurance product. It is natural to think that under the latter scheme, by not being able to adapt the insurance to their particular risk profile, most farmers would indeed be worse-off (except perhaps for those farmers whose risk profile is close enough to the risk profile implicit in the composite product). In addition, it would be important to quantify the relative improvement in farmers' welfare from providing flexibility; if this improvement were too small to be relevant, the additional costs associated to implementing a flexible system would not be warranted. Fortunately, the structural model estimated in the previous section provides us with a framework where to explore these questions.

With this objective in mind, we carry out policy experiments where we take as given the model's structure together with the optimal parameters estimated above (Table 4.4) and allow farmers to choose the optimal combination from a hypothetical set of counterfactual insurance products. Then, by comparing each farmer's resulting expected utility under the alternative scenario to their expected utility under the insurance units scheme, we can derive the welfare changes brought about by providing flexibility.

Table 4.4 — Model Parameter Estimates using Actual Planting Dates

Panel A. Parameter Estimates

Parameter	Estimate	Standard Error	Z-Statistic
Constant: β_{cons}	-143.755	27.398	-5.247
Understanding: β_{und}	-124.376	18.867	-6.592
Soil drainage: β_{soil}	46.754	3.009	15.538
Spatial basis risk: β_{dist}	-0.065	0.002	-30.790
$\theta_{sweet\ potato}$	0.0017	0.0003	5.554
θ_{onion}	0.0013	0.0003	4.673
$\theta_{cantaloupe}$	0.0013	0.0004	3.198
$\theta_{sweet\ pepper}$	0.0013	0.0005	2.713
$\theta_{roma\ tomato}$	0.0030	0.0005	6.448
$\theta_{heirloom\ tomato}$	0.0027	0.0009	2.975
θ_{carrot}	0.0018	0.0003	6.138
$\theta_{round\ zucchini}$	0.0030	0.0004	7.331
θ_{squash}	0.0030	0.0002	12.576
Objective function	0.3687		
Number of obs.	91		

Panel B. Test for Over-Identifying Restrictions

	Actual value	5% value	p-value
J-statistic	33.5507	19.675	0.000

We consider two types of alternative composite insurance products. The first type involves a frequent choice in index insurance implementations, where products are designed for a particular crop, using the best available knowledge about the timing and effect of the risks under consideration. The second type is a hybrid between this model and the flexible products rolled-out in Uruguay, where an insurance product is offered for each calendar month but no flexibility is provided in terms of the particular payout function with respect to the index. In other words, under this second type of alternative products, a farmer is able to purchase insurance for individual months (according to his own assessment of his crops' exposure to each of them), but has to put up with a particular payout structure, being thus unable to adapt insurance payouts to his specific exposure torainfall intensity.

For both of these types, we maintain the index chosen (i.e. accumulated daily rainfall over 10 consecutive days) and consider the most common payout structure seen in crop insurance products for smallholder farmers around the world. This structure relies on three elements (or parameters): the strike, the tick, and the exit. In the case of excess rainfall, a typical product would pay nothing when the value of the index is under the strike and would pay the entire coverage amount if the realized value of the index is above the exit value (i.e. extreme rainfall amounts were recorded). For realized values of the index between the strike and exit, the insurance product would pay the product between the tick and the difference between the index and the strike: $Payout = (Index - Strike)\ x\ Tick$. In order to design the alternative products, we need to decide on the values taken by the strike and exit parameters (the tick can be derived implicitly from the difference between the exit and the strike together with the maximum payout intended for the product, which is fixed at $500 as in the insurance units). For simplicity we determine these values as the percentiles of the historic distribution of the index in our study area. We use two alternative combinations: (1) Strike at the 85th percentile and Exit at the 95th percentile; and (2) Strike at the 90th percentile and Exit at the 99th percentile.

The combinations above result in four sets S of alternative insurance products, which we believe reasonably capture the range of options available to the implementer of an index insurance scheme. Next—and separately for each of these sets—we allow each farmer to select an optimal portfolio of insurance products within the available set and calculate the resulting (optimal) level of expected utility from purchasing this portfolio, EU_f^s. Moreover, from the optimization procedure in the previous section, we count with each farmer's expected utility under an optimal portfolio of insurance units, EU_f^*, and their expected utility under a no-insurance scenario, EU_f^{NI}. Finally, we calculate the welfare improvement (with respect to no insurance being available) under each alternative scenario S, relative to the welfare improvement from having a flexibile insurance units scheme available:

$$\frac{\left(EU_f^s - EU_f^{NI}\right)}{\left(EU_f^* - EU_f^{NI}\right)} \tag{4.11}$$

This is our final welfare measure, where values under 100% indicate that a farmer's welfare under the insurance units scheme is higher than under the alternative, hypothetical set of composite insurance products, while values above 100% indicate that a farmer is worse-off under the alternative scenario.

Table 4.5 shows a summary of results from these exercises. Overall, the value of providing flexibility in terms of an insurance units scheme is substantial. For instance, if the first alternative set of products were available (Set #1), with one insurance product per crop and strike and exit values at the 85th and 95th historical percentiles, all farmers would be worse-off than in the benchmark system of insurance units, with excess expected utilty (over the expected utility achieved under no insurance) around half of that achieved under the more flexible scheme. Certainly, there is a lot of heterogeneity among farmers: for some farmers, the benefit of being able to create a portfolio reflecting their unique risk profile is quite large, allowing them to boost the additional welfare from counting with insurance by about 4 times compared to the alternative set; for other farmers, on the other hand, the welfare derived from either insurance scheme would be almost the same. In terms of the other hypotethical sets of insurance products, the value of flexibility is even greater if one considers the insurance products available in Set #2, which pay under more extreme rainfall conditions than those in Set #1. As for the month-level products considered in Sets #3 and #4, the value of an insurance units scheme is somewhat smaller, though still positive. In these cases, though, a small subset of farmers (18.7% and 28.6%, respectively) would be better-off under the alternative schemes.

Table 4.5 — Results from Alternative Policy Experiments

Alternative products	Description	Average	Median	Min.	Max.	Perc. farmers worse-off
Set #1	Crop-level products Strike: p, Exit: p	57.2%	53.3%	22.0%	97.9%	100.0%
Set #2	One ins. prod. per crop S: 90th pctile, E: 99th pctile	28.2%	25.2%	6.4%	60.4%	100.0%
Set #3	One ins. prod. per month S: 85th pctile, E: 95th pctile	79.1%	84.7%	23.1%	115.7%	81.3%
Set #4	One ins. prod. per month S: 90th pctile, E: 99th pctile	73.9%	72.6%	17.0%	124.8%	71.4%

All in all, the results in this subsection illustrate that there are large benefits to be reaped from implementing an insurance scheme that provides the basic elements for heterogeneous farmers to adapt it to their particular risk needs.

4.6 Discussion and Policy Implications

We describe a flexible commercial index insurance scheme which comprises a number of independent insurance units covering against different aspects of excess rainfall, based on the weather securities approach originally proposed by Hill and Robles (2011). We exploit substantial variation in the insurance portfolios purchased by farmers to provide unique theory-based evidence on how farmers' heterogeneity in risk profiles affects the nature of their demand for insurance. In particular, we provide evidence that differences in crop composition, planting dates, soil drainage, distance to the weather station, and understanding of the product are all significant determinants of the type of insurance portfolio purchased by a farmer. In addition, we show that, under the assumptions and optimal estimated parameters of the model, providing flexibility through an insurance units scheme improves farmer welfare considerably in relation to having provided a range of other hypotethical composite products.

More generally, this chapter illustrates the feasibility of rolling out a real-world system of commercially-backed, flexible insurance units or weather securities. At least in the particular context of horticultural farmers in Uruguay, this system seems to be more appropriate than traditional index insurance systems based on one single, standardized product. Nevertheless, the generalizability of these findings seems plausible for any other contexts with sufficient degrees of farmer heterogeneity.

As discussed above, traditional index insurance products are generally based in the expert design of an optimal insurance policy for a representative farmer or for the expected losses under ideal management conditions. Now, a natural preconception against the flexible insurance approach is that farmers would not be able to construct their optimal portfolios in a way that improves upon an expert's assessment. We show this notion to be unfounded. Farmers do indeed purchase a wide array of different portfolios and, more importantly, do so in a way that is consistent with the underlying risk exposure of their particular crop composition and other individual characteristics related to their individual farming conditions.

Our results have significant policy implications. Farmer heterogeneity matters. The approach of selling index insurance as a single, one-size-fits-all policy is misguided. For all its worth, index insurance cannot be designed to fit all risk profiles. Product flexibility is important and farmers seem to have the ability to adapt it to their needs. However, not all heterogeneity is equal. On one hand, there are certain sources of heterogeneity that are related to information problems, such as limited product understanding or misperceptions about rainfall distributions. In this respect, a flexible insurance system would serve only as a temporary patch; requiring more permanent solutions that tackle the root of the problem such as accompanying the rollout of new insurance with appropriate marketing and campaigns educating about true climatic patterns. On the other hand, other sources of heterogeneity are of a more structural nature, such as the degree of soil drainage a farmer has or his specific choice of crops and planting dates. A flexible insurance system is almost unavoidable in these cases, as it is hard to imagine how to design only a few composite insurance products that could appropriately consider the weather risks of all possible risk profiles.

This product was marketed in a context of better-educated farmers relative to other rural contexts in Latin America, not to mention Africa or Asia. This gives rise to the question of whether farmers' choice of appropriate insurance portfolios would still be attainable in these other contexts. Preliminary evidence in this regard discussed in Hill and Robles (2011) shows that smallholder farmers in Ethiopia do indeed flexibly respond to their risk exposure by purchasing different insurance portfolios. In addition, we venture that insurance units of the sort commercialized in our project are indeed easier to grasp, as their payout structure relates directly to simple relations between weather and crop yields that farmers already have beliefs formed over. In a way, traditional index products with more complex payout structures may require a leap of faith from farmers with low education levels (in trusting that they will match their losses). Furthermore, this can be harmful for the sustainability of the insurance scheme. Most index insurance projects abound in anecdotal evidence about farmers complaining about lack of payouts after experiencing losses, and this is generally related to a lack of understanding about the insurance product they purchased. This type of negative feedback can greatly damage repeat purchases. The binary payout structure of a single index insurance unit of the type analyzed in this chapter is arguably more transparent, allowing farmers to better understand them and avoiding confusion when a payout does not trigger.

Implementing such an insurance system nevertheless requires a shift in the mindset of insurance providers, generally hang up in old business practices of selling one single insurance policy. Moreover, regulatory constraints may exist in some cases too. Despite the challenges and the need to accompany it with appropriate marketing and education campaigns, we show this system to be an overall feasible insurance scheme that should be seriously considered in contexts where heterogeneity in farmer characteristics and farming practices is present.

CHAPTER 5

General Conclusions

There is no doubt that Earth's climate is changing. With it, both the frequency and intensity of extreme weather events are likely to increase in the near future, with dramatic impacts on agricultural systems and livelihoods. While large-scale agricultural producers are better suited to evolve in this shifting landscape, smallholder farmers are both the most affected and the least able to cope with shocks and adapt. In this context, affordable, small-scale crop insurance is bound to be a fundamental piece in the policy toolkit to promote rural resilience and help bring about sustainable transformation to billions of people.

Weather index insurance in particular seems better equipped than its indemnity cousin to play a central role in this process. The cost reduction that comes along from basing insurance on an index are crucial to encourage low-income farmers to affordably insure themselves against weather risks. Now, after two decades of sustained push yet oscillating success for weather index insurance, it is high time to take stock of its triumphs, confront its weaknesses, and assess the way forward.

This dissertation provides some important building blocks to support such a task, discussing topics relating to both the demand and the supply perspectives.

In terms of the demand perspective, a number of insights are provided on smallholder farmers' real-world choices and attitudes towards insurance instruments. These insights should prove helpful when introducing new insurance products, by informing marketing and outreach efforts and providing a roadmap for policymakers to assess the degree of support needed in premium subsidies.

In terms of the supply perspective, the dissertation tackles the very relevant issue of basis risk, contributing a novel methodology to aid with the ex-ante assessment of spatial basis risk and presenting evidence on the feasibility of a flexible system capable of reducing the extent of design basis risk in index insurance products. Both of these should serve to devise better insurance products, with improved risk coverage and a higher appeal for targeted farmers.

5.1 Key findings

Chapter 2 analyzes the case of an insurance product against deficit or excess rainfall implemented in Madhya Pradesh, India. It presents causal evidence on three factors affecting take-up: price, distance to the reference weather station (a proxy for basis risk), and insurance literacy. The evidence is in line with the predictions from an expected-utility theoretical framework: demand is decreasing in price and basis risk, and increasing in product comprehension. Moreover, the chapter assesses the validity of a theoretical prediction about demand increasing with risk aversion at low levels of aversion, yet decreasing at higher levels (due to the uncertain nature of index insurance payouts). The chapter finds some evidence in this regard, although the estimation arguably suffers from lack of power to conclusively assert this. Finally, the chapter provides evidence on the demand for index insurance over time. In particular, the analyses indicate that receiving payouts is more important for repeat purchases than solely having purchased insurance in the past, indicating that trust or other behavioral mechanisms may be at play in farmers risk protection decisions.

Chapter 3, in turn, develops a novel methodology to estimate the extent of spatial basis risk for an arbitrary rainfall index insurance product. It relies on a stochastic rainfall generator model, standard in the hydrological literature, and extends it to accommodate non-traditional patterns of dependence between rainfall distributions at two nearby sites through a bivariate copula. This is important so as to capture the general tendency of extreme precipitation amounts jointly occurring at nearby sites more often than non-extreme amounts. The methodology is then applied to an excess rainfall index insurance product targeting horticultural farmers in Uruguay. The chapter presents evidence of substantial spatial basis risk, though at levels well below the upper bound from a model of demand for index insurance, implying that the product does indeed provide valuable insurance benefits. In addition, spatial basis risk increases concavely with distance, but the rate of increase is generally lower than what farmers perceive it to be. Lastly, the chapter shows that subtle geographic variations can play a major role for spatial basis risk, underscoring the importance of considering other geographic features, beyond distance to farmers' plots, when deciding the placing of reference weather stations (on which to base insurance payouts).

Finally, Chapter 4 describes a flexible commercial index insurance scheme against excess rainfall risk around harvest implemented in Canelones, Uruguay for horticultural crops. The scheme is unique in that it comprises a number of independent insurance units covering against different aspects of excess rainfall, based on the weather securities approach originally proposed by Hill and Robles (2011). This scheme is proposed as an alternative to traditional index insurance products, generally based in the expert design of an optimal insurance policy for a representative farmer under ideal management conditions.

The chapter exploits substantial variation in the insurance portfolios purchased by farmers to provide unique theory-based evidence on how farmers' heterogeneity in risk profiles affects the nature of their demand for insurance. In particular, it provides evidence that differences in crop composition, planting dates, soil drainage, distance to the weather station, and understanding of the product are all significant determinants of the type of insurance portfolio purchased by a farmer. Moreover, while a natural preconception is that farmers would not be able to construct their optimal portfolios, this notion is shown to be unfounded. Farmers do indeed purchase a wide array of different portfolios and, more importantly, do so in a way that is consistent with the underlying risk exposure of their particular crop composition and other individual characteristics related to their individual farming conditions. Finally, the chapter shows that, under the assumptions and optimal estimated parameters of the model, providing flexibility through an insurance units' scheme substantially improves farmer welfare in relation to a range of other hypothetical composite products.

5.2 Policy implications

One of the findings in Chapter 2 is that demand for insurance increases as product comprehension increases. This is an important finding given that weather index insurance products are generally offered to farmers who have limited experience with formal financial products. However, the chapter also finds that, while both price and investments in new weather stations (as a means to reduce the extent of spatial basis risk) are fairly effective in encouraging future demand, insurance literary training seems to be of a more transient nature, with no significant impact on understanding or demand after the first year of its implementation. Price discounts in the first year, however, had a much stronger effect on understanding, consistent with a model of learning by doing or learning by using. This is an important implication to consider into the design of future products.

Chapter 2 also finds that a prior positive experience with the insurance product—as captured by having purchased insurance and having received a payout during the first season—significantly encourages uptake in subsequent seasons. This could also be explained by low levels of trust in the product or the insurance company. Designing insurance products with small yet highly likely payouts during the first years of implementation may be desirable if it contributes to boost demand and trust and thus to establish the product among small farmers. This is an interesting avenue for future research.

An important contribution of this chapter relates to its analysis of the cost-effectiveness of alternative ways to boost demand for insurance. In this regard, providing premium subsidies seems to have been the most cost-effective measure—in terms of its direct effect on observed demand—, at least in the context of our study. Reducing the extent of spatial basis risk by installing automated weather stations comes second, even though its benefits on uptake are considered during only one insurance season. Considering effects over multiple seasons, together with the current downward tendency of automatic weather station prices make this an important intervention to evaluate in future projects. Finally, providing insurance training was the most costly measure when taking its benefits into consideration. In fact, the chapter presents evidence suggesting that simply experiencing the product may be a more effective way to increase insurance knowledge, further supporting premium subsidies as the mechanism of choice.

The following chapter, Chapter 3, has two main policy implications for existing and future index insurance schemes. First, the evidence points to the existence of important information asymmetries in regards to weather perceptions, where farmers perceive spatial correlation in extreme rainfall to be much lower than what historical records indicate. This suggests the need to complement the introduction of new index insurance products with extensive training on the spatial properties of rain, and that of extreme rainfall events in particular. Of course, low levels of education among farmers in developing countries represent a barrier to these type of technical trainings. The point is still crucial, though, since farmers are typically presented with a complex insurance product framed in terms of millimeters of rain and expected to make rational insurance decisions based on their prior knowledge and expectations on complex meteorological phenomena (see also Sibiko, Veettil, and Qaim, 2016).

Second, in order to minimize the extent of spatial basis risk a careful consideration of subtle regional variations in rainfall patterns is essential, in addition to the more commonly studied distance aspect. This should be ideally carried out at the initial design stages of an index insurance scheme, when selecting from existing or determining the placement of new weather stations to be used as a reference for the product. In particular, subtle differences in rainfall patterns between the plot location and its reference rainfall gauge may result in substantial disparities in basis risk between locations. In extreme cases, local geo-climatic features could prove more important for determining basis risk than the actual distance to the reference gauge. Surprisingly, this point has been largely ignored by the index insurance literature until now, beyond settings where this is directly evident such as mountainous topographies.

Despite its advantages, however, the methodology in Chapter 2 can only tackle the estimation of the spatial component of basis risk. As discussed in the chapter, there exist a number of different sources for the mismatch between a farmer’s losses and an index insurance product’s payout. An analysis incorporating these other sources would require longitudinal farm-level data which is not always available, but this remains a very interesting avenue for future research. Such an analysis can contribute to the design of innovative index insurance products that help mitigate the negative aspects of basis risk and reveal the potential for novel financial instruments to enhance farmers’ resilience. This is important as, for instance, if most basis risk were to arise from spatial variability in weather, the commonly proposed solution of increasing weather station density would be appropriate. Alternatively, however, if a large fraction of basis risk were to be explained by idiosyncratic differences in farmers' abilities and technologies to cope with weather risk, policy recommendations would be entirely different.

The final chapter, Chapter 4, illustrates the feasibility of rolling out a real-world system of commercially-backed, flexible insurance units or weather securities. At least in the particular context of horticultural farmers in Uruguay, this system seems to be more appropriate than traditional index insurance systems based on one single, standardized product. While the generalizability of these findings seems plausible for other contexts with sufficient degree of farmer heterogeneity, this remains an important avenue for future research.

The evidence presented in the chapter has significant policy implications. Farmer heterogeneity matters. The approach of selling index insurance as a single, one-size-fits-all policy is misguided. For all its worth, index insurance cannot be designed to fit all risk profiles. Product flexibility is important and farmers seem to have the ability to adapt it to their needs. However, not all heterogeneity is equal. On one hand, there are certain sources of heterogeneity that are related to information problems, such as limited product understanding or misperceptions about rainfall distributions. In this respect, a flexible insurance system would serve only as a temporary patch; requiring more permanent solutions that tackle the root of the problem such as accompanying the rollout of new insurance with appropriate marketing and campaigns educating about true climatic patterns, as discussed above. On the other hand, other sources of heterogeneity are of a more structural nature, such as the degree of soil drainage a farmer has or his specific choice of crops and planting dates. A flexible insurance system is almost unavoidable in these cases, as it is hard to imagine how to design only a few composite insurance products that could appropriately consider the weather risks of all potential risk profiles.

The product described in the chapter was marketed in a context of better-educated farmers relative to other rural contexts in Latin America, not to mention Africa or Asia. This gives rise to the question of whether farmers' choice of appropriate insurance portfolios would still be attainable in these other contexts. Preliminary evidence in this regard discussed in Hill and Robles (2011) shows that smallholder farmers in Ethiopia do indeed flexibly respond to their risk exposure by purchasing different insurance portfolios. In addition, insurance units of the sort commercialized in this project may indeed be easier to grasp, as their payout structure relates directly to simple relations between weather and crop yields that farmers already have beliefs formed over. In a way, traditional index products with more complex payout structures may require a leap of faith from farmers with low education levels (in trusting that they will match their losses). Furthermore, this can be harmful for the sustainability of the insurance scheme. Most index insurance projects abound in anecdotal evidence about farmers complaining about lack of payouts after experiencing losses, and this is generally related to a lack of understanding about the insurance product they purchased. This type of negative feedback can greatly damage repeat purchases. The binary payout structure of a single index insurance unit of the type analyzed in this chapter is arguably more transparent, allowing farmers to better understand them and avoiding confusion when a payout does not trigger.

Implementing such a flexible insurance system nevertheless requires a shift in the mindset of insurance providers, generally hang up in old business practices of selling one single insurance policy. Moreover, regulatory constraints may exist in some cases too. Despite the challenges and the need to accompany it with appropriate marketing and education campaigns, the chapter shows this system to be an overall feasible insurance scheme that should be seriously considered in contexts where heterogeneity in farmer characteristics and farming practices is present.

5.3 Final comments

A number of weather index insurance experiences around the world have shown this type of insurance's potential as a formal, efficient risk management tool for farmers in developing countries. However, to truly bring it at scale globally its limitations have to be addressed. This dissertation provides a number of wortwhile approaches to achieve this. Below, we describe some additional new developments in the index insurance realm that may potentially help this technology to make the jump.

An interesting new alternative is Picture-Based Crop Insurance (PBI), which is currently being tested in the Indian states of Haryana and Punjab. Under PBI, farmers take pictures of their insured plots every week using their own smartphones and a specially-designed app that aids in keeping the frame of view fixed in the same portion of the field. With this time-series of pictures, a farmer can then make a claim for any losses experienced, which can be assessed by agronomic experts or automated machine learning algorithms on the basis of the pictures and other auxiliary information. This type of product can greatly reduce basis risk and encourage uptake by instilling a sense of ownership in the farmer. Initial results are very promising, in terms of both the feasibility of the approach (Kramer et al. 2017a) and its sustainability, with no evidence of moral hazard or adverse selection (as would be expected from the product's resemblance to indemnity-based insurance), nor of picture tampering or fraud (Kramer et al. 2017b).

Importantly, the increasing affordability of automatic weather stations and the expanding technologies for remote sensing of weather variables and crop growth (such as micro-satellites and unmanned aerial vehicles) have an enormous potential to underlie innovative insurance products with reduced basis risk in the near future. Satellite-based products are already available across many continents, and are being received with a lot of enthusiasm among local stakeholders (some prominent examples are the Index-Based Livestock Insurance—IBLI—in Kenya and Ethiopia; IBLI in Mongolia; and RIICE in East and South-East Asia).

Finally, weather index insurance may find a natural partnership in Climate-Smart Agriculture (CSA) technologies, which have gained popularity during the past decade as an efficient means towards climate adaptation by rural farming communities.

In some cases, CSA technologies involve reducing a crop's vulnerability to certain weather risks, thus achieving a similar objective as crop insurance. Due to these similarities between the two families of technologies, there has been a recent strand of work that has focused on evaluating the potential for complementarities between them.

One of the most important examples of complementarity between weather-index insurance and a CSA technology are drought tolerant (DT) seed varieties. Although the main characteristic of such seed varieties is their resistance to mild or moderate lack of soil moisture, crop failure is generally an inevitable result under an extreme drought, with the added consequence of farmers' being worse off due to having to repay the higher cost of DT seeds. Weather index insurance, on the other hand, is not very well suited to handle moderate drought because it tends to be expensive under a high frequency of loss (insurance premiums must be high to account for frequent payouts). Nevertheless, because extreme drought events occur much more rarely and are generally easier to identify through an index (compared with more moderate events that may or may not damage crops), weather index insurance boasts natural comparative advantages to handle this layer of risk. It is natural to see, thus, that a holistic system—wherein farmers rely first on DT seeds to inexpensively cover more frequent and milder drought risks, and in addition rely on reduced-cost catastrophic index insurance against extreme events—could provide farmers with more complete protection against all potential scenarios, thus more efficiently handling drought risk at a much lower cost than any of the above technologies would be able to achieve on its own (Lybbert and Carter 2015; Ward et al. 2015).

Lastly, other ways in which index insurance can partner up with CSA technologies is by encouraging CSA adoption. Many farmers generally refrain from adopting CSA practices due to their uncertainty and higher perceived risks, relative to traditional practices. In these contexts, index insurance may be a tool to inexpensively give the farmer the necessary peace of mind to try out the new technology. Such an approach could either complement or substitute for standard subsidies for encouraging CSA adoption. Understanding the optimal interplay between these two mechanisms is an important research avenue.

Overall, and despite some setbacks, weather index insurance seems to have earned a well-deserved place in the development agenda. In the face of climate change and its effects on rural communities in developing countries, the potential benefits of insuring against risk are incalculable. This dissertation discusses some novel perspectives and provides important considerations for the design and marketing of future index insurance products. It is up to the large number of practitioners and researchers in the field, together with policymakers in developing regions, to find the optimal ways to scale up these important instruments for coping with risk.

Appendix Chapter 2

Demand for Insurance

Demand for insurance is zero $(\alpha = 0)$ when basis risk is extremely high ($r \geq p(1-p)$) for any multiple equal or higher than unity $(m \geq 1)$.

We show that the expected utility with no insurance $EV(Y(S)_{\alpha=0})$ is equal or higher than the utility of having any positive insurance $EV(Y(S)_{\alpha>0})$:

$$EV(Y(S)_{\alpha=0}) \geq EV(Y(S)_{\alpha>0})$$

$$\sum_S P(S)\ \left(V(Y(S)_{\alpha=0}) - V(Y(S)_{\alpha>0})\right) \geq 0.$$

We work on the left-hand side:

$$\sum_S P(S)\ \frac{V(Y(S)_{\alpha=0}) - V(Y(S)_{\alpha>0})}{(Y(S)_{\alpha=0}) - (Y(S)_{\alpha>0})}\left((Y(S)_{\alpha=0}) - (Y(S)_{\alpha>0})\right).$$

Let's define $T(S) \equiv \frac{V(Y(S)_{\alpha=0}) - V(Y(S)_{\alpha>0})}{(Y(S)_{\alpha=0}) - (Y(S)_{\alpha>0})}$, where

$$T(L,0) \equiv \frac{V(W-L) - V(W-L-\alpha pmL)}{\alpha pmL}$$

$$T(L,L) \equiv \frac{V(W-L) - V(W-L-\alpha pmL+\alpha L)}{\alpha pmL - \alpha L}$$

$$T(0,0) \equiv \frac{V(W) - V(W-\alpha pmL)}{\alpha pmL}$$

$$T(0,L) \equiv \frac{V(W) - V(W-\alpha pmL+\alpha L)}{\alpha pmL - \alpha L}.$$

Because of the concavity of $V(.)$, $T(L,0) > T(L,L) > T(0,0) > T(0,L)$. After replacing terms we have

$$= r\,T(L,0)\,\alpha pmL + (p-r)\,T(L,L)\,(\alpha pmL - \alpha L) + (1-p-r)\,T(0,0)\,(\alpha pmL) + r\,T(0,L)\,(\alpha pmL - \alpha L)\,.$$

Now we use $T(L,0) > T(L,L) > T(0,0) > T(0,L)$ and replace terms

$$= \alpha pmL(r\,T(L,0) + (p-r)\,T(L,L) + (1-p-r)\,T(0,0) + r\,T(0,L)\,)$$

$$- \alpha L\big((p-r)\,T(L,L) + r\,T(0,L)\,\big)$$

$$> \alpha pmL(r\,T(L,L) + (p-r)\,T(L,L) + (1-p-r)\,T(0,L) + r\,T(0,L)\,)$$

$$- \alpha L\big((p-r)\,T(L,L) + r\,T(0,L)\,\big)$$

$$= \alpha L\,T(L,L)\big(p^2m - (p-r)\big) + \alpha L\,T(0,L)(pm(1-p) - r)$$

$$= \alpha L\,T(L,L)\big(r - (p - mp^2)\big) - \alpha L\,T(0,L)\big(r - m(p-p^2)\big)$$

For this expression to be non-negative, two conditions are sufficient:

$$r - (p - mp^2) \geq 0 \qquad => \qquad r \geq p(1-mp)$$

$$\big(r - (p - mp^2)\big) \geq \big(r - m(p-p^2)\big) \quad => \qquad m \geq 1.$$

Combining these conditions we conclude that when $r \geq p(1-p)$ the demand for insurance is zero for any $m \geq 1$.

Appendix Table 1 — Product Design

Coverage period	Time period	Description	Index
1	Jun 25 – July 20	This period corresponds to the sowing and germination stage. Sowing usually takes place after June 15. Farmers have the option to wait until the start of the rainy season to decide when to start sowing. After sowing and during the germination phase, the major peril is excessive rain on a single day.	Maximum rainfall on any single day during coverage period
2	Jul 21 – Sep 15	This period combines the vegetative and reproductive phases. Both phases share similar perils—either excess or deficit of total rainfall during the period—although during the vegetative phase, rain deficit seems relatively more important, and during the reproductive phase, excess rain. Given that each phase by itself is relatively short, our evaluation is that it is not practical to create securities for each phase separately.	Total cumulative rainfall during coverage period
3	Sept 16 – Oct 15	This period combines the maturity phase and harvest. The major peril is excess rainfall, especially heavy rain on a single day.	Maximum rainfall on any single day during coverage period

Appendix Table 2 — First-Stage Results for Table 6

	Specification (3)	Specification (5)	
	(1)	(2)	(3)
	Log (distance)	Log (distance)	Log (distance) x Log (price)
New weather station dummy	–0.955***	–0.351	2.464
	(0.235)	(0.615)	(2.227)
New weather station dummy x price		–0.117	–1.430***
		(0.104)	(0.451)
Observations	2,183	2,183	2,183
Unrestricted R^2	0.255	0.255	0.271
Restricted R^2	0.052	0.052	0.068
Wald F-test of joint significance	16.47***	9.54***	10.00***

Standard errors, clustered at the village level, are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Source: Administrative sales data.

Appendix Chapter 3

Appendix Table 3 — Weather Station Data Availability and Summary Statistics

Rainfall gauge	Proportion of days with rainfall (in %)	Average daily rainfall amount (in mm.)	Start date	End date	Number of Obs.
Pre-existing gauges					
Aeropuerto Carrasco	26.9	10.9	1981-01-01	2011-12-31	2,797
Aeropuerto Melilla	27.7	11.2	1981-01-01	2011-12-31	2,797
Cerrillos	16.1	16.1	1981-01-01	2011-12-31	2,797
Chamizo	17.4	18.9	1981-01-01	2011-12-31	2,797
Dr. Soca	14.3	17.8	1981-01-01	2011-12-31	2,797
Libertad	18.4	17.3	1981-01-01	2011-12-31	2,797
Mendoza	18.6	17.7	1981-01-01	2011-12-31	2,797
Prado	26.4	10.9	1981-01-01	2011-12-31	2,797
San Jacinto	15.1	18.2	1981-01-01	2011-12-31	2,797
Tala	14.5	19.4	1981-01-01	2011-12-31	2,797
Villa Rodriguez	18.4	17.2	1981-01-01	2011-12-31	2,797
Insurance reference gauges					
Chacra Policial	28.9	11.7	2013-11-30	2015-06-23	180
Progreso	31.7	13.8	2013-11-30	2015-06-23	180
San Bautista	31.1	11.2	2013-11-30	2015-06-23	180
San Jacinto	32.2	12.1	2013-11-30	2015-06-23	180
Tala	31.7	10.4	2013-11-30	2015-06-23	180
Monitoring gauges					
WS 1	37.0	14.1	2013-10-27	2015-02-10	162
WS 2	41.2	13.3	2013-11-10	2014-12-29	119
WS 3	39.5	14.3	2013-10-28	2015-02-10	162
WS 6	31.5	11.2	2013-10-28	2015-02-10	162
WS 7	33.8	12.4	2013-12-02	2015-02-12	160
WS 8	36.0	13.8	2013-12-02	2015-02-10	161
WS 9	37.6	12.6	2013-11-10	2015-02-05	157

WS 10	33.3	14.5	2013-11-29	2015-02-10	162
WS 11	37.7	14.5	2013-10-27	2015-02-10	162
WS 12	34.6	14.1	2013-11-03	2015-02-10	162
WS 13	35.2	15.8	2013-11-22	2015-02-10	162
WS 15	38.3	10.5	2013-11-09	2015-02-10	162
WS 16	33.3	13.4	2013-11-09	2015-02-10	162
WS 17	34.6	12.6	2013-11-09	2015-02-10	162
WS 19	38.7	10.5	2013-11-28	2015-02-11	163
WS 20	39.1	10.4	2013-12-08	2015-02-11	156
WS 21	38.0	10.4	2013-11-28	2015-02-11	163
WS 22	35.2	11.3	2013-11-09	2015-02-10	162
WS 23	35.0	10.1	2013-11-29	2015-02-11	163
WS 24	33.1	13.8	2013-11-29	2015-02-11	163
WS 26	34.9	15.4	2013-11-22	2015-02-07	129
WS 27	37.5	12.0	2013-11-22	2015-02-10	152
WS 28	37.1	10.2	2013-11-04	2015-02-18	170
WS 29	34.8	12.2	2013-11-04	2015-02-12	164
WS 30	35.4	10.5	2013-11-04	2015-02-12	164
WS 32	36.8	6.8	2013-11-28	2015-02-11	163
WS 33	37.4	9.8	2013-11-16	2015-02-11	163
WS 34	39.4	10.2	2013-11-16	2015-02-13	165
WS 35	35.5	10.0	2013-11-16	2015-02-13	155
WS 36	35.7	10.4	2013-11-16	2015-02-13	157
WS 37	35.2	11.4	2013-11-28	2015-02-13	165
WS 38	40.0	13.7	2013-11-16	2014-03-04	90
WS 39	36.4	11.6	2013-11-04	2015-02-13	165
WS 40	34.6	9.2	2013-12-07	2015-02-13	130

References

Ailliot, P., D. Allard, V. Monbet, and P. Naveau, 2015. “Stochastic Weather Generators: An Overview of Weather Type Models,” *Journal de la Société Française de Statistique* 156(1): pp. 101-113.

Baffault, C., M.A. Nearing, and A.D. Nicks, 1996. “Impact of CLIGEN Parameters on WEPP Predicted Average Annual Soil Loss,” *Transactions of the ASAE* 39(2): pp. 447–457.

Balsa, A.I., N. Gandelman, N. González, 2015. “Peer Effects in Risk Aversion,” *Risk Analysis* 35(1): pp. 27-43.

Bárdossy, A., 2006. “Copula-Based Geostatistical Models for Groundwater Quality Parameters,” *Water Resources Research* 42(11): pp. 1-12.

Bárdossy, A., and E.J. Plate, 1992. “Space-Time Model for Daily Rainfall using Atmospheric Circulation Patterns,” *Water Resources Research* 28: pp. 1247–1259.

Bárdossy, A., and G.G.S. Pegram, 2009. “Copula Based Multisite Model for Daily Precipitation Simulation,” *Hydrology and Earth System Sciences* 13(12): pp. 2299-2314.

Barrett, C.B., B.J. Barnett, M.R. Carter, S. Chantarat, J.W. Hansen, A.G. Mude, D. Osgood, J.R. Skees, C.G. Turvey, M.N. Ward, 2007. “Poverty Traps and Climate Risk: Limitations and Opportunities of Index-Based Risk Financing,” IRI Technical Report No. 07-02.

Barrett, C. and J. McPeak. 2006. “Poverty Traps and Safety Nets.” In *Poverty, Inequality and Development: Essays in Honor of Erik Thorbecke*, edited by A. de Janvry and R. Kanbur, 131–154. New York: Springer.

Bjørnstad, O.N., and W. Falck, 2001. “Nonparametric Spatial Covariance Functions: Estimation and Testing,” *Environmental and Ecological Statistics* 8(1): pp. 53-70.

Buishand, T.A., 1978. “Some Remarks on the Use of Daily Rainfall Models,” *Journal of Hydrology* 36: pp. 295–308.

Cai, J., 2016. “The Impact of Insurance Provision on Household Production and Financial Decisions,” *American Economic Journal: Economic Policy* 8(2): pp. 44-88.

Cai, J., A. de Janvry, and E. Sadoulet. 2013. “Social Networks and the Decision to Insure: Evidence from Randomized Experiments in China.” Mimeo. University of Michigan.

Cao, M., A. Li, and J. Wei, 2004. “Precipitation Modeling and Contract Valuation: A Frontier in Weather Derivatives,” *The Journal of Alternative Investments* 7(2): pp. 93-99.

Capéraà P., A.L. Fougères, and C. Genest, 1997. “A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas,” *Biometrika* 84(3): pp. 567–577

Ciach, G.J., and W.F. Krajewski, 2006. “Analysis and Modeling of Spatial Correlation Structure in Small-Scale Rainfall in Central Oklahoma,” *Advances in Water Resources* 29: pp. 1450–1463.

Clarke, D.J., 2016. “A Theory of Rational Demand for Index Insurance,” *American Economic Journal: Microeconomics* 8(1): pp. 283-306.

Clarke, D.J., and S. Dercon, 2016. Dull Disasters? How planning ahead will make a difference. Oxford University Press.

Clarke, D., and M. Kalani. 2011. Microinsurance Decisions: Evidence from Ethiopia. Mimeo, Oxford University.

Clarke, C., O. Mahul, K.N. Rao, and N. Verma. 2012. *Weather-Based Crop Insurance in India*. Policy Research Working Paper 5985. Washington, DC: World Bank.

Cole, S., G. Bastian, S. Vyas, C. Wendel, and D. Stein, 2012. “The Effectiveness of Index-Based Micro-Insurance in Helping Smallholders Manage Weather-Related Risks,” London: Social Science Research Unit, University of London.

Cole, S., X. Giné, J. Tobacman, P.B. Topalova, R.M. Townsend, and J.I. Vickery, 2013. “Barriers to Household Risk Management: Evidence from India,” *American Economic Journal: Applied Economics* 5(1): pp. 104–135.

Cole, S., X. Giné, and J. Vickery, 2013. “How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment,” Harvard Business School Working Paper 13-080.

Cole, S., D. Stein, and J. Tobacman. 2014. “Dynamics of Demand for Index Insurance: Evidence from a Long-Run Field Experiment.” *American Economic Review Papers & Proceedings*, forthcoming.

Danaher, P.J., and M.S. Smith, 2011. “Modeling Multivariate Distributions Using Copulas: Applications in Marketing,” *Marketing Science* 30(1): pp. 4-21.

Dercon, S., R.V. Hill, D. Clarke, I. Outes-Leon, and A.S. Taffesse, 2014. "Offering Rainfall Insurance to Informal Insurance Groups: Evidence from a Field Experiment in Ethiopia," *Journal of Development Economics* 106: pp. 132-143.

Dercon, S., and J. Hoddinott, 2004. "Health, Shocks, and Poverty Persistence," in *Insurance against Poverty*, edited by S. Dercon, 123–136. Oxford, UK: Oxford University Press

Dubé, J.P., 2004. "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science* 23(1): pp. 66-81.

Dubrovský, M., J. Buchtele, and Z. Žalud, 2004. "High-Frequency and Low-Frequency Variability in Stochastic Daily Weather Generator and Its Effect on Agricultural and Hydrologic Modelling," *Climatic Change* 63(1-2), pp. 145-179.

Elabed, G., M.F. Bellemare, M.R. Carter, and C. Guirkinger, 2013. "Managing Basis Risk with Multiscale Index Insurance," *Agricultural Economics* 44(4-5): pp. 419-431.

Embrechts, P., A. McNeil, and D. Straumann, 2002. "Correlation and Dependence in Risk Management: Properties and Pitfalls," in M.A.H. Dempster (ed.), Risk Management. Cambridge University Press, Cambridge.

Genest, C., and A.C. Favre, 2007. "Everything You Alwayss Wanted to Know about Copula Modeling but Were Afraid to Ask," *Journal of Hydrologic Engineering* 12(4): pp. 347-368.

Genest, C., B. Rémillard, and D. Beaudoin, 2009. "Goodness-of-Fit Tests for Copulas: A Review and a Power Study," *Insurance: Mathematics and Economics* 44: pp. 199–213.

Giné, X., R.M. Townsend, and J.I. Vickery, 2008. "Patterns of Rainfall Insurance Participation in Rural India," *World Bank Economic Review* 22(3): pp. 539–566.

Goodwin, B.K., and A. Hungerford, 2015. "Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts," *American Journal of Agricultural Economics* 97 (3): pp. 879-896.

Greatrex, H.L., 2012. "Satellite Rainfall Information in Africa," Statistical Services Centre, University of Reading, UK, Edition 1.

Gumbel, E. J., 1960. "Distributions des Valeurs Extremes en Plusieurs Dimensions," Publications de l'Institute de Statistique de l'Université de Paris 9: pp. 171–173.

Hansen, L.P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50(4): pp. 1029-1054.

Hazell, P., J. Anderson, N. Balzer, A. Hastrup Clemmensen, U. Hess, and F. Rispoli, 2010. "The Potential for Scale and Sustainability in Weather Index Insurance for Agriculture and Rural Livelihoods," Rome: International Fund for Agricultural Development.

Hazell, P., C. Pomareda, and A. Valdes, 1986. Crop Insurance for Agricultural Development. Baltimore: Johns Hopkins University Press.

Hendel, I., 1999. “Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns,” *The Review of Economic Studies* 66(2): pp. 423-446.

Hill, R., J. Hoddinott, and N. Kumar. 2013. “Adoption of Weather-Index Insurance: Learning from Willingness to Pay among a Panel of Households in Rural Ethiopia.” *Agricultural Economics* 44(4-5): pp. 385-398.

Hill, R., and M. Robles, 2011. “Flexible Insurance for Heterogeneous Farmers: Results from a Small-Scale Pilot in Ethiopia,” IFPRI Discussion Paper 01902.

Holt, C.A., and S.K. Laury, 2002. “Risk Aversion and Incentive Effects,” *American Economic Review* 92(5): pp. 1644-1655.

IPCC, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White, eds. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ison N.T., A.M. Feyerherm, L. Dean Bark, 1971. “Wet Period Precipitation and the Gamma Distribution,” *Journal of Applied Meteorology* 10: pp. 658–665.

Jensen, N., A. Mude, and C. Barrett, 2014. “How Basis Risk and Spatiotemporal Adverse Selection Influence Demand for Index Insurance: Evidence from Northern Kenya,” MPRA Paper 60452, University Library of Munich, Germany.

Joe, H., 1997. Multivariate Models and Multivariate Dependence Concepts. Chapman & Hall, New York.

Junker, M., and A. May, 2005. “Measurement of Aggregate Risk with Copulas,” *Econometrics Journal* 8(3): pp. 428-454.

Just, R.E., and Q. Weninger, 1999. “Are Crop Yields Normally Distributed?,” *American Journal of Agricultural Economics* 81(2): pp. 287-304.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry, 2012. “Agricultural Decisions after Relaxing Credit and Risk Constraints.” NBER Working Papers 18463.

Katz, R.W., 1977. “Precipitation as a Chain-Dependent Process,” *Journal of Applied Meteorology* 16: pp. 671–676.

Kellner, R., and N. Gatzert, 2013. “Estimating the Basis Risk of Index-Linked Hedging Strategies using Multivariate Extreme Value Theory,” *Journal of Banking & Finance* 37: pp. 4353-4367.

Kharoubi-Rakotomalala, C., and F. Maurer, 2013. "Copulas In Finance Ten Years Later," *Journal of Applied Business Research* 29(5): pp. 1555-1566.

Kim, J., G.M. Allenby, and P.E. Rossi, 2002. "Modeling Consumer Demand for Variety," *Marketing Science* 21(3): pp. 229-250.

Kleiber, W., R.W. Katz, and B. Rajagopalan, 2012. "Daily Spatiotemporal Precipitation Simulation Using Latent and Transformed Gaussian Processes," *Water Resources Research* 48(1).

Kost, A., P. Läderach, F. Myles, S.E. Cook, and L. Gómez, 2012. "Improving Index-Based Drought Insurance in Varying Topography: Evaluating Basis Risk Based on Perceptions of Nicaraguan Hillside Farmers," PLoS One 7(12): e51412.

Kousky, C., and R.M. Cooke, 2009. "The Unholy Trinity: Fat Tails, Tail Dependence, and Micro-Correlations," Resources For the Future Discussion Paper 09-36-REV.

Kramer, B., F. Ceballos, K. Hufkens, E. Melaas, A. Mishra, M.S. Toor, and M. Robles, 2017a. "Picture-Based Crop Insurance: Using Farmers' Smartphone Pictures to Minimize the Costs of Loss Verification," IFPRI Project Note. Washington, DC: International Food Policy Research Institute.

Kramer, B., F. Ceballos, M. Krupoff, A. Mishra, M. Robles, and M.S. Toor, 2017b. "Picture-Based Insurance: Is It Sustainable? Effects on Willingness to Pay, Adverse Selection and Moral Hazard," IFPRI Project Note. Washington, DC: International Food Policy Research Institute.

Lettenmaier, D., 1995. "Stochastic Modeling of Precipitation with Applications to Climate Model Downscaling," in von Storch, H., Navarra, A., Eds., Analysis of Climate Variability, Applications of Statistical Techniques, Springer-Verlag, Berlin: pp. 197–212.

Lopez Cabrera, B., M. Odening, and M. Ritter, 2013. "Pricing Rainfall Futures at the CME," *Journal of Banking & Finance* 37(11): pp. 4286-4298.

Low, H., and C. Meghir, 2017. "The Use of Structural Models in Econometrics," *Journal of Economic Perspectives* 31(2): pp. 33-58.

Lybbert, T., and M.R. Carter. 2015. "Bundling Drought Tolerance and Index Insurance to Reduce Rural Household Vulnerability to Drought." In Sustainable Economic Development: Resources, Environment, and Institutions, edited by A. Balisacan, U. Chakravorty, and M.-L. Ravago, 401–414. San Diego, CA, US: Elsevier Academic Press.

Matul, M., A. Dalal, O. De Bock, and W. Gelade. 2013. *Why People Do Not Buy Microinsurance and What We Can Do about It.* Briefing Note 17. Geneva: Microinsurance Innovation Facility.

Mehrotra, R., R. Srikanthan, and A. Sharma, 2006. “A Comparison of Three Stochastic Multi-Site Precipitation Occurrence Generators,” *Journal of Hydrology* 331: pp. 280–292.

Mobarak, A. M., and M. Rosenzweig. 2012. *Selling Formal Insurance to the Informally Insured*. Working Paper No. 97, Economic Growth Center Discussion Paper No. 1007. Economics Department. New Haven, CT, US: Yale University.

Mosthaf, T., A. Bárdossy, and S. Hörning, 2015. “Interpolating Non-Parametric Distributions of Hourly Rainfall Intensities Using Random Mixing,” presented at the EGU General Assembly 2015 Conference, mimeo.

Nelsen, R.B., 2006. An Introduction to Copulas. Springer, New York.

Nelson, G.C., H. Valin, R.D. Sands, P. Havlík, H. Ahammad, D. Deryng, J. Elliott, S. Fujimori, T. Hasegawa, E. Heyhoe, P. Kyle, M. Von Lampe, H. Lotze-Campen, D. Mason d’Croz, H. van Meijl, D. van der Mensbrugghe, C. Müller, A. Popp, R. Robertson, S. Robinson, E. Schmid, C. Schmitz, A. Tabeau, and D. Willenbockel. 2013. “Climate Change Effects on Agriculture: Economic Responses to Biophysical Shocks.” *Proceedings of the National Academy of Sciences.* doi:10.1073/pnas.1222465110.

Odening, M., O. Musshoff, and W. Xu, 2007. “Analysis of Rainfall Derivatives Using Daily Precipitation Models: Opportunities and Pitfalls,” *Agricultural Finance Review* 67(1): pp. 135-156.

Okhrin, O., M. Odening, and W. Xu, 2013. “Systemic Weather Risk and Crop Insurance: The Case of China,” *Journal of Risk and Insurance* 80: pp. 351–372.

Pakes, A., and D. Pollard, 1989. “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica* 57(5): pp. 1027-1057.

Patton, A.J., 2012. “A Review of Copula Models for Economic Time Series,” *Journal of Multivariate Analysis* 110: pp. 4-18.

Pegram, G.S., 2009. “A Nested Multisite Daily Rainfall Stochastic Generation Model,” *Journal of Hydrology* 371(1): pp. 142-153.

Racsko, P., L. Szeidl, and M. Semenov, 1991. “A Serial Approach to Local Stochastic Weather Models,” *Ecological Modeling* 57: pp. 27–41.

Ramirez, O.A., S. Misra, and J. Field, 2003. “Crop-Yield Distributions Revisited,” *American Journal of Agricultural Economics* 85(1): pp. 108-120.

Richardson, C.W., 1981. “Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation,” *Water Resources Research* 17: pp. 182–190.

Ritter, M., O. Musshoff, and M. Odening, 2014. “Minimizing Geographical Basis Risk of Weather Derivatives Using A Multi-Site Rainfall Model,” *Computational Economics* 44(1): pp. 67-86.

Serinaldi, F., 2008. "Analysis of Inter-Gauge Dependence by Kendall's τ_K, Upper Tail Dependence Coefficient, and 2-Copulas with Application to Rainfall Fields," *Stochastic Environmental Resources Risk Assessment* 22: pp. 671–688.

Serinaldi, F., 2009a. "Copula-Based Mixed Models for Bivariate Rainfall Data: An Empirical Study in Regression Perspective," *Stochastic Environmental Research and Risk Assessment* 23(5): pp. 677-693.

Serinaldi, F., 2009b. "A Multisite Daily Rainfall Generator Driven by Bivariate Copula-Based Mixed Distributions," *Journal of Geophysical Research: Atmospheres (1984–2012),* 114(D10).

Serinaldi, F. and C.G. Kilsby, 2014. "Simulating Daily Rainfall Fields over Large Areas for Collective Risk Estimation," *Journal of Hydrology* 512: pp. 285-302.

Sibiko, K.W., P. Veettil, and M. Qaim, 2016. "Small Farmers' Preferences for Weather Index Insurance: Insights from Kenya," AAAE Fifth International Conference, Addis Ababa, Ethiopia, African Association of Agricultural Economists (AAAE).

Smith, R., and R. Blundell. 1986. "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply." *Econometrica* 54:679–685.

Smith, V. H., and B.K. Goodwin. 2010. "Private and Public Roles in Providing Agricultural Insurance in the United States." In *Public Insurance and Private Markets*, edited by J. R. Brown, 173–210. Washington, DC: American Enterprise Institute.

Srikanthan, R., and G.S. Pegram, 2007. "Stochastic Generation of Spatially Consistent Daily Rainfall," MODSIM07 - Land, Water and Environmental Management: Integrated Systems for Sustainability, Proceedings.

Stein, D. 2011. "Paying Premiums with the Insurer's Money: How Loss Aversion Drives Dynamic Insurance Decisions." Mimeo.

Todorovic, P., and D.A. Woolhiser, 1975. "A Stochastic Model of n-day Precipitation," *Journal of Applied Meteorology* 14: pp. 17–24.

Townsend, R.M., 1994. "Risk and Insurance in Village India," *Econometrica* 62(3): pp. 539–591.

Trivedi, P.K., and D.M. Zimmer, 2007. "Copula Modeling: An Introduction for Practitioners," Foundations and Trends in Econometrics 1(1): pp. 1-111.

Wang, H.H., and H. Zhang, 2003. "On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach," *The Journal of Risk and Insurance* 70(1): pp. 111-124.

Ward, P.S., D.J. Spielman, D.L. Ortega, N. Kumar, and S. Minocha, 2015. "Demand for Complementary Financial and Technological Tools for Managing Drought Risk," IFPRI Discussion Paper 01430. Washington, DC.

Wilks, D.S., 1998. “Multisite Generalization of a Daily Stochastic Precipitation Generation Model,” *Journal of Hydrology* 210: pp. 178–191.

Woolhiser, D.A., 1992. “Modeling Daily Precipitation – Progress and Problems,” in Statistics in the Environmental & Earth Sciences, Walden A.T., Guttorp P. (Eds.), Halsted Press, London: pp. 71–89.

World Bank, 2013. *Turn Down the Heat: Climate Extremes, Regional Impacts, and the Case for Resilience*. Report for the World Bank by the Potsdam Institute for Climate Impact Research and Climate Analytics. Washington, DC.