Land, Labor and Technology: Essays in Development Economics

Land, Labor and Technology: Essays in Development Economics

A dissertation presented

by

Asanga Nilesh Fernando

to

The Department of Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachusetts

May 2015

Dissertation Advisors:
Professor Asim Khwaja
Professor Lawrence Katz

Author:
Asanga Nilesh Fernando

Land, Labor and Technology:
Essays in Development Economics

Abstract

Many of the world's rural poor make a living from agriculture. Consequently, the productivity of agriculture and non-agricultural employment opportunities are important determinants of rural poverty and the subject matter of the three essays in this dissertation. The first chapter in this dissertation estimates the long-term causal effect of inheriting land in rural India. Using quasi-experimental methods, I find that inheriting land greatly influences occupational trajectories and can suppress consumption to an extent that may overwhelm its direct benefit. The second chapter uses a field experiment to understand whether barriers to information influence agricultural productivity. We find that the introduction of a mobile phone-based agricultural information service greatly influences reported sources of information, input adoption decisions and agricultural productivity. The final chapter studies the effect of the external provision of agricultural information on social interactions and agricultural outcomes in village India. Using a field experiment, I find that the introduction of a mobile phone-based agricultural extension service influences the structure and content of social interactions with peers both within and outside the original study population. Respondents receiving valuable agricultural information are more likely to interact with their peers and share information from the service. These changes in social interactions also influence the agricultural outcomes of peers. These results suggest that technological innovations may increase the returns to in-person exchange of information and, in so doing, influence agricultural outcomes.

Contents

Abstract . . . iii
Acknowledgments . . . ix

Introduction **1**

1 Shackled to the Soil: The Long-Term Effects of Inherited Land on Labor Mobility and Consumption **4**

1.1 Introduction . . . 4
1.2 Context: Agricultural Land in Rural India . . . 9
1.2.1 Customs and Laws Governing the Inheritance of Agricultural Land . 9
1.2.2 Land Markets in Rural India . . . 10
1.3 Conceptual Framework . . . 11
1.4 Data . . . 14
1.5 Empirical Strategy . . . 16
1.6 Results . . . 20
1.6.1 Summary Statistics . . . 20
1.6.2 First Stage and Reduced Form Estimates . . . 22
1.6.3 Occupational Choice, Migration and Consumption . . . 27
1.7 Mechanisms . . . 33
1.7.1 Mechanisms: Access to Credit . . . 33
1.7.2 Mechanisms: Culture as a Friction on Labor Markets . . . 35
1.7.3 Mechanisms: Transaction Costs in the Market for Land . . . 40
1.8 Robustness Tests . . . 42
1.8.1 Selection Concerns from Urban Migration . . . 42
1.8.2 Addressing Instrument Validity: Conditional Independence Assumption 44
1.8.3 Addressing Instrument Validity: Exclusion Restriction Assumption . 48
1.9 Conclusion and Discussion . . . 51

2 The Value of Advice: Evidence from the Adoption of Agricultural Practices **54**

2.1 Introduction . . . 54
2.2 Context and Intervention Description . . . 58

2.2.1 Agricultural Extension . . . 58
2.2.2 Avaaj Otalo: Mobile Phone-Based Extension . . . 60
2.3 Experimental Design & Empirical Strategy . . . 62
2.3.1 Summary Statistics and Balance . . . 67
2.4 Experimental Results . . . 68
2.4.1 First Stage: Take-Up and Usage of AO . . . 70
2.4.2 Impact on Sources of Information for Agricultural Decisions . . . 73
2.4.3 Overall Impact on Input Adoption . . . 75
2.4.4 Impact on Seed Selection . . . 76
2.4.5 Pest Management Practices . . . 76
2.4.6 Fertilizers . . . 78
2.4.7 Sowing and Productivity . . . 79
2.4.8 Impact on Agricultural Knowledge . . . 80
2.4.9 Heterogeneous Treatment Effects . . . 82
2.4.10 Peer Effects . . . 84
2.4.11 Willingness to Pay . . . 86
2.4.12 Cost-Benefit Analysis . . . 87
2.5 Threats to Validity . . . 90
2.5.1 Attrition . . . 90
2.5.2 Experimenter Demand Effects . . . 90
2.6 Conclusion . . . 91

3 Social Interactions, Technology Adoption and Information Exchange: Evidence from a Field Experiment 94
3.1 Introduction . . . 94
3.2 Context: Mobile Phone-Based Agricultural Extension in Rural India . . . 98
3.3 Conceptual Framework . . . 99
3.4 Data and Empirical Strategy . . . 101
3.5 Experimental Results . . . 103
3.5.1 Summary Statistics and Balance . . . 103
3.5.2 Usage of Mobile Phone-Based Extension Service . . . 107
3.5.3 Social Interactions and Information Exchange . . . 107
3.5.4 Peer Effects: Structure of Social Interactions and Agricultural Outcomes 111
3.5.5 Peer Effects: Agricultural Outcomes . . . 114
3.6 Discussion of Mechanisms . . . 116
3.7 Threats to Validity . . . 117
3.7.1 Attrition . . . 117
3.8 Conclusion . . . 117

References 119

Appendix A Appendix to Chapter 1 125

A.1 Data Appendix and Summary Statistics 125

Appendix B Appendix to Chapter 1 134

B.1 Additional Results 134

Appendix C Appendix to Chapter 1 141

C.1 Additional Robustness Tests 141

Appendix D Appendix to Chapter 2 155

D.1 Supplementary Tables and Figures 155

Appendix E Appendix to Chapter 3 165

E.1 Supplementary Tables and Figures 165

List of Tables

1.1 Summary Statistics for Head-Level, Sibling-Level and Child-Level Data . . . 21
1.2 The First Stage and Reduced Form Estimates 24
1.3 The Effect of Inherited Land on Occupational Choice, Migration and Household Consumption 31
1.4 The Effect of Inherited Land on Borrowing 34
1.5 Heterogeneous Effects of Inherited Land by Birth Order 37
1.6 The Effect of Parent's Inherited Land on Child Outcomes by Birth Order . . 41
1.7 Heterogeneous Effects of Inherited Land by Transaction Costs in the Market for Land 43
1.8 Balance Check for Instrument 47
1.9 The First Stage and Reduced Form Effects in States with Matrilineal Inheritance Rules 50

2.1 Summary Statistics and Balance 69
2.2 Usage of Avaaz Otalo (AO) Information Service 71
2.3 Effects on Sources of Agricultural Information 74
2.4 Effects on Input Adoption Decisions 77
2.5 Effects on Sowing Decisions and Agricultural Productivity 79
2.6 Effect on Agricultural Knowledge 81
2.7 Heterogeneous Effects by Education and Income 83
2.8 Peer Effects on AO Usage, Sources of Information and Sowing 85
2.9 Willingness to Pay for AO Information Service 88

3.1 Summary Statistics and Balance - General 104
3.2 Summary Statistics and Balance - Sources of Information and Social Interactions 106
3.3 AO Usage and Content 108
3.4 Effect of AO on Social Interactions and Exchange of Information 109
3.5 Peer Effects - Social Interactions and Exchange of Information 113
3.6 Peer Effects - AO Usage, Sources of Information and Sowing 115

List of Figures

1.1 Rule-Based and Empirical Inheritance Shares of Family Land 18
1.2 Distribution of Inherited Land and Changes in Landholdings 25
1.3 Visualization of First Stage 26
1.4 Non-Parametric Visualizations of Reduced Form Equations 28
1.5 Non-Parametric Visualizations of Reduced Form by Birth Order 39

2.1 Experimental Design 64
2.2 Project Timeline 65
2.3 AO Usage by Month 72
2.4 Demand Curve from Willingness to Pay Experiments 89

Acknowledgments

Many people have helped me in this undertaking. I'd especially like to thank my parents, who have made immense sacrifices to give me terrific opportunities. They gave me tremendous freedom to pursue my interests and were endlessly accommodating of many changes of course. I'd also like to thank my advisors: Asim Khwaja, Larry Katz Rick Hornbeck, and Shawn Cole. In particular, I'd like to thank Asim for numerous insights about chapter 1 and Shawn for nearly a decade of constant support and friendship. Id also like to thank Jim Boyce and Vivek Bhandari for their advice and encouragement in pursuing graduate studies in a purposeful way.

I'd especially like to thank Niharika Singh for her friendship, support and patience throughout graduate school. Many friends at Harvard have made this journey more fun and rewarding. John Marshall and Nils Hagerdal have been terrific friends. John has been a model of what is achievable without procrastination, while Nils' good nature seems without limit. Anitha Sivasankran has been an extraordinary friend and colleague. Other friends too sat through many a practice talk. I'd especially like to thank Joey Neggers, Sam Stolper, Janhavi Nilekani, Shayak Sarkar, Daria Pelech and Raissa Fabregas. I'd also like to thank Nicole Tateosian for her support and perpetual cheeriness.

Many families have been kind to me over the years and have often made a foreign country feel like home. This is especially true of Katie, Dorri and Yao Li, Mike and Ursula Apte, Ana Rona and Gabriel Santos. My closest friends, Gabriel Ceriotti and Zac Apte have been nothing short of immense over the years. Nadisha Sirisena has always made home feel familiar after extended periods away.

Much of the work in chapters 2 and 3 would not have been possible without the terrific support from field staff at CMF over many years. In particular, I'd like to thank Tarun Pokiya, Sachin Srivastava, Avdhut Fadanwis, Niriksha Shetty, Tanaya Devi, Ishani Desai, Rubina Sheikh, Janki Parmar, Manish Sharma, Dhruti Purohit, Prashant Parmar, Abhay Desai, Ramesh Machhi and the many field staff who collected data on our behalf.

In addition, my research has benefited from numerous comments from Josh Angrist, Bill

Clark, Michael Kremer, Lakshmi Iyer, Rohini Pande, Nathan Nunn, Supreet Kaur, Rema Hanna, Jeremy Tobacman and participants of the Harvard Development and Labor lunches.

Finally, I thank Andrew Foster for sharing data for chapter 1 and I'd like to acknowledge support from USAID, the Agricultural Technology Adoption Initiative and the Sustainability Science Program at the Harvard Kennedy School.

To my mother and two retiring Sri Lankan legends.

Introduction

Understanding the factors that influence the movement of labor out of the agricultural sector during the process of development is an enduring question studied famously by Lewis (1954) and Harris and Todaro (1970). This transition carries with it the promise of spectacular reductions in poverty as witnessed most recently in the case of China. Over the last two decades, the share of the Chinese workforce engaged in agriculture has dropped by approximately 30%, while the share of the population living in poverty has fallen by 80% (IBRD, 2014).[1]. However, in many developing countries, rural populations continue to be reliant on agriculture with their welfare subject to the vagaries of weather and its effect on agricultural production.

India, the focus of this dissertation, is a case in point. The Indian labor force has experienced a much smaller shift out of agriculture (15%) of the last two decades, a fact symptomatic of a general absence of rural industrialization. During this period, the headcount poverty ratio in India reduced by 34% which, while impressive, leaves much to be desired in comparison to the Chinese growth experience. These comparisons have led commentators to suggest an important role for the development of the non-agricultural sector in poverty reduction and indicate the continuing relevance of Lewis' original inquiry (Bardhan, 2012).

Given the importance of this question and the accumulated evidence on disparities in the returns to labor between the agricultural and non-agricultural sectors across the

[1]The figures reported here are for 1994 and 2010 and the headcount ratio is calculated as the fraction of the population under $1.25 per day

developing world (Gollin *et al.*, 2014b), I've sought to understand what influences the mobility of labor out of the agricultural sector at a microeconomic level. In addition to this direct economic rationale, this question was motivated by John Rawls characterization of 'fair equality of opportunity', the extent to which the occupational trajectories of the rural poor are influenced by their initial circumstances and how markets, states and communities should act in response Rawls (1971).

The first chapter of this dissertation follows in this spirit and seeks to understand how the inheritance of agricultural land influences occupational trajectories. While the inheritance of productive assets is ordinarily thought to be beneficial, where market frictions and embedded cultural obligations limit the exchange or ability to relinquish such assets, they may restrict access to high-return economic opportunities and undermine asset benefits causing a microeconomic parallel to the 'resource curse. Using variation arising from sibling sex composition and Hindu inheritance customs that favor sons, I test this hypothesis by estimating the long-term causal effect of inheriting land in rural India. Consistent with standard models, inheriting land facilitates borrowing and increases household consumption. Yet, where the ability to fully utilize land through markets is severely constrained by frictions, either cultural obligations or land market transaction costs, the effect on consumption is entirely attenuated and negative for a subset of the sample. Those who inherit land are significantly less likely to migrate to urban areas and enter non-agricultural work in rural areas; effects that are accentuated by such frictions.

The remaining chapters of this dissertation are policy-oriented and prescriptive in nature, with the intention of developing evidence on what can be done to address persistent rural poverty in the agricultural sector. Technological innovations that provide valuable information to farmers hold tremendous promise in improving agricultural productivity and narrowing the observed differences in agricultural productivity (Gollin *et al.*, 2014a). In the second chapter, co-authored with Shawn Cole, we report the results of a randomized evaluation of the introduction of a mobile phone-based agricultural consulting service, Avaaj Otalo (AO), to farmers in Gujarat, India. We find that demand for agricultural advice is substantial and farmers offered the service turn less often to traditional sources of agricultural advice. Management practices change as well: farmers invest more in recommended agricultural inputs resulting in dramatic increases in yield for cumin (26.3%), and improvements in cotton yield (3.5%) for a sub-group that received frequent reminders to use the service. Peers of treated farmers report limited changes in their information sources and cropping decisions. Farmers appear willing to follow advice without understanding why it is correct: we do not observe gains in agricultural knowledge. We estimate that each dollar invested by a farmer in the service generates a return of $10. These findings highlight the importance of managerial practices in facilitating technology adoption and improving the productivity of agriculture.

The final chapter of this dissertation asks whether such information and communication technologies (ICT's) influence the pre-existing structure of social interactions. Specifically, I estimate the effect of a mobile phone-based agricultural extension service on the exchange of information in village India and its influence on agricultural outcomes. Treated respondents make extensive use of the service but are initially no more likely to share information with their peers than control respondents. However, after production outcomes have been observed, treated respondents are both more likely to share information and recommend inputs to their peers. Yet, treated respondents are less likely to receive information from their peers and observe their fields in person. The sources of shared information change dramatically: treated respondents are substantially more likely to report sharing mobile phone-based information in comparison to traditional sources. Treated respondents are differentially more likely to visit the homes of treated peers to discuss agricultural topics, suggesting that the provision of external information may increase the returns to in-person social interactions. These changes also influence 'real economic outcomes: individuals outside the original study with treated peers in their network are less likely to report crop loss due to pest attacks and are more likely to cultivate cumin. These results suggest that technological innovations can influence the structure and content of social interactions and influence agricultural outcomes.

Chapter 1

Shackled to the Soil: The Long-Term Effects of Inherited Land on Labor Mobility and Consumption

1.1 Introduction

> *"I see young men, my townsmen, whose misfortune it is to have inherited farms ... ; for these are more easily acquired than got rid of. Who made them serfs of the soil?"*
>
> - Henry David Thoreau, Chapter 1, *Walden*, 1854

Asset endowments can expand access to economic opportunity. Models of poverty traps predict that initial endowments may ease borrowing constraints and allow the rural poor to take advantage of high-return opportunities outside agriculture (Banerjee and Newman, 1993; Galor and Zeira, 1993).[1] In so doing, assets may facilitate the process of 'structural transformation' predicted by influential theories of economic development (Lewis, 1954;

[1]Recent evidence suggests that value-added in the agricultural sector is less than half of that in the non-agricultural sector across 100 developing countries, leading the authors to conclude that labor is 'greatly misallocated' in these countries (Gollin *et al.*, 2014b)

Harris and Todaro, 1970) that is borne out in the recent empirical literature.[2] However, where selling, renting, or leaving behind assets is restricted by cultural obligations and market frictions, they may prevent the poor from taking advantage of high-return opportunities by limiting their spatial and occupational mobility. As a consequence, assets may have no effect or even reduce the level of household consumption: a microeconomic parallel to the 'resource curse' hypothesis (Auty, 1993).[3]

Using a combination of data I collected and an existing dataset spanning 16 Indian states, I test this hypothesis by estimating the long-term causal effect of inheriting agricultural land – a highly illiquid and important asset for the rural poor – on the occupational and spatial mobility of labor and on household consumption. My identification strategy leverages a custom among Hindu families that results in sons inheriting equal shares of their parents' land. Conditional on the number of siblings an individual has and their parents' landholdings, sibling sex composition serves as an instrument for inherited land. I find that an additional brother reduces the amount of land inherited by more than an acre, or *one-third* of median landholdings in rural India.

In contrast to theories emphasizing the importance of initial endowments in the context of credit constraints, I find that 14 years later (median) inheriting land both restricts migration to urban areas (-0.02% per acre) and reduces the likelihood of entering non-agricultural work in rural areas (-1.8% per acre), even though it eases borrowing constraints.[4] However, these estimates mask substantial non-linearities: the point estimates are more than ten times as large (-3.4% per acre for migration; -21% per acre for non-agricultural occupation) for smaller inheritances that are below the median of the land distribution (3 acres). Inheriting land *increases* household consumption on average (2.7% per acre) but where a cultural

[2] Experimental and non-experimental evidence shows that asset transfers allow the poor to attain a higher level of consumption and may in addition support entry into the non-agricultural sector (Besley and Burgess, 2000 ;Bandiera *et al.*, 2011; Blattman *et al.*, 2013)

[3] Auty (1993) argues that countries reliant on natural resource wealth are less able to diversify industrial production, restricting production in sectors in which they may development a comparative advantage and limiting growth relative to countries with less resource wealth.

[4] Specifically, inheriting an additional acre increases the probability of taking out a loan in the last five years by 1.5% and the value of loans by 15.2%.

obligation and land market frictions limit the ability to fully utilize land through markets, it has *no effect* on consumption and, in some cases, the effect is *negative*.

The cultural obligation in question requires that the eldest son in a Hindu family support his parents in their old age, often resulting in the expectation of occupational succession. In contrast to their latter-born counterparts, inheriting land is even more likely to restrict urban migration and entry into non-agricultural work for first-born sons.[5] While the household consumption of later-borns increases by 3% for each acre inherited, for first-borns the effect of inheriting land is indistinguishable from zero. The implied loss in consumption resulting from this friction is almost 9% for a median inheritance.[6] Furthermore, the importance of this custom persists across generations: I find that the probability of a first-born *child* migrating is *decreasing* in their parents' landholdings.

Transaction costs are another important friction limiting participation in land markets across rural India. Across multiple measures of village-level transaction costs in the market for land, I find that inheriting land in villages with higher transaction costs leads to a significantly larger effect on persisting in agriculture and attenuates the consumption benefits of inheriting land.

Collectively, these results highlight the competing effects of inherited land on consumption and the mechanisms through which they operate. Inheriting land has the expected wealth effect that leads to an increase in consumption. However, where cultural obligations and frictions are salient, inheriting land limits the occupational and spatial mobility of inheritors thereby restricting access to high-return opportunities. This 'shackling' effect of land is also dynamic in nature: first-borns inheriting at an earlier age are even more likely to remain in agriculture, and their level of consumption is *decreasing* in the extent of inherited landholdings. The balance of these effects determine the overall effect of inheriting land on consumption.

---

[5] I use 'first-born' to refer to respondents who are the first-born son in their family and 'latter-born' to refer to all other sons.

[6] This consumption loss also allows me to bound the size of any non-pecuniary benefit a first-born would need to experience from observing this custom to leave their *welfare* comparable to that of latter borns.

I conduct a series of robustness tests to address identification concerns. A selection issue arises from unobserved *family* migration to urban areas. While this form of migration is considered very rare in rural India, I perform simulations which show that the reduced form results are robust to selective migration. In support of the conditional independence assumption of the instrument, I show that outcomes which precede the inheritance of land – including individual and parent characteristics – are not correlated with the instrument. Additionally, I show that the main estimates are robust to a number of controls intended to account for son-preferring fertility behavior including sex selective abortion, differential care and differential stopping rules.

The independent effects of sibling sex composition on human capital acquisition and dowry payments present another threat to identification. Areas with historically matrilineal inheritance customs serve as a placebo test as they should have a weaker first stage but any other effects of sibling sex composition should still influence the outcomes under consideration. Reassuringly, the reduced form effects in these areas are significantly different from areas with patrilineal inheritance customs. Similarly, I find no reduced form effects of sibling sex composition on the main outcomes of interest for individuals whose parents were landless – individuals who cannot, by definition, inherit land. In each of these cases, I also show that the first stage for alternative causal channels – human capital and dowry payments – does not vary significantly across these contexts.

This paper primarily contributes to a literature seeking to understand the frictions and barriers that restrict the movement of labor across sectors in the developing world. A recent essay on the contributions of the Lewis model notes that such frictions, particularly as they relate to other factor markets, are poorly understood (Gollin, 2014). Other contributions in this literature focus on barriers in credit, insurance, information and transportation that restrict the allocative efficiency of labor markets (Blattman *et al.*, 2013 ; Bianchi and Bobba, 2012 ; Bryan *et al.*, 2011 ; Gollin and Rogerson, 2010). This paper complements these findings by looking at frictions in land and labor markets and providing evidence of long-term effects by examining occupational transitions over more than a decade.

Second, this paper highlights the importance of asset specificity in anti-poverty programs and informs the debate on land reform.[7] Relatively illiquid assets such as cattle or land are often the focus of anti-poverty programs and some, such as large-scale targeted 'ultra poor' programs in Bangladesh, may even require that beneficiaries do not liquidate their assets (Das *et al.*, 2013). These assets may still serve as collateral in the presence of credit constraints, but the inability to easily sell the asset may constrain rather than support occupational transitions thereby harming household consumption. To this point, Das *et al.* (2013) find that transfers of cattle to women in 'ultra-poor' households in Bangladesh constrain their ability to work outside the household. Using the same dataset, Gulesci (2012) finds that this program has general equilibrium effects on the wage rates of non-beneficiary households.

Third, these findings are consistent with a literature estimating how changes in land markets influence labor supply. In particular, Chernina *et al.* (2013) find that titling reforms in Russia that enabled the sale of previously communally owned land supported internal migration by easing credit constraints. These findings are also consistent with the literature on titling in land markets, where reductions in transaction costs allow for the efficient reallocation of labor across space (Field, 2007; De Janvry *et al.*, 2012). Similarly, Wang (2012) finds that property reform in China allowing state employees to purchase homes at a subsidized price lead to increased job mobility through entrepreneurship.

Finally, this paper contributes to an understanding of the connection between the allocative efficiency of labor markets and inherited assets. While the context and the part of the distribution of wealth under consideration differ greatly to Piketty (2014), the findings are a corollary of his examination of the intergenerational transmission of advantage. In this case, as a consequence of the fact that the assets bequeathed by parents are illiquid rather than liquid, inheritors may be inadvertently made worse off in the future. However, in both cases, the implication is that inherited assets may influence the allocative efficiency of labor

[7] A survey of research on land markets concludes that estimates of their welfare effects are inhibited by identification challenges resulting from the non-random assignment of land and a lack of longitudinal data (Deininger and Feder, 2001). Recent work uses 19th century land lotteries in the US state of Georgia to estimate the causal effect of land, but the authors are unable to evaluate the impact on measures of household consumption for lack of data (Bleakley and Ferrie, 2010;Bleakley and Ferrie, 2013).

markets.

The next section provides context on the inheritance of land and land markets in rural India. Section 3 motivates the empirical strategy with a conceptual framework. Section 4 describes the data sources used in the empirical analysis. Section 5 describes the empirical strategy, while Section 6 presents the results from this analysis. Section 7 discusses heterogeneity in the results and the mechanisms underlying the observed effects. Section 8 describes robustness checks that test the validity of the identifying assumptions of the empirical strategy. Section 9 concludes.

1.2 Context: Agricultural Land in Rural India

1.2.1 Customs and Laws Governing the Inheritance of Agricultural Land

The majority of Hindu communities throughout rural India are characterized by patrilineal land inheritance customs. The ethnographic literature suggests that such customs largely hold sway over recent progressive reforms – in many cases agricultural land is exempt from such reforms – continuing to restrict the ability of women to inherit agricultural land. To this point, in her exhaustive study of gender and land rights in India, Agarwal (1994) concludes that women seldom inherit land.[8] However, the north eastern states of India and the southern state of Kerala are an exception where matrilineal inheritance customs continue to be more prevalent.[9]

In patrilineal areas, agricultural land is typically inherited after the death of the father, with sons inheriting equal shares of their father's land.[10] While recent reforms to the Hindu

[8]In particular, she states: "Ethnographic information, although it is extremely fragmentary, consistently indicates that women in traditionally patrilineal communities of South Asia rarely realize the rights that contemporary laws have promised them. Custom still dominates practice. Hence the vast majority of women do not inherit landed property as daughters, most dont do so even as widows and few women inherit in other capacities. To the extent women inherit is usually under very restricted conditions." ((Agarwal, 1994))

[9]The north eastern states include Assam, Arunachal Pradesh, Manipur, Mizoram, Sikkim, Tripura and Meghalaya

[10]Foster and Rosenzweig (2002), in the their study of rural household division, also suggest that equal division among sons is the norm. Technically, inheritance claims extend to four generations of agnates, implying

Succession Act have sought to give women equal claims to land, data I collected from 1,037 landed households engaged in agriculture in the western Indian state of Gujarat supports equal division and continued primacy of custom over law.[11] Among these respondents, 82% described standard practice as equal division among male sons while 15% claimed all siblings inherited equal amounts (See Appendix A2.1).

1.2.2 Land Markets in Rural India

Deininger *et al.* (2009) suggest that both micro studies focusing on collections of villages and nationally representative datasets point to very limited participation in sales and rental markets for agricultural land in rural India.[12] Similarly, Skoufias (1995) uses ICRISAT data in India to show that 75% of households are unable to meet their 'desired cultivated area' – predicted landholdings based on livestock and family labor endowments – using land markets. Reasons for limited market participation in India include rental restrictions (Deininger *et al.*, 2008), regulatory restrictions preventing the sale of land for non-agricultural purposes and high stamp duties on land transactions (Morris and Pandey, 2007).

Generic land market imperfections include poorly defined property rights resulting in uncertainty over ownership claims (Deininger and Goyal, 2012) and information asymmetries in assessing quality of land and effort of tenants (Deininger and Feder, 2001). An important barrier to land sales and rental – of particular consequence to this paper – is a desire for farmers to continue in the tradition of their ancestors (Jodhka, 2006). While the latter study focuses on the Indian state of Punjab, a survey of landed adults aged 18-30 across 13 states in rural India (Sharma, 2007) found that 60% of respondents had no intention of selling

---

that grandsons also have a claim upon birth. In practice these shares are typically claimed and registered after the more senior member in the agnatic line dies.

[11] The Hindu Succession Act (HSA) of 1951 sought to unify differing legal traditions deriving from *Shastric* texts but fell short of giving both sons and daughters equal claims to ancestral property. Amendments were made to the HSA by Kerala (1976), Andhra Pradesh (1986) and Tamil Nadu (1989) to enable women to have equal inheritance rights.

[12] In the dataset used for the majority of the empirical analysis below, over a 20-year period 7.34% of households sold land and 13.6% bought land. In the past year, 2.89% of households leased in land, 8.63% leased out land, 4.89% engaged in any type of sharecropping and just 0.4% of the sample mortgaged their land.

their land, with 34% suggesting that farming was a 'mark of their identity' and they would like their children to cultivate their land as had been done for generations.

1.3 Conceptual Framework

The generic effect of inheriting land is an outward shift of the budget constraint affording an individual a more desirable consumption bundle: a 'wealth effect'. Additionally, in the context of credit constraints individuals may be able to leverage land as collateral and take advantage of high-return opportunities in the non-agricultural sector. However, if land markets are severely constrained by frictions, inheriting land also involves an opportunity cost. An inability to part with land through sales or rental markets – or, at an extreme, vacate it – may limit an individual's spatial mobility or ability to diversify into other occupations within rural areas. Estimating the marginal effect of inherited land will ordinarily combine both these effects.

To clarify these competing effects and motivate the empirical strategy, consider a conceptual experiment with three groups: a 'land' group that is randomly assigned an acre of land, a 'cash' group that is randomly assigned the equivalent value in cash and a control group which receives nothing. Assume that the 'land' group is prohibited from selling, renting or leaving the land; it is assumed to not be in their interest to leave it fallow.[13] However, there exists a market where the 'cash' group may purchase land. We could then estimate the effect of our treatments on household consumption:

$$\begin{aligned}
\beta_w &= \mathbb{E}(Consumption \mid Cash = 1) - \mathbb{E}(Consumption \mid Control = 1) && (1.1)\\
\beta_{w-c} &= \mathbb{E}(Consumption \mid Land = 1) - \mathbb{E}(Consumption \mid Control = 1) && (1.2)\\
\beta_c &= \mathbb{E}(Consumption \mid Land = 1) - \mathbb{E}(Consumption \mid Cash = 1) && (1.3)
\end{aligned}$$

β_w is the marginal effect of the cash endowment on household consumption and β_{w-c} is the marginal effect of an acre of land on household consumption. By assumption $\beta_w \geq \beta_{w-c}$

[13]It is worth noting here that many asset transfer programs prohibit the resale of assets to ensure that beneficiaries are provided with a basis for a livelihood rather than a temporary wealth shock (Das *et al.*, 2013).

as the latter effect involves an opportunity cost, c. This cost c can be thought to derive from foregone returns in the non-agricultural sector imposed by the requirement that the 'land' group is unable to fully leverage this asset through markets.

Theoretically it is possible that β_{w-c} could be positive, zero or even negative. For example, consider an individual in the 'land' group forced to make a living off an acre of land, while their counterfactual outcome in the control group would have been to get a job at a call center in the city. If the opportunity cost of remaining in agriculture were larger than the wealth effect we would get the perverse result that our control group has *higher* average household consumption than the 'land' group.[14] Finally, assuming an additive marginal effect structure, we could estimate the size of this cost c by comparing the outcomes of the 'land' group to the 'cash' group, i.e. β_c in equation (1.3).

In practice, implementing the experiment described above may not be feasible both on account of its cost and, more importantly, because the institutional constraints required to capture these effects cannot reasonably be imposed on experimental subjects. In contrast, using plausibly exogenous variation in inherited landholdings I am able to recover estimates of these marginal effects by exploiting heterogeneity in factor market frictions.[15] However, in contrast to the conceptual experiment, there is no 'cash' group and I estimate the effect on the intensive margin for land.

To make the first point of departure clear, the empirical strategy detailed in section 5 produces the equivalent of the 'land' group and the 'control' group. However, in this case the 'land' group is *not* prohibited from selling, renting or leaving the land. Instead, those who own land face different values of $\theta \in [0,1]$ a parameter that captures frictions in factor markets. Where θ is higher, it is more difficult to sell, rent or leave inherited land. In the empirical strategy detailed below θ is approximated by cultural obligations and transaction

[14] Alternatively, in a constrained optimization framework we can motivate this idea by assuming that selling land involves a non-pecuniary cost α such that $\alpha > \beta_c$. This implies that the 'land' group may be just as well off in a welfare sense but there are observable implications in terms of occupational choice and consumption.

[15] The fact that land is inherited may also make the non-pecuniary costs of parting with this land more salient as a consequence of having farmed it for generations.

costs in the market for land.

Considering the extremes, an individual inheriting land that faces $\theta = 0$ can be compared to someone assigned to the 'cash' group in the conceptual experiment : they can just sell the land and get the equivalent cash value. However, for those who own land and face $\theta = 1$ it would be as though they were assigned to the 'land' group from the conceptual experiment. In this case, it is clear that causal estimates will recover a weighted average of $\hat{\beta}_w$ and $\hat{\beta}_{w-c}$ in estimating the marginal effect of household consumption and its analogous effect on occupational choice. In addition, we can use the heterogeneity in θ to estimate $\hat{\beta}_c$, the opportunity cost of inheriting land:

$$\begin{aligned} \hat{\beta}_c &= [\mathbb{E}(Consumption \mid Land = 1, \theta = 1) - \mathbb{E}(Consumption \mid Control = 1, \theta = 1)] \\ &\quad - [\mathbb{E}(Consumption \mid Land = 1, \theta = 0) - \mathbb{E}(Consumption \mid Control = 1, \theta = 0)] \end{aligned} \tag{1.4}$$

Second, I estimate the difference in consumption for individuals with varying sizes of landholdings, i.e. the intensive margin. To map this to the conceptual experiment, imagine an individual who gets two acres of land relative to someone who gets an acre. While the pure wealth effect of 2 acres of land is larger than 1 acre (i.e. $\beta_{w,2acres} > \beta_{w,1acre}$) the change in the opportunity cost is indeterminate (i.e. $\beta_{c,2\ acre} \geq \beta_{c,1\ acre}$). An increase in wealth may afford an individual alternatives that were not available to someone with less wealth (e.g. the cost of transportation to a more remunerative market). Another way of conceiving this is that θ itself may depend on the size of landholdings and, as such, the level of consumption with 2 acres of land could be larger, equal to or smaller than with 1 acre of land.[16] While the empirical strategy estimates local average treatment effects, non-parametric estimations of the reduced form can reveal non-linearities in the estimated effects.

[16]Larger landholdings may absorb more of an individuals time endowment resulting in less opportunity to diversify labor supply within rural areas: this implies that θ is increasing in the extent of landholdings.

1.4 Data

The data used in this paper primarily draw from the 1999 wave of the ARIS/REDS[17] data set (hereafter, 'REDS') collected by the National Council for Applied Economic Research (NCAER). The REDS data is a national probability survey intended to be representative of the rural population of India residing in 17 major states and 100 districts. It's distinctive features include a complete enumeration of respondent's siblings and children – not limited to those present at the household at the time of surveying – and data on a respondent's inherited landholdings and parent's landholdings. Additionally, the survey contains detailed data on consumption, non-farm and agricultural investments and labor supply.

The REDS panel was collected in four waves conducted between 1971 and 2006 and has previously been used in a number of prominent studies of the Green Revolution in India (Foster and Rosenzweig, 1995;Foster and Rosenzweig, 1996). The first round of the survey randomly sampled 4527 households in 259 villages, stratifying by farm size and wealth. The survey was originally intended to evaluate the impact of an agricultural development program, but was expanded beyond program districts in 1982 with the intention of making it nationally representative. All of the original villages were surveyed in the 1999 wave, excluding 8 sample villages from Jammu and Kashmir (owing to problems of local insurgency). Because of household divisions and the inclusion of a new random sample of households in each village, the number of households in the 1999 round increased to 7474.[18]

The main sample used in the analysis presented below uses data from the 1999 wave and drops household heads whose parents owned no land (1,654 households) as by definition any landholdings they possess are not governed by the inheritance laws described above. While household heads from landless families are dropped from the main analysis, they are used in placebo tests to evaluate the validity of the exclusion restriction. Among the

[17] Additional Rural Incomes Survey/Rural Economic and Demographic Survey

[18] After merging in data across parts of the survey that includes separately elicited information on all siblings and children, the total number of households in the dataset drops to 7393.

remaining 5,793 households, 338 female headed households are dropped because the data does not indicate whether the sibling data refers to the female head or her spouse, and a further 592 households of minorities are dropped as the study is restricted to Hindu households.[19]

The primary unit of analysis in this paper is the household head. However, data on all siblings and children of the household head are collected for a limited set of outcomes including education, migration, inherited and current landholdings. The very low level of permanent migration to urban areas in the period preceding the survey and the fact that most respondents were born prior to the introduction of affordable sex selection technologies make the 1999 wave particularly well suited to this analysis.[20] In addition, this wave was the first to include detailed information on household structure, inheritances and agricultural labor.

In addition to the REDS data, this paper also makes use of the Indian Human Development Survey (IHDS) which was conducted in 2004-2005. The IHDS is a nationally representative, multi-topic survey of 41,554 households in 1503 villages and 971 urban neighborhoods across India. This survey is primarily used in analysis supporting the exclusion restriction owing to the larger sample size and urban sample. However, the IHDS does not collect information on land inheritances or family data on heads aside from the total number of siblings making it ill suited to the main analysis.

Finally, I collected data on household perceptions, understanding and the administration of inheritance rules from a sample of 1,200 households in rural Gujarat. These households were randomly selected from village lists of cotton farmers as a part of a separate study on technology adoption in agriculture (Cole and Fernando, 2014). This data was collected

[19] Note, the incorporation of households splitting over time after the 1982 round suggests that successive rounds cannot be considered nationally representative.

[20] In spite of low levels of permanent migration to urban areas, a potential selection concern arises from unobserved *family* migration to urban areas. While I observe whether all siblings and children of household heads migrate, I would not observe a household if entire families migrated to urban areas which may influence the estimates that follow. While this form of migration is considered very rare in rural India (Munshi and Rosenzweig, 2007), I perform a series of simulations that assess the robustness of the reduced form results to selective migration in Section 8.1, which are described in further detail in Appendix C2.

through paper based surveys in coordination with the Centre for Micro Finance in August, 2013.

1.5 Empirical Strategy

In the empirical analysis, I estimate the causal effect of inherited landholdings L_{ij} on household consumption and labor mobility, Y_{ij}. In the case of the latter outcome, let Y_{ij} be a dummy for holding a non-agricultural occupation. The structural equation of interest is:

$$Y_{ij} = \alpha_j + \rho L_{ij} + X_{ij} + \nu_{ij} \tag{1.5}$$

Where α_j is a district fixed effect and X_{ij} is a set of controls which include characteristics of i and his family background that might influence occupational choice.[21] The concern here is that L_{ij} is correlated with ν_{ij}. For example, people who own more land may also have higher ability, A_{ij}, and this may in turn be positively correlated with exiting agriculture. When A_{ij} omitted, OLS will biased upwards relative to ρ. In order to address this concern, I make use of an inheritance rule that results in land being divided equally between sons after a father's death to instrument for L_{ij}.

If the number of brothers an individual has is the product of a random process and the rule is binding, then, conditional on the number of siblings he has, it must also be the case that his inheritance share is the product of a random process.[22] The functional form for this share is non-linear and equal to: $\textit{Predicted Share} = [\frac{1}{1+\textit{Brothers}}]$. Panel A and B in Figure 1.1 demonstrate the validity of this functional form assumption by showing how the empirical shares of inherited land vary with sibling sex composition.[23] Panel A plots inheritance

[21] If we instead change the outcome variable to consumption, ρ is a combination of β_w in equation (1.1) and β_{w-c} in equation (1.2) as the estimate is averaged over values of θ, the measure of frictions in the conceptual framework.

[22] Sibling sex composition has previously been used in studies examining the effect of child bearing on labor supply and the 'quantity-quality' fertility trade-off (Angrist and Evans (1998); Angrist *et al.* (2010)).

[23] Appendix Table A2.2 and A2.3 provide additional support for equal division. Virtually all (93%) households among the 1,037 surveyed in Gujarat report that parents do not deviate from the equal shares rule because of mitigating circumstances such as higher human capital or employment. Furthermore, a large majority (71%)

shares observed in the data by a household head's number of brothers (black dots). The empirical shares closely track what is predicted by the inheritance rule (red-dashed line). In contrast, Panel B shows that no such relationship exists in the analogous visualization by number of sisters.[24]

To illustrate the use of this instrument, consider the case of a respondent with one sibling. If that sibling is a brother then the individual inherits half the family land and the value of the instrument, *Predicted Share*, equals 0.5 . If, on the other hand, the respondent has as sister, the instrument equals 1 and he inherits all the family land. The identifying assumptions are that conditioning on the respondent's total number of siblings and his parent's landholdings, his predicted inheritance share is independent of potential outcomes and only affects these outcomes through the inheritance of land. The validity of these assumptions – the conditional independence assumption and the exclusion restriction assumption – are addressed in Section 8.

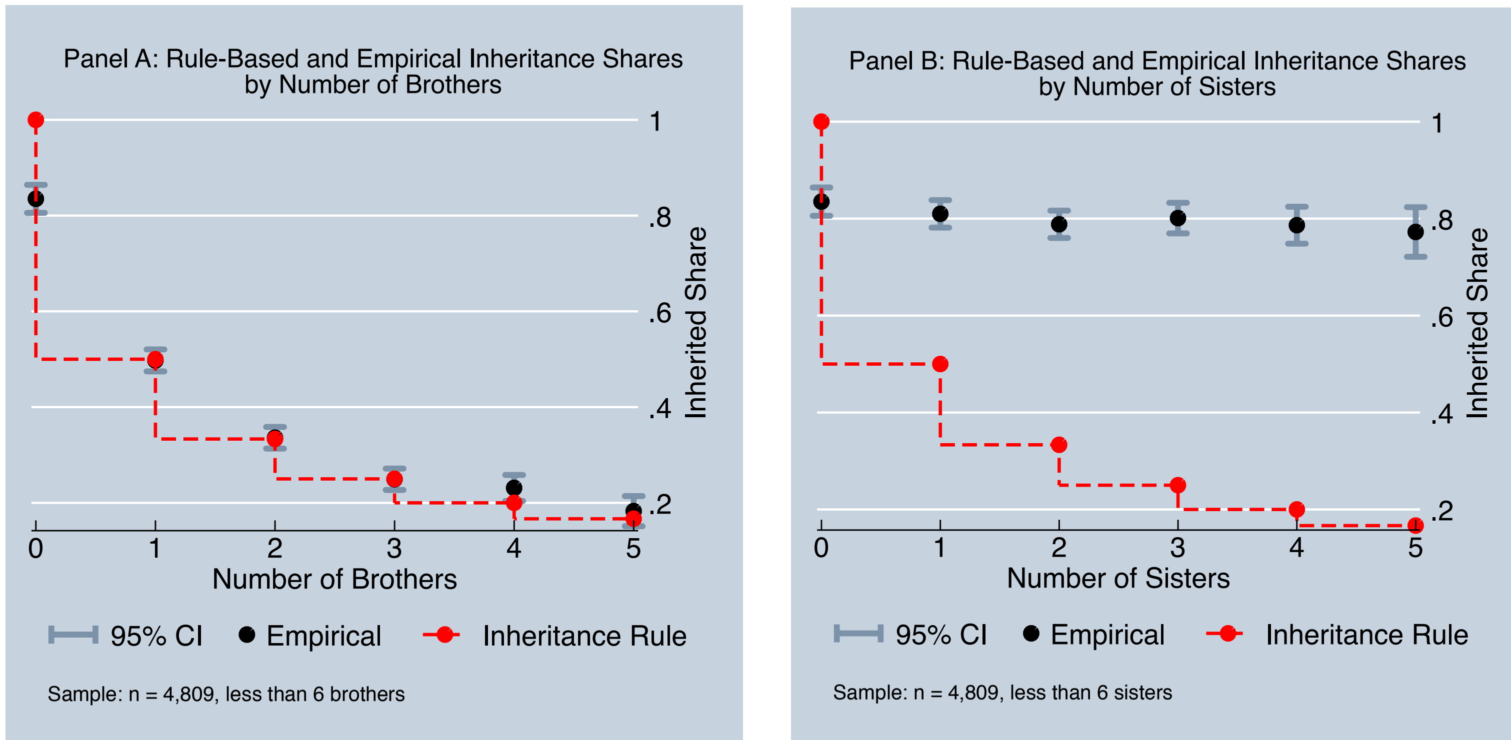
The 2SLS estimate computes a weighted average of the Wald estimator across individuals with varying numbers of siblings and family landholdings, weighted by the strength of the covariate specific first stage. Conceptualizing this in the LATE framework (Angrist and Imbens, 1995), the 2SLS estimate gives us the local average treatment effect for the compliers: the individuals who would have stood to inherit a larger share of their parent's land had one or more of their brothers been sisters. As such, the first stage (1.6) and second stage equations (1.7) are:

$$L_{ij} = \alpha_j + \pi \ Pred\ Share_{ij} + \gamma_k \sum_{k=1}^{K} I(sibs_{ij} = k) + \gamma_l \sum_{l=1}^{L} I(fam\ land_{ij} = l) + \epsilon_{ij} \tag{1.6}$$

reported that brothers inherit the same quality of land. Column (1) in Appendix B2 shows, using within family estimates, that birth order does not predict the probability of inheriting land.

[24]'Predicted share' is the preferred specification for the instrument as the functional form has a clear parallel to the visual in Panel A of Figure 1.1. Relative to instrumenting with the predicted *level* of land, the share has the advantage of being uncorrelated with controls for family land, being far less sensitive to outliers and having a larger first stage F-statistic as discussed in the following section. While using the natural logarithm of the level of land attenuates some of these concerns, estimates giving the marginal effect of an acre of inherited land are preferred to elasticities. In addition, use of the 'predicted share' instrument permits the comparison of marginal effects in placebo tests estimating the independent effects of sibling sex composition in the absence of it influencing inherited land (e.g. respondents from landless families and those residing in urban areas).

FIGURE 1.1: RULE-BASED AND EMPIRICAL INHERITANCE SHARES OF FAMILY LAND



Notes:

These figures plot the inheritance shares predicted by the instrument and the empirical inheritance shares observed in the data. Panel A shows empirical inheritance shares by the number of brothers a respondent has, while Panel B shows it by the number of sisters. The 'Inheritance Rule' in each panel is the value of the instrument, i.e. 1/(1+Brothers). The Empirical share is calculated as the land inherited by the head divided by the land owned by his family. For Panel A this value is regressed on a set of dummies for the number of brothers (sisters for Panel B) and individual has, controlling for the number of sisters (brothers for Panel B). Household head's with 5 brothers or less (domain of Panel A) account for 98.49% of the sample. Household head's with 5 sisters or less (domain of Panel B) account for 98.05% of the sample. Data Source: ARIS-REDS Dataset

$$Y_{ij} = \alpha_j + \rho \hat{L}_{ij} + \gamma_k \sum_{k=1}^{K} I(sibs_{ij} = k) + \gamma_l \sum_{l=1}^{L} I(fam\ land_{ij} = l) + \eta_{ij} \tag{1.7}$$

Where α_j is a district fixed effect, $Pred\ Share_{ij}$ is the instrument as described above, $\sum_{k=1}^{K} I(sibs_{ij} = k)$ is a set of dummy variables for the number of siblings k in $i's$ family, $\sum_{l=1}^{L} I(fam\ land_{ij} = l)$ is a set of dummy variables for parent's landholdings, Y_{ij} is the outcome variable of interest, and $\hat{L}_{ij}$ are the first stage fitted values.

The causal effect of inherited land may also vary with underlying heterogeneity. For example, consider the case of whether a specific factor market friction is more or less binding (i.e. variation in θ from the conceptual framework). To test these hypotheses, I incorporate interaction effects into the structural equation. In this case for individual i, in village j in district k the structural equation is:

$$Y_{ijk} = \alpha_k + \rho_1 L_{ijk} + \rho_2 Q_{jk} + \rho_3 (L_{ijk} * Q_{jk}) + \gamma X_{ijk} + \nu_{ijk} \tag{1.8}$$

Where Y_{ijk}, L_{ijk} and X_{ijk} are as above, α_k is a district fixed effect and Q_{jk} is a dummy variable set to 1 if a factor market friction exists at the village level and 0 if not. ρ_1 is the causal effect of inherited land on holding a non-agricultural occupation, and ρ_3 tests whether the estimated effect varies across villages in which the factor market friction is present and those where it is not.

As Q_{jk} is an approximation to θ discussed in the conceptual framework, if we consider Y_{ijk} as household consumption instead, ρ_1 is now an estimate of β_w, the comparison between the 'cash' group and the 'control' group, while the sum of ρ_1 and ρ_3 is an estimate of β_{w-c}, the comparison between the 'land group' and the 'control' group. Finally, ρ_3 is an estimate of β_c; the opportunity cost of inheriting a marginal acre of land averaged over the intensive margin for land. Given two endogenous variables, I require at least two instruments: $Pred\ Share_{ijk}$ and $Pred\ Share_{ijk} * Q_{jk}$. This results in two first stages to instrument for both

L_{ijk} and $L_{ijk} * Q_{jk}$ that share a common RHS as below:

$$\begin{aligned}(L_{ijk}\ ,\ L_{ijk} * Q_{jk}) \quad &= \quad \alpha_k + \pi_1 Pred\ Share_{ijk} + \pi_2(Pred\ Share_{ijk} * Q_{ijk}) + \pi_3 Q_{jk} \\ &+ \gamma_z \sum_{z=1}^{Z} I(sibs_{ijk} = z) + \gamma_l \sum_{l=1}^{L} I(fam\ land_{ijk} = l) + \epsilon_{ijk} \qquad (1.9)\end{aligned}$$

1.6 Results

In the following sections I use the terms 'household head' and 'respondent' interchangeably. Additionally, I use the term 'predicted share' to refer to the instrument.

1.6.1 Summary Statistics

The first two columns of Table 1.1 report summary statistics for household head-level data, the primary unit of analysis. The remaining columns report summary statistics for sibling-level and child-level data, for which fewer outcomes are available. The mean age of a household head is 49 years, they have 5.8 years of education and spend roughly $80 per month on household expenses.[25] Nearly 30% of the population describe their primary activity status as being in a non-agricultural occupation.[26] Of these respondents 32% hold a salaried job, 22% are engaged in non-agricultural wage labor and 18% report operating a non-farm business.[27] Mean inherited landholdings are 4.12 acres while the median is 2 acres, suggesting a distribution of landholdings with a long right tail as in Panel A of Figure 1.2.

On average respondents became the head of their household at age 33 (median 32) or 16 years ago (median 14) - a proxy for their age of inheritance. In spite of this, Panel B shows that 70% of respondents have experienced no change in the current landholdings over their

[25] Appendix A1 provides details of the variables discussed in the analysis.

[26] Appendix C6 tests whether the main results are robust to alternative definitions of this variable. In particular, alternatively defining this variable as the occupation from which a respondent gets the majority of their income yields similar results.

[27] See Appendix A3 for details of salaried positions and non-farm businesses. The REDS data does not contain details on non-agricultural wage work, however for roughly 3000 respondents reporting non-agricultural wage labor in the IHDS, 36% are involved in construction work.

TABLE 1.1: SUMMARY STATISTICS FOR HEAD-LEVEL, SIBLING-LEVEL AND CHILD-LEVEL DATA

	Head-Level		Sibling-Level		Child-Level	
Dependent Variable	Mean (1)	Standard Deviation (2)	Mean (3)	Standard Deviation (4)	Mean (5)	Standard Deviation (6)
Age (Years)	49.263	14.149	48.376	35.705	23.248	11.216
Sex	1.000	0.000	1.000	0.000	0.549	0.498
Education (Years)	5.847	5.253	4.623	4.386	5.481	4.709
Inherited Land	0.791	0.406	0.674	0.469	0.033	0.179
Currently Owns Land	0.926	0.262	0.766	0.423	0.322	0.467
Household Size	6.257	3.447	5.085	2.887	4.015	4.360
Rural to Urban Migrant	-	-	0.011	0.103	0.018	0.132
Non Agricultural Occupation	0.298	0.458	-	-	-	-
Yearly HH Consumption (Rs.)	42262.01	37895.06	-	-	-	-
Years since Headship Assumed	15.961	13.776	-	-	-	-
Father of Head Alive	0.187	0.390	-	-	-	-
Loan (Last 5 years)	0.196	0.397	-	-	-	-
No. of Brothers	1.925	1.432	-	-	-	-
No. of Siblings	3.773	2.183	-	-	-	-
Predicted Share (0-1)	0.452	0.265	-	-	-	-
First Born	0.377	0.485	-	-	-	-
Inherited Land (Acres)	4.120	6.734	-	-	-	-
Current Land (Acres)	5.136	8.090	-	-	-	-
N	4809	-	14773	-	16130	-

Notes:

This table presents summary statistics for the data used from the 1999 Wave of the ARIS-REDS survey. Means are reported in columns 1, 3, and 5, while standard deviations are reported in columns 2,4 and 6. Columns 1 and 2 contain summary statistics for all male Hindu household heads whose parents owned land. Column 3 and 4 contain summary statistics for sibling level data. The data examined here corresponds to all male siblings of household heads (including the head) used in the main analysis that reached the age of 10 prior to their death. Columns 5 and 6 contain summary statistics for child level data. This includes all children of household heads in the main analysis who are at least 6 years old (i.e. of schooling age). Data Source: ARIS-REDS Dataset.

inherited landholdings and 84% of respondents have experienced increases or decreases of less than 2 acres over this period. For the median respondent, all current landholdings are inherited while the average share of inherited land in total landholdings is 83%. Nearly 19% of the sample still have a living father. Among this subset 20% of respondents have inherited land, while for those with no living father 93% of respondents have inherited land.[28] The average household has 6 persons residing in it and nearly one-fifth of the sample has taken out a loan in the last five years. Respondents have 4 siblings on average, 52.3% of whom are male.

The means and standard deviations for sibling-level data are reported in columns (3) and (4). The average age of siblings is comparable to that of household heads, but a slightly lower fraction (67%) inherit land. Rural to urban migration is very low at just 1.1%. Column (5) shows that the children of household heads are aged 23 and have 5.4 years of education on average. A very small fraction (3%) report inheriting land, as a consequence of their parents still being alive, and the rate of urban migration is comparable to that of the previous generation.

1.6.2 First Stage and Reduced Form Estimates

First Stage

As discussed in Section 5, Panel A of Figure 1.1 reveals that the data on empirical inheritance shares closely approximate what is predicted by the inheritance rule.[29] A visualization of the first stage using the 'predicted share' instrument is presented in Figure 1.3. This figure plots the coefficients from a regression of inherited land on a set of dummies for each value of the instrument, with fixed effects for the number of siblings, districts and family

[28] 60% of respondents became the head of their household after their father died. While the survey does not give the exact time at which they inherited land, consistent with the summary statistics land is typically inherited upon a father's death. As is evident, however, household headship does not necessitate the inheritance of land or a father's death. Foster and Rosenzweig (2002) using the the REDS panel find that household division (i.e. headship) is predicted by age of the respondent, household size and a co-resident wife but not by family landholdings.

[29] The empirical inheritance shares are calculated as self-reported inherited land divided by total land owned by parents, both of which are directly reported in the survey.

landholdings.[30]

A clear pattern emerges where inherited landholdings rise on average as the predicted share increases. Column (2) in Table 1.2 confirms this graphical intuition and assesses five other parameterizations of the first stage. In each case, the instruments are highly correlated with inherited landholdings and yield first stage F-statistics ranging from 22 to 142. [31]

Column (1) reports the coefficient on the linear specification of the instrument which yields a first stage F-statistic of 128.[32] Column (2) reports the coefficient on 'predicted share' = $\left[\frac{1}{1+Brothers}\right]$, and is precisely estimated with a F-statistic of 126. Column (3) in turn reports the coefficient on 'predicted land' = $\left[\frac{Family\ Land}{1+Brothers}\right]$, which has a weaker first stage (F-statistic = 101), likely due to the fact that family landholdings are themselves highly correlated with 'predicted land' leading to a lower partial F-statistic.[33] Column (4) reports the coefficient on the natural logarithm of 'predicted land' which yields a F-statistic of 142. Finally, in column (5) the coefficients on a set of dummies for the number of brothers are reported, these instruments have a F-statistic of 22.

The coefficients reveal that the marginal effect of a brother on inherited landholdings is decreasing as the equal division rule would suggest. Substantively, an additional brother leads to a reduction in inherited landholdings of 1.24 acres on average or one-third of median landholdings in rural India. Therefore sibling sex composition induces substantial variation in inherited landholdings.

[30]The coefficients that are plotted the those on a set of dummies equal to 1 if the share is equal to 1 (no brothers), 0.5 (2 brothers) and so on. Individuals with more than 4 brothers are omitted for graphical clarity. They constitute less than 5% of the sample.

[31]As discussed in the following section, Appendix C1 shows that these alternative instrument specifications yield similar 2SLS estimates.

[32]All first stage specifications include district, sibling and family land fixed effects.

[33]Additionally, using an instrument specification that uses the level of inherited land will result in families with larger landholdings receiving disproportionate weight in the overall LATE as a marginal brother will induce more variation in the instrument for such families relative to those with smaller landholdings. In combination with a highly-nonlinear reduced form, this can yield estimates that do not reflect the the actual distribution of land that has a long right tail as in Panel A of Figure 1.2 but are rather driven by functional form assumptions.

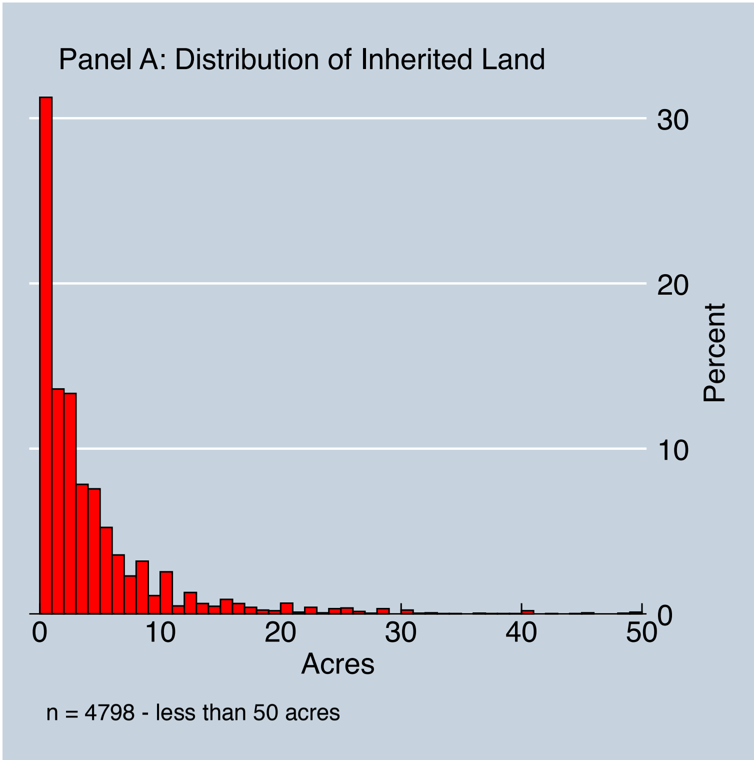
TABLE 1.2 : THE FIRST STAGE AND REDUCED FORM ESTIMATES

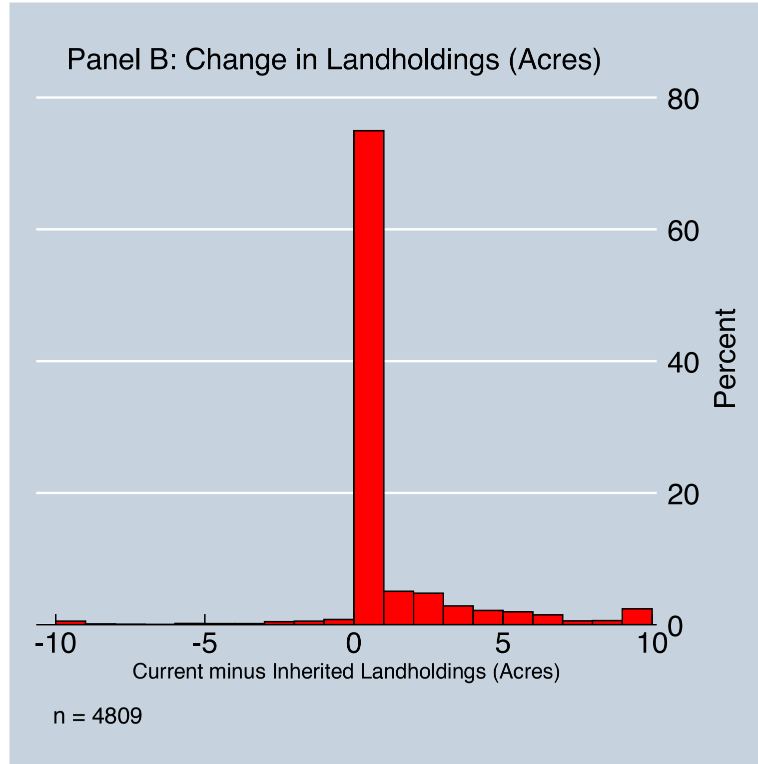
Instrument Specification	No. of Brothers (Linear) (1)	Pred Share (0-1) (2)	Pred Land (Acres) (3)	Log(Pred Land) (Acres) (4)	Brother Dummies (5)
Panel A: First stage for Inherited Land (Acres)					
Instrument	-1.248***	7.174***	0.542***	1.757***	-
	(0.110)	(0.639)	(0.054)	(0.148)	
1 Brother	-	-	-	-	-2.704***
					(0.390)
2 Brothers	-	-	-	-	-4.425***
					(0.444)
3 Brothers	-	-	-	-	-5.313***
					(0.488)
4 Brothers	-	-	-	-	-6.481***
					(0.559)
5 Brothers	-	-	-	-	-6.689***
					(0.757)
Depvar mean	4.120	4.120	4.120	4.120	4.120
First Stage F-Statistic	128.137	125.952	101.458	141.607	22.433
N	4809	4809	4809	4809	4809
Panel B: Reduced form for Non-Agricultural Occupation					-
Instrument	0.025***	-0.126***	-0.001	-0.073***	-
	(0.006)	(0.035)	(0.001)	(0.011)	
Depvar Mean	0.298	0.298	0.298	0.298	
N	4809	4809	4809	4809	-
Panel C: Reduced form for Log(Household Consumption)					
Instrument	-0.041***	0.195***	0.012***	0.091***	-
	(0.008)	(0.043)	(0.002)	(0.013)	
Depvar Mean	10.442	10.442	10.442	10.442	-
N	4809	4809	4809	4809	-
No. of Siblings FE	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y

Notes:

The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Panel A reports the coefficient on the instrument(s) for the first stage. Panel B reports the coefficient on the instrument from reduced form regression for non-agricultural occupation, while Panel C reports the reduced form for the log of household consumption. In column 1 the instrument is specified as the (linear) number of brothers, in column 2 it is 'Predicted Share' = 1/(1+Brothers), in column 3 it is 'Predicted Land' = Family Land/(1+Brothers), in column 4 it is Log(Predicted Land), and in column 5 it is the a set of dummies for the number of brothers (8 dummies in total, I report the coefficients for up to 5 brothers which account for 98.11% of sample). The dependent variable Panel B is Non-Ag occupation and is defined as the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable in Panel B is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. Brothers are defined as male siblings who grew up to at least the age of 10. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01Data Source: ARIS-REDS Dataset.

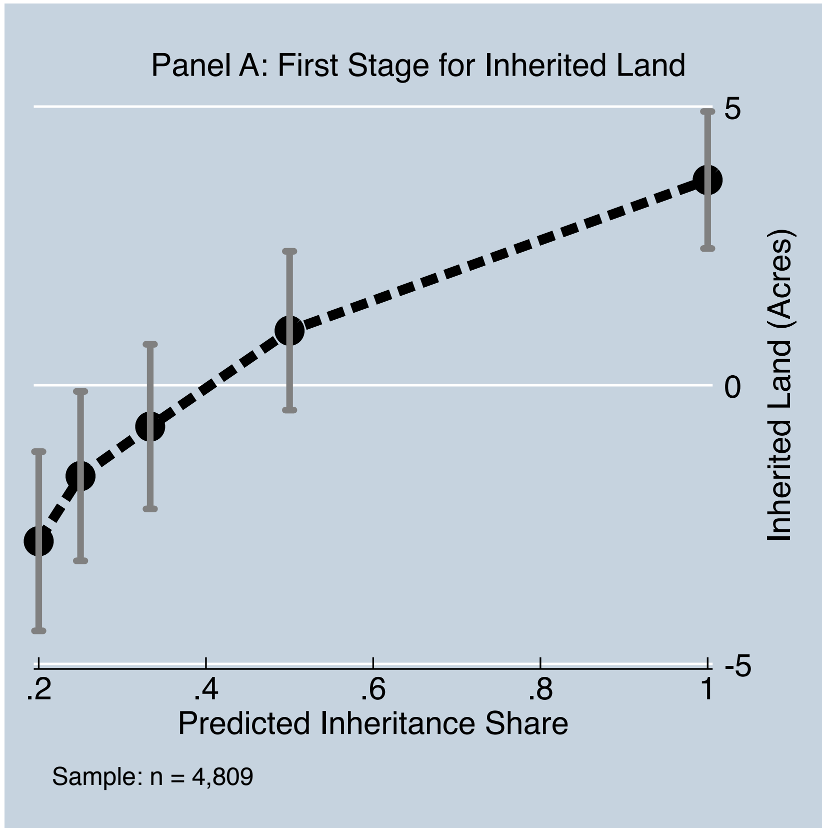
FIGURE 1.2 : DISTRIBUTION OF INHERITED LAND AND CHANGES IN LANDHOLDINGS





Notes:
Panel A plots the distribution of inherited landholdings. Panel B plots the difference between current and inherited landholdings. In the full sample (n = 4,809) 20.8% of households report that they inherited no land, 17.2% report inherited landholdings of 0-1 acres, 15.2% report inherited landholdings between 1-2 acres, 10.0% report between 2-3 acres, 8.3% report between 3-4 acres, and the remainder (> 4 acres) account for 28.2% of the sample. The mean is 4.12 acres and the median is 2 acres. Panel A limits the domain to those who inherit less than or equal to 50 acres of land. Panel B codes increases of greater than 10 acres as 10, and decreases of greater than 10 acres and -10. 70% of respondents report no change in inherited landholdings, while 84.4% report changes of less than 2 acres. The 95th percentile is a change of +6 acres while the 5th percentile is 0. Data Source: ARIS-REDS Dataset.

FIGURE 1.3: VISUALIZATION OF FIRST STAGE



Notes:
This figure plots the first stage for inherited landholdings. For visual clarity, the figure limits the domain to those with less than or equal to 4 brothers (94.7% of sample), i.e. predicted shares between 0.2 and 1. The graphs plots the coefficients (black dots) from a regression of the dependent variable -- land inherited by the respondent measured in acreas -- on a set of dummy variables for each value of the inheritance share, omitting the constant. This regression includes district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The 95% confidence interval is calculated using robust standard errors and is plotted with the gray bars. Data Source: ARIS-REDS Dataset.

Reduced Form

Across four different instrument specifications, Panel A of Table 1.2 reports the reduced form coefficients for whether the head is primarily engaged in a non-agricultural occupation, while Panel B reports the reduced form coefficients for household consumption.[34] Column (1) reports the coefficients on the linear specification of the instrument, column (2) reports the coefficient on 'predicted share', column (3) reports the coefficient on 'predicted land' and column (4) reports the coefficient on the natural logarithm of 'predicted land'.

Across instrument specifications, the coefficients reveal that having more male siblings (or as a consequence: a smaller predicted share or level of inheritance) has a positive effect on the probability of leaving agriculture and a negative effect on consumption. An exception to this uniformity is the reduced form coefficient for non-agricultural occupation using the 'predicted land' instrument. While this is of concern, using the *level* of 'predicted land' is sensitive to outliers: both winsorizing 2.5% of the variable or using the natural logarithm of 'predicted land' as in column (4) yield highly significant reduced form estimates. In addition, the 'predicted land' instrument is highly correlated with the family land control, the omission of which also results in a significant reduced form effect.

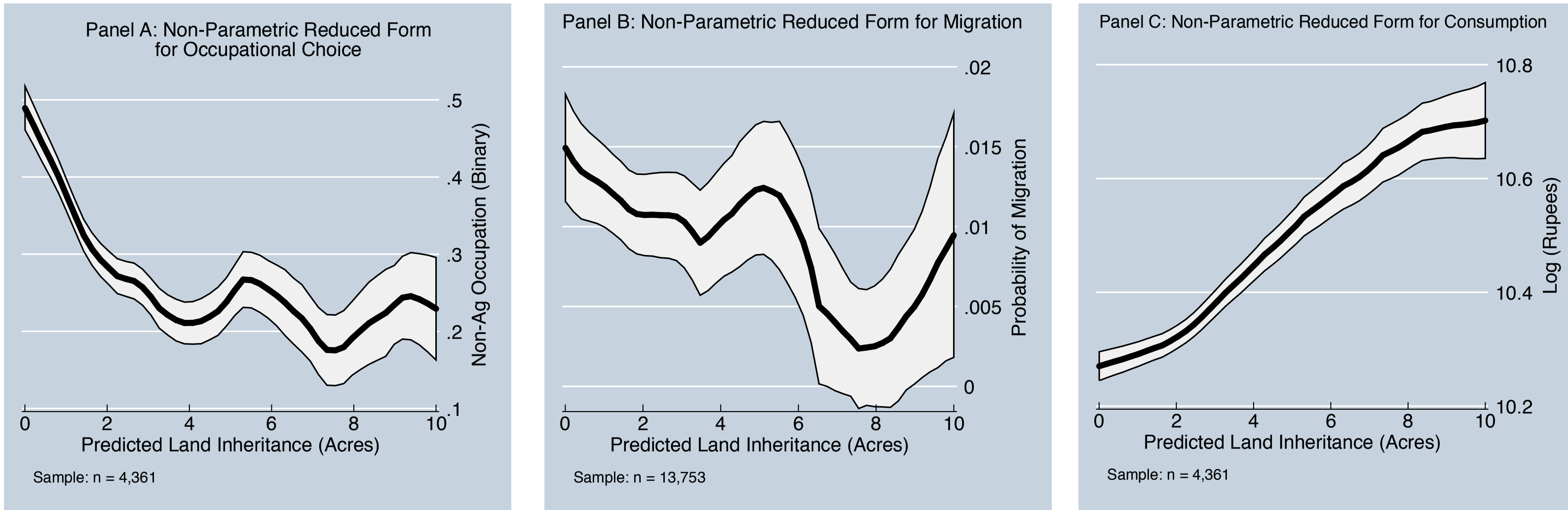
Figure 1.4 plots the smoothed values and the 95% confidence bands from kernel-weighted local polynomial regressions of the reduced form for non-agricultural occupation (Panel A), rural-urban migration (Panel B) and household consumption (Panel C) using 'predicted land' instrument without family land controls. These figures show evidence of non-linearities in the estimated effects that are discussed in Section 7.3 in more detail.

1.6.3 Occupational Choice, Migration and Consumption

In the following sections I estimate (1.5) using two-stage least squares (2SLS), where the first stage specification is equation (1.6). Note, both the first and second stage equations

[34] All reduced form specifications include district, sibling and family land fixed effects.

FIGURE 1.4: NON-PARAMETRIC VISUALIZATIONS OF REDUCED FORM EQUATIONS FOR OCCUPATIONAL CHOICE, CONSUMPTION AND MIGRATION



Notes:
These figures plot the reduced form for occupational choice (Panel A), urban migration (Panel B) and household consumption (Panel C). All figures plot smoothed values from kernel-weighted local polynomial regressions of the dependent variable in question on the predicted land instrument, which is equal to Family Land/(1+Brothers). All figures use the epanechnikov kernel with the STATA calculated rule of thumb (ROT) bandwidth. The white bands show the 95% confidence interval, which are calculated using the default normalized weighted residual sum of squares from a local polynomial fit of a higher order using a pilot bandwidth of 1.5 * ROT. All figures limit the domain to 10 acres which accounts for 90% of the sample. Data Source: ARIS-REDS Dataset.

include sibling, family landholdings and district fixed effects.[35]

Occupational Choice

In the presence of capital market imperfections, theories of occupational choice (Banerjee and Newman, 1993) predict that inherited land may act as collateral in accessing credit. It follows that the landed will be well positioned to take advantage of higher return investments in the non-agricultural sector, paving their way out of the subsistence or agricultural sector.

Table 1.3 reports the OLS (column 1) and 2SLS (column 2) estimates for the effect of inherited land on holding a non-agricultural occupation.[36] The OLS estimate for the effect of land on the probability of transitioning out of agriculture is relatively small, at just -0.4%. The 2SLS estimate, however, reveals a much larger effect having in part addressed omitted variable bias. The causal effect of inheriting an additional acre of land, contrary to the predictions of standard models of occupational choice, *reduces* the probability of transitioning out of agriculture by -1.8% per acre on average.[37] When restricting the sample to individuals whose families are in the first quartile of the family land distribution, this effect increases dramatically to -21% per acre. The non-linearity in this effect suggested by this heterogeneity is apparent from non-parametric visualizations of the reduced form relationship.

---

[35] All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies).

[36] 'Non-Agricultural Occupation' is coded as 1 if the primary status reported by the respondent in the REDS survey is not self-cultivation or agricultural labor. Appendix C6 shows that these results are robust to using an alternative definition that directly computes primary occupation from reported sources of income.

[37] Appendix B5 shows that this effect is driven by those with more land being less likely to own a non-farm business. While the effects on entry in to salaried work and non-agricultural wage work are negative, they are imprecisely estimated.

Panels A in Figure 1.4 plots the smoothed values and the 95% confidence bands from kernel-weighted local polynomial regressions of a dummy for holding a non-agricultural occupation on predicted land.[38] The estimated slope is much steeper across 0-4 acres and levels off thereafter. This suggests that inherited landholdings are a particularly important determinant of occupational choice among those with smaller landholdings. They are also likely to be the individuals with the most to gain from higher returns to their labor in the non-agricultural sector, as their agriculture is characterized by subsistence rather than large profits.[39]

Migration to Urban Areas

Inherited land may also facilitate movement of labor across space, to take advantage of higher returns to labor in urban areas (Bryan *et al.*, 2011;Beegle *et al.*, 2011). While the REDS data only surveys household heads in rural India, it records the movements of their siblings over space and records their inherited landholdings.[40] This permits the estimation of (1.7) using data on all siblings. As before, the variation is across households as within a family all male siblings stand to inherit the same amount of land.

Columns (3) and (4) in Table 1.3 report the OLS and 2SLS estimates for the effect of inherited land on rural to urban migration. The effect of inherited land on the rate of urban migration is negative and significant at -0.02% per acre. For those whose families own less

---

[38]In order to show the reduced form along a continuous support, $PredictedLand = \left[\frac{Family\ Land}{(1+Brothers)}\right]$ is used instead of the main instrument *Predicted Share*. The family land control is dropped in this specification.

[39]This claim is supported by the stylized facts that emerge from the analysis in Appendix A4. Using data from the Indian Human Development Survey, this appendix reports estimates for occupational wage gaps in rural India and how they differ for farmers with varying sizes of landholdings. The analysis is similar to Gollin *et al.* (2014b) in computing wage gaps between the agricultural and non-agricultural sectors, but differs in that it restricts this analysis to rural areas. The equivalent daily wage is higher in agriculture for those with landholdings above 3 acres relative to non-farm business and non-agricultural wage labor. However, for those with landholdings below 3 acres, daily wages in non-farm businesses (37%), salaried work (74%) and non-agricultural wage work (9.2%) are higher on average. These estimates control for human capital, age, sex and district-level unobservables, but there could still be individual-level unobservables that drive these wedges.

[40]Note, these results are not from a household roster, which only gives details for co-resident siblings, but from a complete enumeration of all siblings irrespective of whether they are co-resident. Summary statistics for these 14,773 siblings are reported in columns (3) and (4) of Table 1.1.

TABLE 1.3 : THE EFFECT OF INHERITED LAND ON OCCUPATIONAL CHOICE, MIGRATION AND HOUSEHOLD CONSUMPTION

Dependent Variable	Non-Ag Occupation Binary Variable		Rural-Urban Migration Binary Variable		Household Consumption Log(Rs.)	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
Inherited Land	-0.003**	-0.018***	-0.000	-0.002***	0.025***	0.027***
(Acres)	(0.001)	(0.005)	(0.000)	(0.001)	(0.003)	(0.006)
Mean of Dep. Var.	0.298	0.298	0.011	0.011	10.442	10.442
First Stage F-statistic	-	125.952	-	186.771	-	125.952
N	4809	4809	14773	14773	4809	4809
Panel B: Family Landholdings Below First Quartile (less than 3 acres)						
Inherited Land	-0.030***	-0.210***	-0.000**	-0.034*	0.055***	0.080
(Acres)	(0.011)	(0.069)	(0.000)	(0.021)	(0.014)	(0.064)
Mean of Dep. Var.	0.414	0.414	0.016	0.016	10.244	10.244
First Stage F-statistic	-	22.690	-	6.881	-	22.690
N	1363	1363	3720	3720	1363	1363
No. of Siblings FE	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y
Data	Head-Level		Sibling-Level		Head-Level	

Notes:

This table reports estimates of the long-term effect of inherited land on occupational choice, migration and household consumption. Columns 1, 3 and 5 report OLS coeffiient estimates while columns 2, 4 and 6 report 2SLS estimates. The sample in columns 1,2,5 and 6 is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. The sample in column 3 and 4 are all male siblings of these household heads (including the heads) aged above 10 years. Note, this data is reported for all siblings not just siblings residing in the household at the time of the survey. Panel A includes all households, while Panel B limits the analysis to households whose family had less than 3 acres. The dependent variable in cols 1 and 2 is Non-Ag occupation and is defined as the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable in column 3 and 4 is a dummy variable for whether or not the sibling migrated to an urban area in the same district or outside of it. The dependent variable col 5 and 6 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The instrument specfication used is Predicted Share = 1 / (1+ Brothers). The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$Standard errors are clustered at the family level for sibling-level regressions. Data Source: ARIS-REDS Dataset.

than 3 acres this effect is nearly 20 times as large at -3.4% per are. Panel B in Figure 1.4 plots the smoothed values and the 95% confidence bands from kernel-weighted local polynomial regressions of rural-urban migration on predicted land. Once again, we see that while the estimate negative relationship is steep for amount of land up to 4 acres, beyond this point the estimates are imprecise and qualitatively unclear.

Household Consumption

While the prior estimates suggest that inherited land is an important determinant of leaving agriculture and migration, the implications of these findings for the level of consumption remain unclear. It may be the case that those with more land are more likely to remain in agriculture because it leaves them better off than leaving (i.e. the 'wealth effect', β_w in equation (1.1) dominates the cost β_c in (1.3) of inheriting land). Alternatively, inherited land may reduce spatial and occupational mobility to such an extent that they cause a lower level of consumption (i.e. $\beta_{w-c} <0$). The estimates that follow combine these effects as they average over frictions that exist in factor markets (i.e. θ from section 3).

Column (6) in Table 1.3 reports the causal effect of inherited land on the log of yearly household consumption.[41] An additional acre of land increases household consumption in the long run by 2.7% on average. When restricting the analysis to respondents from families who own less than 3 acres of land, the point estimate is an imprecisely estimated 8% . Panel C in Figure 1.4 shows the reduced form relationship graphically using local polynomial regressions. The estimated effect of predicted land on consumption is flatter from 0-2 acres and begins to increase at a higher rate thereafter.[42]

[41]Household consumption is the preferred metric for empirical studies assessing poverty in India (Deaton and Dreze, 2002). Deaton (1997) suggests that income based measures may be more vulnerable to imputations, recall bias and seasonality. In addition, in this context the measurement of profits is made even more complicated by the valuation of household labor in both farm and non-farm businesses.

[42]It is also possible that the effects may operate through the expectation of inheriting land in addition to the actual inheritance of land. Four-fifths of respondents with a living father are yet to inherit land. While controlling for this does not influence the 2SLS estimates, the estimation of effects on this sub-sample imperfectly captures the effect of expectations of inherited land. Appendix B1 estimates the effects on leaving agriculture (column 1), household consumption (column 2) and years of education (column 3) for individuals whose father is still alive. The evidence suggests that expectations do play a role in influencing occupational choice but not

1.7 Mechanisms

The results in the preceding section suggest that the inheritance of land limits spatial mobility and the likelihood of leaving agriculture but on average increases their level of consumption. This section looks at how these effects vary with underlying frictions (i.e. θ in the conceptual framework) in factor markets. In so doing, the regressions provide estimates of β_c in equation (1.4) and suggest mechanisms through which these effects operate.

1.7.1 Mechanisms: Access to Credit

Initial endowments of land may facilitate a transition out of agriculture through increasing the ability of the landed to borrow and invest in high-return opportunities in the non-agricultural sector. However, if land does not effectively serve as collateral or financial institutions are absent this may not be the case. Column (1) in Table 1.4 shows that the OLS estimate for the effect of land on the probability of having taken out a loan in the last 5 years is 0.2%. Column (2) shows that the 2SLS estimate increases to 1.5% and in each case the estimated coefficients are precisely estimated at the 1% level. The downward bias in the OLS estimate suggests that those who posses land may also have superior access to credit through other channels. Additionally, an additional acre of land increases the value of loans taken out on average by 15.2% (measured in log rupees).[43] These results suggest that inherited land increases the ability of the landed to borrow and its effects on occupational choice, migration and household consumption obtain in spite of this.

human capital acquisition or household consumption.

[43]While occupations may themselves influence the need to borrow, the number of loans does not vary appreciable between those primarily engaged in agriculture and non-agricultural work. Appendix A6.1 shows that a roughly equal amount of loans were taken out for agricultural and non-agricultural investment purposes. Additionally, Appendix A6.2 and A6.3 show that roughly 1/5 loans required collateral and in 83% of these loans land was used as collateral.

TABLE 1.4: THE EFFECT OF INHERITED LAND ON BORROWING (2SLS ESTIMATES)

Dependent Variable	Took out Loan (Last 5 yrs) Binary Variable		Total Value of Loans Log(Rs.)	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
Inherited Land	0.002**	0.015***	0.030***	0.152***
(Acres)	(0.001)	(0.004)	(0.009)	(0.036)
No. of Siblings FE	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Mean of Dep. Var.	0.184	0.184	1.700	1.700
First Stage F-statistic	-	125.952	-	125.952
N	4809	4809	4809	4809

Notes:

This table reports estimates of the long-term effect of inherited land on access to credit. Columns 1 and 3 report OLS coefficient estimates while columns 2 and 4 report 2SLS estimates. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. The dependent variable in col 1-2 is a dummy variable coded as 1 if the head took out a loan in the past 5 years, while in col 3-4 it is the log of the total value of all loans taken out in the last 5 years. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The instrument specfication used here is Predicted Share = 1 / (1+ Brothers). The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01Data Source: ARIS-REDS Dataset.

1.7.2 Mechanisms: Culture as a Friction on Labor Markets

A number of frictions may prevent individuals from renting, selling or leaving behind inherited land, effectively tying inheritors to their land (See Section 2.2). In the preceding 20 years, just 17% of the respondents report buying, selling or mortgaging their land. In addition, leaving land vacant may leave it vulnerable to expropriation, particularly given that just 40% of households in these villages own formal titles to their land. [44]

In this context of limited land markets and formal property rights, I consider a social custom among Hindu families that may make it even more difficult to vacate inherited land. For the eldest son, taking care of his parents in their old age is a 'sacred duty' in Hindu scripture known as *Pithru Rina* (Kumari Bhat and Dhruvarajan, 2001). Upon the retirement of the father, the eldest son is obligated to take over the affairs of the household. In agricultural households this often implies responsibility for the family land and occupational succession, although the land is split equally thereafter regardless of birth order.[45]Jayachandran and Pande (2013) find that parents invest more heavily in the first-born son in India; they are nearly 0.2 standard deviations taller for their age and have almost 2 years more schooling on average relative to second-borns, with the differences growing even larger with parity.[46]

Absent an obligation, this increased human capital endowment should leave household heads that are first-born sons considerably better off. However, by virtue of this custom first-born sons (37% of the sample) may be less able to vacate their land (i.e. θ, the measure of frictions, approaches 1) resulting in any effects on leaving agriculture and household consumption operating more stringently in comparison to latter-born sons (i.e. θ approaches

[44]Goldstein and Udry, 2008 show that the of expropriation risk may influence agricultural investments in land, how such concerns may also distort labor market supply as in (Field, 2007)

[45]As a consequence of this responsibility, the eldest son may be exposed to farming earlier and develop more 'farm-specific' human capital. I interpret any such differences as a consequence of this custom. A similar hypothesis is put forward by Laband and Lentz (1998), who argue that 'farm-specific' human capital may be a reason for the much higher observed rate of occupational succession by sons of farmers in the US.

[46]In contrast, Jensen and Miller (2011) show that parents may also strategically limit the education of children expected to remain at home in order to prevent them from migrating to the city.

0).[47] Table 1.4 estimates equation (1.8) from the empirical section, interacting land with a dummy for whether the respondent is the first-born sibling. As such, the interaction between predicted share and a dummy for first-born son becomes an additional instrument in the first stage as in equation (1.9) in order to recover separate causal estimates for first-born sons and latter-born sons.

Column (1) in Table 1.5 shows that the effect of land on transitioning out of agriculture is -4.4% (significant at the 1% level) for each acre for first-born respondents and virtually zero for latter-born siblings. Turning to the intensive margin, the amount of time spent by first-born heads in agricultural labor on their own farms (column 2) is differentially increasing in the extent of inherited land. The same is true of investments in hired labor (column 3) and improvements in their land including terracing and bunding (column 4). Conversely, latter-borns expend their labor in non-farm enterprises and the amount is increasing in their inherited landholdings (column 5), although this estimate is not significant at traditional levels.[48] Using sibling-level data, column (6) shows that the probability of a first-born sibling migrating to urban areas is also decreasing in their inherited landholdings, but this relationship does not exist for their latter-borns.

Finally, column (7) considers the differential effects on household consumption that this friction imposes on inheritors. First-born respondents have higher consumption on average consistent with parents investing heavily in children obligated to support them. However, we see that inherited land has a positive effect on household consumption on average (3.4% per acre), but for first-born respondents there is a differential negative effect on consumption

[47] First-born in this case refers to the first-born in the family, regardless of sex. If this definition were sex-specific it would not be independent of the instrument, as those with fewer male siblings are more likely to be the first-born son. This implies that some eldest sons are classified as 'latter-borns' because they have older sisters. Columns (3)-(6) in Appendix C6 shows that controlling for latter-born eldest sons leads to small changes in the coefficients but not their qualitative interpretation.

[48] An analysis of sibling-level data also suggests that latter-born siblings may be reallocating land within the family towards their first-borns siblings. Appendix B2 reports the coefficients on the birth order dummies using within-family regression. Latter-borns are less likely to experience an increase in their current landholdings over their inherited landholdings. In the main sample, 26% of household heads report that they experienced an increase in their landholdings over the prior two decades. Of these respondents, nearly 40% report receiving 'gifts' of land, a category distinct to inheriting, leasing or purchasing land.The majority of these contracts are oral rather than written, they do not involve a fee and have no specified term.

TABLE 1.5: HETEROGENEOUS EFFECTS OF INHERITED LAND BY BIRTH ORDER (2SLS ESTIMATES)

Dependent Variable	Non-Ag Occupation Binary (1)	Ag Labor Total Man days (2)	Hired Ag Labor Log(Rs.) (3)	Land Improvement Log(Rs.) (4)	Nonfarm Labor Total Man days (5)	Rural-Urban Migration Binary (6)	Household Consumption Log(Rs.) (7)
Land	-0.002	1.231***	0.174***	-0.014	0.430	-0.001	0.034***
	(0.006)	(0.402)	(0.047)	(0.023)	(0.989)	(0.001)	(0.007)
First Born	0.180***	-6.710*	-0.926**	-0.454**	15.377	0.009	0.172***
	(0.057)	(3.670)	(0.449)	(0.228)	(9.352)	(0.007)	(0.063)
Land*First Born	-0.044***	1.700*	0.210*	0.120**	-4.530**	-0.003*	-0.029**
	(0.014)	(0.886)	(0.109)	(0.055)	(2.182)	(0.002)	(0.015)
No. of Siblings FE	Y	Y	Y	Y	Y	Y	Y
Age FE	Y	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y
Depvar Mean	0.298	28.981	4.758	0.925	23.145	0.011	10.442
First Stage F-Stat	27.390	27.390	27.390	27.390	27.390	42.46	27.390
N	4809	4809	4809	4809	4809	14773	4809
Data	Head-Level	Head-Level	Head-Level	Head-Level	Head-Level	Sibling-Level	Head-Level

Notes:
This table tests whether the effect of inherited land on occupational choice, children's migration and household consumption vary by birth order and the amount of land inherited. All coefficients reported are 2SLS estimates. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. The sample in column 3 and 4 are all male siblings of these household heads (including the heads) aged above 10 years. Note, this data is reported for all siblings not just siblings residing in the household at the time of the survey. Each column reports the 2SLS coefficients on inherited land, first born - a dummy coded as 1 if the respondent was the first born child in his family - and their interaction. The two endogenous variables are instrumented with two instruments : Predicted Share =(1/1+Brothers) and the interaction between Predicted Share and First Born. Non-Ag occupation (col 1) is defined as the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. 'Ag labor' (col 2) are the total number of days of agricultural labor performed by the head in the prior season. 'Hired Ag Labor' (col 3) is the log of the value of hired agricultural labor. Land improvement (col 4) is the log of the total expenditure on improvements in land (e.g. terracing, bunding, fencing, leveling, reclamation etc…) undertaken in the last 10 years. Non farm labor' (col 5) is the total number of days of labor in a non-farm enterprise in the prior season. The dependent variable in column 6 is dummy variable for whether or not the sibling migrated to an urban area within the district or outside it. The dependent variable col 7 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies), age of the household head (20-100, 5 year intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children, aged 20-25 from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The specification involving child-level data (column 6) includes a dummy for the sex of the child. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors are clustered at the family for child-level regressions. Data Source: ARIS-REDS Dataset.

(-2.9% per acre) that results in the net effect of land on consumption to be indistinguishable from zero.[49] This differential effect is also an estimate of β_c, the opportunity cost of land. Taking these estimates literally suggests that for an inheritance of 3 acres, these frictions imply a loss of consumption of 8.7%.

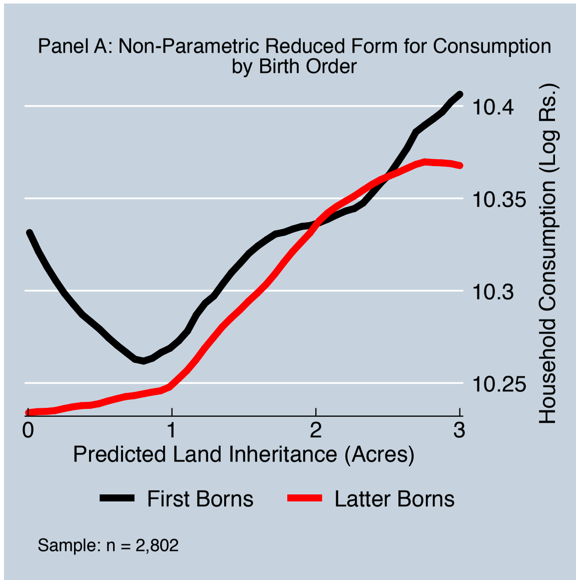
This estimate of β_c also places a bound on how large the non-pecuniary gain from observing the custom – e.g. in terms of status – would need to be in order for *welfare* effect of land to be the same as for latter-borns. Figure 1.5 plots kernel-weighted local polynomial regressions of household consumption on predicted land by birth order. These figures show a dip in consumption for first-borns for inheritances of one acre but no corresponding dip for latter-borns, suggesting that β_c is particularly high for marginal inheritances and reduces thereafter, which is also supported by the non-linearity of the reduced form for household consumption.

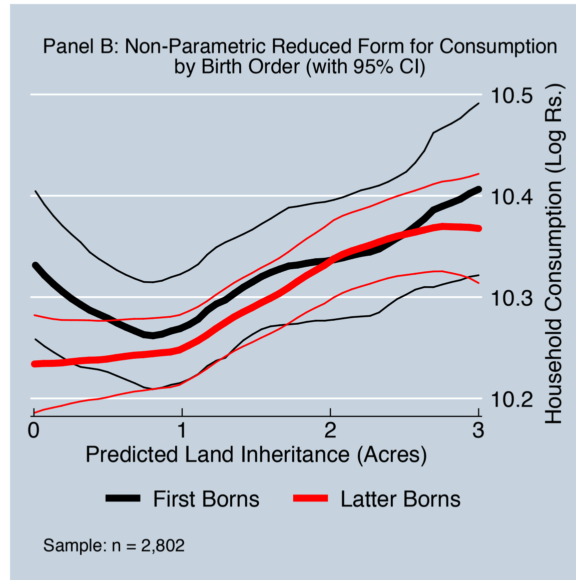
Timing of Headship

Appendix B3 suggests that the timing of the headship (a proxy for the age of inheritance) also matters. First-borns who become heads at a younger age are even more likely to remain in agriculture (column 3) and land has a *net-negative* effect on their level of consumption. In contrast, for latter-borns the age of headship does not appear to matter for either occupational choice or consumption (column 5 & 6). These estimates suggest that for a subset of respondents $\beta_{w-c} < 0$, delivering the perverse result that those with less land are better off in terms of consumption. This may be a consequence of inheriting land earlier in life being especially important in terms of influencing occupational trajectories and precluding profitable opportunities through, for example, migrating to urban areas.

[49] It is important to note here that omitted variable bias is likely to run counter to this result, as research (Jayachandran and Pande, 2013) shows that parents invest more resources (nutrition, education) in first-borns.

FIGURE 1.5: NON-PARAMETRIC VISUALIZATIONS OF REDUCED FORM FOR CONSUMPTION BY BIRTH ORDER





Notes:

These figures plots the reduced form for household consumption. All figures plot smoothed values from kernel-weighted local polynomial regressions of the dependent variable in question on the predicted land instrument, which is equal to Family Land/(1+Brothers). All figures use the epanechnikov kernel with the STATA calculated rule of thumb (ROT) bandwidth. Panel B includes the thinner bands which are the 95% confidence interval. These are calculated using the default normalized weighted residual sum of squares from a local polynomial fit of a higher order using a pilot bandwidth of 1.5 * ROT. Both figures limit the domain to 3 acres which accounts for 58.3% of the sample. Data Source: ARIS-REDS Dataset.

Persistence of Culture

Changing attitudes to custom may result in the attenuation of effects estimated in the previous section. However, the enumeration of details of all children of respondents allows us to look at whether these effects persist over generations.[50]. Specifically, if culture persists over generations first-born children will have a similar obligation to take care of their parents.

Using the same 16,130 child-level observations in the REDS survey, I estimate equations (1.5) and (1.8) to identify the causal effect of land inherited by the head on the outcomes of his children and how these effects vary by the birth order of the *children*. In this case, standard errors are clustered at the family level. In column (1) of Table 1.6 we see that the causal effect of parent's landholdings on the average education of their children is positive. However, in column (2) we see that while first-born children are more educated than their latter-born siblings on average, their relative advantage decreases in the size of their parent's inherited landholdings, although this effect is not significant at traditional levels (t-statistic = 1.38).

A similar story emerges for migration in columns (3) and (4). Migration is on average facilitated by greater landholdings, but there is a differential effect for first-born children. These results suggest culture persists as a mechanism through which the inheritance of land restricts mobility: first-born children become shackled to the land, restricting their movement over space.

1.7.3 Mechanisms: Transaction Costs in the Market for Land

Frictions in land markets may result in inherited land having an even larger influence on leaving agriculture and household consumption relative to where these frictions are less salient. In Table 1.7, I estimate whether the long-term effects of inherited land vary

[50]Note, these results are not from a household roster, which only give details for co-resident children, but from a complete enumeration of all children irrespective of whether they are co-resident. Summary statistics for these 16,310 children are reported in Column (3) of Appendix Table A1

TABLE 1.6: THE EFFECT OF PARENT'S INHERITED LAND ON CHILD OUTCOMES BY BIRTHODER (2SLS ESTIMATES)

Dependent Variable	Education (Years)		Rural-Urban Migration (Binary)	
	(1)	(2)	(3)	(4)
Inherited Land	0.073**	0.091**	-0.001	-0.000
(Acres)	(0.036)	(0.038)	(0.002)	(0.001)
First Born	-	0.451*	-	0.013
		(0.245)		(0.008)
Land*First Born	-	-0.075	-	-0.004**
		(0.054)		(0.002)
Age FE	Y	Y	Y	Y
Sex FE	Y	Y	Y	Y
No. of Siblings FE	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Mean of Dep. Var.	5.481	5.481	0.018	0.018
First Stage F-statistic	157.903	37.047	157.903	37.047
N	16130	16130	16130	16130

Notes:

This table reports estimates of the long-term effect of inherited land on the education and migration of children of household heads by child birth order. The sample is restricted to all children of Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the child- level. The dependent variable in cols 1-2 is the years of education of the child. The dependent variable in cols 3-4 is a dummy variable for whether or not the child migrated to an urban area within the district or outside it. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The instrument specfication used here is Predicted Share = 1 / (1+ Brothers). The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are clustered at the family-level and are shown in parentheses, asterisks denote significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Data Source: ARIS-REDS Dataset.

with transaction costs in the market for land. The reported coefficients are 2SLS estimates (equation (1.8)), using predicted share and its interaction with a measure of transaction costs as instruments for inherited land and it's interaction with transaction costs.

To measure transactions costs, columns (1) and (2) use a z-score index that combines 7 village-level measures of costs in the market for land. This index includes fees for registering landholdings, the cost of a Record of Rights certificate (RoR) and the travel time taken to get to the registrar's office.[51] Column (1) shows that higher transactions costs serve to exacerbate the effect of land on restricting occupational choice although this effect is imprecisely estimated. Column (2) shows that higher transactions costs reduce the beneficial effects of inherited land on household consumption. In columns (3) and (4) the measure of transaction costs are whether there are additional fees that need to be paid in order to register one's landholdings in the village. Out of 246 villages, 60% have additional registration fees which on average amount to Rs. 1,019.83 ($20 in 1999). The presence of additional registration fees similarly serves to tie individuals that inherit land even more to agriculture and in so doing reduce the consumption benefits from inheriting land.[52]

1.8 Robustness Tests

1.8.1 Selection Concerns from Urban Migration

A potential selection concern arises from the fact that all surveyed household heads in the REDS dataset reside in rural areas. I observe the location of *all* siblings of household heads – irrespective of whether they reside in rural or urban areas – and their inherited landholdings which allows me to estimate the negative effect of inherited land on urban migration. However it may still be the case that *entire* families inherited large amounts of

[51] A Record of Rights (RoR) certificate shows proof of ownership and can be used to obtain access to credit.

[52] The nature of the crop-specific production function may also influence θ - the measure of frictions in factor markets. Appendix B3 shows that household heads inheriting land *and* cultivating paddy – a labor intensive crop – relative to wheat, are even less likely to transition out of agriculture. However, this analysis is complicated by endogenous crop choice and the fact that rice, on the whole, a more profitable crop.

TABLE 1.7: HETEROGENEOUS EFFECTS OF INHERITED LAND BY TRANSACTION COSTS IN THE MARKET FOR LAND (2SLS ESTIMATES)

Dependent Variable	Non-Ag Occupation Binary (1)	HH Consumption Log(Rs.) (2)	Non-Ag Occupation Binary (3)	HH Consumption Log(Rs.) (4)
Land	-0.024**	0.032***	-0.011	0.030***
	(0.010)	(0.003)	(0.010)	(0.010)
Transaction Cost	0.005	0.019	0.080*	-0.109**
	(0.048)	(0.024)	(0.048)	(0.053)
Land*Cost	-0.003	-0.008**	-0.024**	0.001
	(0.011)	(0.003)	(0.011)	(0.012)
No. of Siblings FE	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Depvar Mean	0.298	10.442	0.298	10.442
Transaction Costs Measure	Transaction Costs Index (z-score)		Additional Registration Fees (Dummy)	
First Stage F-Stat	46.624	46.624	37.168	37.168
N	4809	4809	4809	4809

Notes:
This table tests whether the long-term effects of inherited land on occupational choice and household consumption vary with measures of transaction costs in the market for land. All coefficients reported are 2SLS estimates. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Each column reports the 2SLS coefficients on inherited land, a measure of transaction costs in the market for land, and their interaction. The two endogenous variables are instrumented with two instruments: Predicted Share = (1/1+Brothers) and the interaction between Predicted Share and the measure of transaction costs. The dependent variable col 2 & 4 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. Non-Ag occupation (cols 1 & 3) is defined as the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. In cols 1 & 2 the measure of transaction costs is a z-score that combines 7 measures of transaction costs in the market for land that vary at the village level in 1999. This index includes fees for registering a landholdings, the cost of a Record of Rights certificate (RoR), the travel time taken to get to the registrar's office, the number of days taken for registration, a dummy coded as 1 if the RoR cannot be obtained in the village or tehsil/taluka of residence (i.e. administrative block), stamp duty paid for registration, and a dummy for whether there are additional registration fees. A z-score is computed for each component of this index (across villages) and the average z-score is the 'Transaction Costs Index'. In cols 3 & 4 the measure of transaction costs are whether there are additional fees that need to be paid in order to register one's landholdings in the village. Out of 246 villages, 60% have additional registration fees which on average amount to Rs. 1,019.83 (~ $20 in 1999). All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The specifications also included a control for distance to the closest town (km) and the interaction between this and landholdings. The excluded group are heads who are only children, from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

land that facilitated their migration to urban areas and I do not observe this in the data. Omitting these 'missing migrants' may result in an overestimate of the negative effect of land on migration or occupational choice. However, studies suggest that the wholesale rural-to-urban migration of families is extremely rare in India (Munshi and Rosenzweig, 2007) and Foster and Rosenzweig (2007) estimate that just 3-5% of *all* males aged 15-24 migrated in each of the three decades preceding the 1999 wave of the REDS survey.

Nevertheless, I simulate the effect of a 10% rural-to-urban migration rate on the reduced form relationship for non-agricultural occupation.[53] Across a series of covariate values – family landholdings and number of siblings – that are most favorable to overturning the reduced form effect, I find that the implied sex composition of such migrants required to overturn the reduced form is very different to what is empirically observed in the IHDS dataset which surveys such migrants.[54] In particular, even assuming that all 'missing migrant' families have landholdings in the 95th percentile, virtually all such migrants would also need to be the only son in their family for the confidence interval to not contain the actual reduced form estimate. In contrast, the IHDS data reveals that just 11% of such rural-to-urban migrants are only sons.

1.8.2 Addressing Instrument Validity: Conditional Independence Assumption

A number of studies (Sen, 1990; Gupta, 2005) document the fact that sex ratios, particularly in north-west India, are substantially skewed towards males reflecting a preference for sons. In this context, parents may influence the sex composition of their children through differential care for daughters or through sex-selective abortion. However, the majority of the REDS sample were born prior to the widespread availability of ultrasound technologies, which provided a low-cost way to facilitate sex selective abortion in rural India.[55] Nevertheless, I

[53] This is the census-based *individual* urban migration rate for the three decades preceding the REDS survey.

[54] Appendix C2 contains a detailed discussion of these simulations and the choice of covariate values.

[55] In the REDS data, 99.7% of household heads were born prior to 1980 when ultrasound technologies became widely available in India. While other methods such as amniocentesis may have preceded this, their availability in rural areas was more limited (Arnold *et al.*, 2002)

carry out a series of robustness checks that control for proxies of the demand for sex-selection established in the literature.[56] [57] Alternatively, given a preference for sons, differential stopping behavior may result in families with otherwise similar resources having different numbers of children based on when they achieve their desired number of sons. While both of these cases are of a concern to the conditional independence of the instrument, the robustness tests that follow suggest that this assumption is not violated.

Sex Selective Preferences: 'Balance' Test

A priori, it is unclear which household attributes are correlated with a preference for sons. One possibility is that households which prefer sons are also those willing to invest additional parenting effort in supporting them, implying that the coefficient on inherited land is biased upward. Alternatively, households which prefer more sons and are willing to influence the sex composition of their children may be less educated. Such parents may consequently also lack the human capital to guide their sons into more lucrative occupations and this would bias the estimate downwards.

The 'balance' tests in Table 1.8 provide further support for the conditional independence assumption for the instrument. Characteristics of the household head and his family are regressed on the instrument, predicted share, while controlling for the number of siblings and family landholdings as in specification (1.6). Columns (2)-(5) report the coefficient on the instrument by number of siblings a respondent has, while column (6) includes the entire sample. The estimates suggest that the instrument is independent of a number of respondent

[56] Specifically, I check whether the results are robust to the inclusion of controls for the demand for sex selection as in Vogl (2013). The estimates in column (2) of appendix C4 include 223 fixed effects for the exact permutation of the sex of older siblings. The estimated effects are largely similar to those in the main specification. The exception is the point estimate for consumption which is not precisely estimated, presumably on account of the reduction in covariate specific variation (the First stage F-statistic reduces by one-third).

[57] Differential investment or care may still pose a concern if the types of families who selectively provide less nutrition and other care for female children also have unobserved characteristics that influence the future success of their sons. In order to address this concern, I check in column (3) of appendix C4 whether the results are robust to a set of 18 dummies (0-10+ years, 6 month intervals) that control for the average spacing between siblings. The assumption here is that differential care would lead to an increase in the average spacing between births of siblings, but there is virtually no change in the estimated coefficients.

characteristics including age, education, birth order and average spacing between siblings for the respondents. In addition, the literature on sex selective abortion in India finds that a mother's education is an important (positive) correlate of sex-selective abortion (Pörtner, 2010). While, we observe such an imbalance for respondents with four siblings, this effect is reversed for those with three siblings. More generally, the pooled sample (column 6) reveals that there is no systematic relationship between sibling sex composition and mother's education across sibling cohort sizes.[58]

Differential Stopping Rules

As suggested above, another threat to the conditional independence of the instrument stems from son-preferring, differential stopping behavior (SP-DSB). On average, women in India are more likely to belong to families with a larger number of siblings and have less education on average (Jensen, 2003). While the inclusion of sibling fixed effects takes care of some of these concerns, it may still be the case that families with different fertility constraints or preferences end up with a similar number of children, resulting in an apples to oranges comparison in the regressions of interest.

To address concerns from stopping rules, I use an instrument that only uses variation from siblings born prior to the respondent — those who are, by definition, unaffected by stopping rules – and estimate my results restricting the sample to a subset where a fertility constraint is more likely to have been satisfied.[59] In both cases the estimates are largely consistent with those from the preferred specification. Appendix C4 discusses these tests in detail.

[58]Although there are imbalances for a few characteristics in specific sibling cohort sizes, they vary qualitatively across cohort sizes suggesting the absence of systematic bias across these characteristics, and are not significant in the pooled sample (column 6). The only imbalance that is significant for the pooled sample is the time at which the respondent became the head of the household, with those with a higher predicted share becoming heads at a slightly younger age. This result appears to be driven by individuals with four siblings, and is only marginally significant in the pooled sample. Appendix C3 reports the 2SLS estimates for the effect of inherited land on occupational choice and household consumption by sibling cohort size. The estimated effects do not vary qualitatively across sibling cohort sizes, once again reducing concerns about the imbalances in Table 1.8.

[59]The restriction is the subset of respondents whose youngest sibling is female. By definition, such families could not have engaged in son-preferring differential stopping behavior

TABLE 1.8: BALANCE CHECK FOR INSTRUMENT

		Reduced Form Estimates				
	Mean/ S.D.	Coefficient on Instrument† by Sibling Cohort Size				
Dependent Variable	Full Sample	2 siblings	3 siblings	4 siblings	5 siblings	Full Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Age of Head	49.263	-1.138	6.058**	-3.261	3.191	0.947
	14.149	(2.460)	(2.742)	(3.198)	(3.740)	(1.085)
Father's Education	1.428	0.202	-0.860**	0.567	0.063	0.103
(Years)	2.759	(0.421)	(0.414)	(0.671)	(0.788)	(0.205)
Mother's Education	0.407	0.198	-0.314*	0.610*	0.343	0.146
(Years)	1.440	(0.187)	(0.163)	(0.319)	(0.455)	(0.102)
Father in Agriculture	0.866	0.021	0.027	0.105	-0.017	0.008
(Primary Occupation)	0.340	(0.062)	(0.067)	(0.096)	(0.082)	(0.027)
Dowry Received	4.929	0.159	-0.073	-0.726	-1.703**	-0.377
Log(Rs. +1)	3.977	(0.578)	(0.631)	(0.828)	(0.794)	(0.261)
Age when Headship Assumed	33.397	-1.939	0.804	-6.436**	-3.547	-1.450*
(Years)	9.988	(1.786)	(1.902)	(2.736)	(2.725)	(0.790)
Age of Marriage	22.193	-1.103	0.693	-0.184	-0.575	-0.367
(Years)	5.173	(0.761)	(0.797)	(1.042)	(1.083)	(0.347)
Birth Order	2.585	-0.045	-0.078	-0.335	0.140	-0.112
	1.796	(0.153)	(0.225)	(0.395)	(0.475)	(0.107)
Sibling Spacing	4.912	-11.072	-0.794	0.238	-0.751**	1.699
(Years)	34.732	(8.557)	(0.494)	(1.129)	(0.382)	(4.083)
No. of Siblings FE	-	N	N	N	N	Y
Family Land FE	-	Y	Y	Y	Y	Y
District FE	-	Y	Y	Y	Y	Y
N	4809	729	811	777	726	4809

Notes:

This table presents summary statistics (mean and standard deviation) in Column 1 and assesses the conditional independence assumption of the instrument in Col 2-6, by seeing if it is independent of a number of characteristics of the household head and the head's parents. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. †Columns 2-6 report the coefficient estimate on the instrument, Predicted Share = 1/(1+Brothers), from a reduced form regression of the dependent variable on the instrument. Columns 2-5 assesses balance for household heads with varying numbers of siblings, while Column 6 includes all household heads. 'Father in agriculture' is coded as 1 if the primary occupation of the head's father is agriculture. 'Dowry received' reports the natural logarithm of the value of dowry payments given to the head or his parents at the time of marriage. 'Age when Headship Assumed' reports the age at which the respondent assumed headship of the household. 'Birth Order' is an integer value that is rising in parity (1 if eldest) and 'Sibling Spacing' computes the average interval between sibling births in number years. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children, from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. Brothers are defined as male siblings who grew up to at least the age of 10. Results are robust to alternative definitions and using ever born siblings. Robust standard errors are given in parentheses, asterisks denote significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Data Source: ARIS-REDS Dataset.

1.8.3 Addressing Instrument Validity: Exclusion Restriction Assumption

A key concern with estimation strategies relying on instrumental variables is the validity of the exclusion restriction assumption. In this context, the independent effects of sibling sex composition on human capital and dowry payments are of particular concern. In the former case, either a desire for diversification through reducing income covariance or decreasing returns to investing in the education of successive male children may result in respondents from families with more brothers having less education on average. In the case of dowry payments, families must typically pay a substantial sum for the dowry of a daughter when she marries. This may result in families with a higher composition of female children being more credit constrained thereby potentially influencing the cash reserves their sons are able to draw upon.

Controlling for education and dowry payments does not change the 2SLS estimates. [60] However, column (1) in Appendix C5 shows that the first stage for years of education is indeed negative implying that those with more land have more education, and that this bias runs counter to the finding that less land may be beneficial. In contrast, the parents of respondents with fewer sisters incur less dowry payments, suggesting that those who have less land may also have less credit constrained parents.[61] The first stage for land is 10-15 times as large as the first stage for these alternative channels, suggesting that 2SLS coefficients are less sensitive to these exclusion restriction violations.[62] Nevertheless, in

[60]In column (2) of Appendix C5 I add a set of 12 dummy variables (0-13, 1 year intervals) that control for the education of the respondent to the main specification (equation (1.6)). These controls mostly leave the estimates unchanged relative to the main specification. Similarly, in column (2), I add to the main specification a set of 19 dummy variables (Rs. -50,000 - Rs. 50,000, Rs. 5000 intervals) that control for net dowry receipts. Once again, the 2SLS estimates are left largely unchanged.

[61]Column (2) in Appendix C5 reports the first stage for a dummy coded as 1 if net dowry receipts are above the median (Rs. 0). Net dowry receipts are defined as the net sum of all dowry paid and received for all siblings of the respondent. The use of a median threshold rather than the natural logarithm – used for all other rupee values – is on account of many values being negative. However, while 37% of the respondents have 0 net dowry, the 5th percentile is Rs. -27,200 and the 95th percentile is Rs. 30,000, although the mean is just Rs. 1,471, leading to OLS estimation in levels to be greatly influenced by outliers.

[62]Conley *et al.* (2012) show that in the just identified case, the bias in the 2SLS estimate resulting from a violation of the exclusion restriction is proportional to $\frac{\gamma}{\pi}$ where γ is the size of the exclusion restriction violation and π is the first stage coefficient for the endogenous variable of interest. Given that in this case π is 125.95 in the preferred specification, Appendix C8 shows that γ has to be extremely large in order for the confidence

the next subsections I discuss how a number of placebo tests that can be used to support the claim that the primary channel through which the instrument influences outcomes is through the inheritance of land.

Placebo Test: Areas with Historically Matrilineal Inheritance Customs

Table 1.9 tests whether the findings in this paper are present in regions of India with historically matrilineal inheritance customs. Such customs entitle women to inherit more land on average relative to other parts of India. As such, the first stage for inherited land in these areas may be much weaker, but any other effects of sibling sex composition that operate through alternative causal channels should still influence outcomes in these areas. However, it may also be the case that sibling sex composition influences education or dowry receipts differently across matrilineal and patrilineal areas reducing the credibility of the placebo test. To address this concern, column (1) tests whether the first stage for education varies across matrilineal and patrilineal states using the REDS data.

If anything, it appears that in matrilineal states having fewer brothers leads to even more education for the respondent on average, relative to patrilineal states. As expected, the first stage for inherited land in matrilineal states (column 3) is much weaker and the point estimate is significantly different from patrilineal states. The reduced form effects of predicted share on non-agricultural occupation (column 3) and household consumption (column 4) are significantly different in matrilineal states, supporting the assumption that the instrument primarily operates through the land inheritance channel.

Placebo Test: Landed versus Landless Families

Respondents whose parents owned no land provide another means to investigate the causal channel of the instrument. The REDS dataset contains 1,315 such individuals who are not a

interval to cover zero using the 'union of confidence intervals' approach to checking the sensitivity of the estimates to violations of the exclusion restriction. The tests show that virtually all of the reduced form effect for non-agricultural occupation (Panel A) and household consumption (Panel B) has to operate through other channels, which is at odds with the ratio of the first stages for other candidate channels.

TABLE 1.9: THE FIRST STAGE AND REDUCED FORM EFFECTS IN STATES WITH MATRILINEAL INHERITANCE RULES (2SLS ESTIMATES)

	Patrilineal vs. Matrilineal Areas				Landed vs. Landless Parents			
Dependent Variable	Education (Years) (1)	Land Owned (Acres) (2)	Non-Ag Occupation (Binary) (3)	HH Consumption Log (Rs.) (4)	Education (Years) (5)	Net Dowry (Binary) (6)	Non-Ag Occupation (Binary) (7)	HH Consumption Log (Rs.) (8)
Predicted Share	1.158***	1.554***	-0.073***	0.110***	1.529***	-0.098***	-0.107***	0.198***
	(0.224)	(0.259)	(0.023)	(0.030)	(0.361)	(0.028)	(0.033)	(0.040)
Dummy for Restricted Sample	3.134	0.772	-0.711***	-0.741**	-1.038***	0.033	0.103***	-0.024
(Matrilineal/Landless Parents)	(1.945)	(0.607)	(0.257)	(0.308)	(0.339)	(0.028)	(0.031)	(0.037)
Predicted Share*Dummy	0.399	-1.245***	0.132**	-0.153**	-0.621	-0.017	0.137***	-0.199***
	(0.545)	(0.377)	(0.065)	(0.078)	(0.579)	(0.038)	(0.053)	(0.063)
Father's Occupation FE	Y	Y	Y	Y	N	N	N	N
Age & Education FE	Y	Y	Y	Y	N	N	N	N
District FE	Y	Y	Y	Y	Y	Y	Y	Y
No. of Siblings FE	Y	Y	Y	Y	Y	Y	Y	Y
Family Land FE	N	N	N	N	Y	Y	Y	Y
Can Reject Null Hypothesis in Restricted Sample?	Y	Y	N	N	N	Y	N	N
Depvar Mean	5.077	1.652	0.430	13.413	5.540	0.703	0.353	10.377
First Stage F-statistic	25.08	33.667	-	-	17.974	12.176	-	-
N	11181	11181	11181	11181	6124	6124	6124	6124
Data Source	Indian Human Development Survey, 2004-2005				ARIS-REDS, 1999 Wave			

Notes:
This table tests whether the first stage and reduced form vary differentially in states with matrilineal and patrilineal land inheritance rules and between household heads whose parents were landless and landed towards supporting the exclusion restriction assumption. In columns 1-4 the data is limited to Hindu male household heads in the 2004-2005 wave of the Indian Human Development Survey, who reside in rural areas. In columns 5-8 the sample is limited to Hindu male household heads whose parents were either landed or landless in the 1999 ARIS-REDS's survey. For columns 1-4 the' restricted sample' is the subset of households Kerala, Assam, Arunachal Pradesh, Meghalaya, Manipur, Mizoram, Tripura, Nagaland or Sikkim. These are areas reported as having Matrilineal or Bilateral Inheritance laws in Agarwal (2004). For columns 5-8 it is the subset of households whose parents were landless. First stage F-statistics and sample sizes are reported separately for the full and restricted samples. The 'Dummy for Restricted Sample' corresponds to a dummy coded as 1 if the observation is from the restricted sample and 0 otherwise. Each column reports the coefficients from the reduced form regression of the dependent variable on the instrument, Predicted Share = (1/1+Brothers), the 'Dummy for Restricted Sample' and their interaction. The dependent variable in Column 1 and 5 are the years of education of the head of the household. The dependent variable in Column 2 is current land owned (acres). The dependent variable in Column 3 and Column 7 is a dummy variable if the head has a non agricultural occupation. Non-Ag occupation is defined by the primary status reported by the respondent in the survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable column 4 and column 8 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. The dependent variable in column 6 is a dummy variable for whether net dowry receipts are above or below the median. The former is calculated as the net sum of all dowry payments and receipts for the parents, which are reported for each sibling of the head of the household. 37% of the sample do not report paying or receiving dowry. All specifications include district fixed effects (99 dummies) and the number of siblings (14 dummies). Specifications 1-4 include fixed effects for head's education (0-15 years, 1 year intervals, 14 dummies), age of the household head (20-100, 5 year intervals, 15 dummies), father's occupation (89 dummies), and father's education (0-15, years, 1 year intervals, 14 dummies). Parent's landholdings are not reported in the IHDS data. Specifications 5-8 include fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies). The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Data Sources: ARIS-REDS Dataset and Indian Human Development Survey.

part of the main sample of 4,809 respondents. By definition, these individuals cannot inherit land from their parents, however any independent effects of sibling sex composition should still influence their outcomes. Once again, it may be the case that sibling sex composition and dowry operate differently for landless families. Columns (5) and (6) of Table 1.9 show that the point estimate on inherited land from the first stage regressions for education and dowry are not systematically different across individuals from landed and landless families.

In contrast, column (7) and column (8) show that the reduced form effects of the instrument on non-agricultural occupation and consumption, respectively, are significantly different.[63] As in the case of the matrilineal placebo, the differential effects are large in magnitude, and precisely estimated. Once again, these estimates lend support to the claim that the effects of sibling sex composition on occupational choice and household consumption primarily operate through the inheritance of land.[64]

1.9 Conclusion and Discussion

Using a novel instrumental variables strategy that relies on variation arising from sibling sex composition and Hindu inheritance customs, the primary contribution of this paper is to shed light on the importance of frictions – both market and non-market – and inherited land in explaining the observed persistence of labor in the agricultural sector.

Contrary to theories of poverty traps emphasizing the importance of initial endowments, I find that inherited land does not facilitate the movement of labor out of the agricultural sector. Rather, for the majority of the population I find the opposite: inheriting land reduces the likelihood of exiting agriculture both within rural areas and through migration to urban

[63] It is worth noting here that household heads with more brothers – those with a lower predicted share – are more likely to report that they set up a new household when they became the head of their family as opposed to assuming their parent's household. However, this does not differ across respondents from landed and landless families.

[64] Appendix C7 details two further placebo tests: First, the reduced form effect for non-agricultural occupation (column 2) and household consumption (column 3) are qualitatively different but imprecisely estimated for those inheriting after reforms to inheritance laws that allow sisters to inherit land. Second, the null hypothesis (zero effect) *cannot* be rejected for the the reduced form effect for household consumption in urban areas (column 4).

areas. For those inheriting below median landholdings the effect sizes are more than ten times as large.

Further, I find that cultural obligations and land market transaction costs undermine the benefits of inherited land. In the presence of such frictions, inheriting land results in two competing forces. On the one hand larger landholdings result in a wealth effect, expanding the consumption of inheritors. On the other, they influence the long-term occupational trajectory of inheritors – with younger inheritors more affected – nudging them towards the agricultural sector. This latter effect results in foregone higher returns in the non-agricultural sector, which may be particularly large for smaller farmers. Depending on the balance of these effects, the net effect of inherited land on consumption can be *zero* or even *negative*.

Individuals wishing to part with inherited land passed through a family for generations may face unique constraints. However, beneficiaries of asset transfer programs may be prohibited from selling these assets or face constraints in being able to lease them out through markets. As such, these programs may have similar unanticipated consequences for labor mobility. In the context of land reform, this would appear to be especially true for small amounts of land (less than 3 acres) which I find to greatly influence labor mobility and where the gains to switching to non-agricultural work are largest.

The findings of this paper suggest that frictions in land markets may be an important source of labor market misallocation in rural India. As a consequence, the most productive farmers may be unable to enter agriculture resulting in suboptimal agricultural productivity. Future work may test this implication by estimating how land market reforms and changes to inheritance laws influence agricultural productivity and labor mobility.[65] A second implication is that interventions that improve the returns to 'bad farmers' remaining in agriculture will create additional distortions in labor markets. Spatial and temporal variation in large scale programs intended to improve agriculture would provide a test of this implication.

[65]For example, in India the computerization of land registries and amendments to the Hindu Succession Act may have important consequences for labor market outcomes through influencing the channels discussed in this paper.

An important caveat to my findings is that consumption may not adequately capture welfare. Incorporating measures of productivity as discussed above and subjective measures of well-being will strengthen the evidence for misallocation. Finally, given the preponderance of patrilineal inheritance laws across the developing world, a similar instrumental variables strategy can be used to understand the importance of inherited assets for labor market outcomes across a number of contexts.[66]

[66] For example, Kuran, 2012 studies the implications of Islamic inheritance laws, where male heirs typically inherit twice the amount that daughters inherit.

Chapter 2

The Value of Advice: Evidence from the Adoption of Agricultural Practices[1]

2.1 Introduction

Agricultural productivity varies dramatically around the world. For example, India is the second largest producer of cotton in the world, after China. Yet, Indian cotton productivity ranks 78th in the world, with yields only one-third as large as those in China. While credit constraints, missing insurance markets, and poor infrastructure may account for some of this disparity, a variety of observers have pointed out the possibility that suboptimal agricultural practices and poor management practices may also be to blame (Jack, 2011).

This is not a novel idea. For decades, the Government of India, like most governments in the developing world, has operated a system of agricultural extension, intended to spread information on new agricultural practices and technologies, through a large work force of public extension agents. However, evidence of the efficacy of these extension services is quite limited. In India, dispersed rural populations, monitoring difficulties and a lack of

[1]Co-authored with Shawn Cole (Harvard Business School)

accountability hamper the efficacy of traditional extension systems: fewer than 6% of the agricultural population reports having received information from these services.[2]

This paper examines whether the introduction of a low-cost information and communications technology (ICT), able to deliver timely, relevant, and actionable information and advice to farmers at dramatically lower cost than any traditional service can improve agricultural management. We evaluate Avaaj Otalo (AO), a mobile phone-based technology that allows farmers to call a hotline, ask questions and receive responses from agricultural scientists and local extension workers. Callers can also listen to answers to questions posed by other farmers.

Working with the Development Support Centre (DSC), an NGO with extensive experience in delivering agricultural extension, the research team randomly assigned toll-free access to AO to 800 households, with an additional 400 households serving as a control group.[3] The households were spread across 40 villages in Surendranagar district in Gujarat, India, and randomization occurred at the household level.

The AO service also included weekly push content, delivering time sensitive information such as weather forecasts and pest planning strategies directly to farmers. This paper presents the results from two rounds of household surveying conducted one (midline) and two (endline) years after the baseline survey.[4] In addition, we report the results from a survey of the peers of study respondents that was conducted by phone.[5]

Demand for agricultural information is substantial: more than 80% of the treatment group called into the AO line over two years. The average treatment respondent made

[2] This estimate is from the 59th round of the National Sample Survey (NSS) and asks farmers about their information sources for 'modern agricultural technologies'. See Glendenning *et al.*, 2010 for a detailed discussion of this data.

[3] Of the 800 households assigned to AO, 400 were assigned to also receive traditional agricultural extension services. This will allow us to evaluate the complementarity of in-person and ICT-based training.

[4] In a previous version of this paper (Cole and Fernando, 2014) we reported the results of this intervention on a subset of the sample 7 months after it had been administered. In this paper we report the long-term results for the *entire* sample.

[5] 'Peers' refer to network contacts in a respondent's network. At baseline we asked all respondents to give us the names of their three top agricultural contacts.

20 calls and used the service for more than 2.5 hours. We show that AO had a range of important, positive effects on farmer behavior. It significantly changed farmers' sources of information for sowing and input-related decisions. In particular, farmers relied less on commissions-motivated agricultural input dealers for pesticide advice and less on their prior experience for fertilizer-related decisions. Instead, farmers dramatically increase their usage of and trust in mobile phone-based information across a number of agricultural decisions.

Importantly, treated farmers were significantly more likely to adopt agricultural practices and inputs recommended by the service. These inputs choices include recommended seed varieties, fertilizers, pesticides and irrigation practices. These input choices also translate into higher crop yields: they are 3.5% higher for cotton and 26.3% higher for cumin.

Variation in the extent of respondents' social networks treated by our intervention allows us to estimate peer effects. We find that individuals with more treated peers – members in their social network who are also assigned to treatment – make limited changes in their sources of information, plant more cumin and report lower pest-related cotton losses.

Finally, we conduct a series of willingness to pay experiments to estimate demand for AO. Average willingness to pay for a 9-month AO subscription across multiple price elicitation methods is roughly \$2. However, the price of providing the service over this period is \$7. This implies a \$12 subsidy per farmer to run the service over two years. In support of such a subsidy, we estimate that each *dollar* invested in AO generates a return of more than \$10, with the return for a two-year subscription at more than \$200.

This study makes the following contributions. We demonstrate that informational inefficiencies are real,[6] and that farmers are aware they lack information: there is considerable demand for high quality agricultural information. We present the first rigorous evidence that a low-cost agricultural extension service (costing as little as US \$.83 per farmer per month) can change behavior. We provide some evidence of the existence of a "digital divide," as richer and more educated individuals use the service more. This is true even

[6]Informational inefficiencies in the context of technology adoption have been defined as a situation in which farmers may not be aware of new agricultural technologies, or how they should be utilized (Jack, 2011)

though the treatment group is relatively homogeneous, and even though the technology was delivered for free, and specifically designed to be accessible to an illiterate population. As a methodological contribution, we demonstrate that surveying by mobile phones can be conducted effectively and cheaply (the average "all-in" cost of a phone survey was \$2.51, compared to over \$10 for a paper survey), in a developing country context.[7]

First, this paper contributes to an understanding of the mechanisms underlying the dramatic variation in productivity of firms and farms in developing countries, and the role of management practices in improving productivity. These large productivity differences have in part motivated the recent literature on non-aggregative growth (Banerjee and Duflo, 2005; Hsieh and Klenow, 2009). While a large literature focuses on the microeconomics of technology adoption (for a survey, see Foster and Rosenzweig, 2010a), we instead focus on whether a consulting-like service can facilitate improved production practices. (Duflo *et al.*, 2011.) Our treatments differ from much previous work in this space in that participants receive a continuous flow of demand-oriented information, rather than a one-off provision of supply-driven information. See McKenzie and Woodruff, 2012 for a discussion of training and consulting evidence for small firms in developing countries.

More specifically, this paper advances the literature on the efficacy of agricultural extension (Feder *et al.*, 1987, Gandhi *et al.*, 2009,Duflo *et al.*, 2011). The existing literature finds mixed evidence of efficacy, though it is not clear whether this is due to variation in programs offered, or methodological challenges associated with evaluating programs without plausibly exogenous variation (Birkhaeuser *et al.*, 1991). This paper complements recent evidence on the historical efficacy of agricultural extension in promoting the adoption of new agricultural technologies in India (Bardhan and Mookherjee, 2011), and provides guidance as to lower-cost solutions for delivering advice. To our knowledge, our study is the first rigorous evaluation of mobile phone-based extension and, more generally, the first evaluation of a demand-driven extension service delivered by any means.

[7]In a related study, we test the validity of mobile-phone based surveying by randomly assigning one module of a household survey to be administered either by mobile phone, or by paper survey.

This paper is organized as follows. The next section places this paper's contribution in the literature, and provides context and the details of the AO intervention. Section 3 presents the experimental design and the empirical strategy, while Section 4 presents the results from the two years of survey data. Following this, Section 5 considers threats to the validity of the results, and Section 6 concludes.

2.2 Context and Intervention Description

2.2.1 Agricultural Extension

According to the World Bank, there are more than 1 million agricultural extension workers in developing countries, and public agencies have spent over $10 billion dollars on public extension programs in the past five decades (Feder, 2005). The traditional extension model, "Training and Visit" extension, has been promoted by the World Bank throughout the developing world and is generally characterized by government-employed extension agents visiting farmers individually or in groups to demonstrate agricultural best practices (Anderson and Birner, 2007). Like many developing countries, India has a system of local agricultural research universities and district level extension centers, producing a wealth of specific knowledge. In 2010 the Government of India spent $300 million on agricultural research, and a further $60 million on public extension programs (RBI, 2010).

Yet, traditional extension faces several important challenges that limit its efficacy.

Spatial Dimension: Limited transportation infrastructure in rural areas and the high costs of delivering information in person greatly limit the reach of extension programs. The problem is particularly acute in interior villages in India, where farmers often live in houses adjacent to their plots during the agricultural cycle, creating a barrier to both the delivery and receipt of information.

Temporal Dimension: As agricultural extension is rarely provided to farmers on a recurring basis, the inability of farmers to follow-up on information delivered may limit their willingness to adopt new technologies. Infrequent and irregular meetings limit the ability to provide timely information, such as how to adapt to inclement weather or unfamiliar pest infestations.

Institutional Rigidities: In the developing world, government service providers often face institutional difficulties. The reliance on extension agents to deliver in-person information is subject to general monitoring problems in a principal-agent framework (Anderson and Feder, 2007). For example, monthly performance quotas lead agents to target the easiest-to-reach farmers, and rarely exceed targets. Political capture may also lead agents to focus outreach on groups affiliated with the local government, rather than to marginalized groups for whom the incremental benefit may be higher. Even when an extension agent reaches farmers, the information delivered must be locally relevant, and delivered in a manner that is accessible to farmers with low levels of literacy.

The importance of these constraints is difficult to overstate (Birkhaeuser *et al.*, 1991, Saito and Weidemann, 1990.) A recent nationally representative survey shows that just 5.7% of farmers report receiving information about modern agricultural technologies from public extension agents in India (Glendenning *et al.*, 2010.) This failure is only partly attributable to the misaligned incentives of agricultural extension workers; more fundamentally, it is attributable to the high cost of reaching farmers in interior rural areas.

Finally, a potential problem is that information provision to farmers is often "top-down." This may result in an inadequate diagnosis of the difficulties currently facing farmers, as well as information that is often too technical for semi-literate farming populations. This problem may affect adoption of new technologies as well as optimal use of current technologies.

In the absence of expert advice, farmers seek out agricultural information through word of mouth, generic broadcast programming, or agricultural input dealers, who may be poorly informed or face incentives to recommend the wrong product or excessive dosage (Anderson and Birner, 2007).[8]

[8]An audit study of 36 input dealerships in a block near our study site provides a measure of the quality of advice provided by commissions-motivated input dealers. The findings suggest that the information provided is rarely customized to specific pest management problems of the farmer, and often takes the form of ineffective pesticides that were traditionally useful, but are no longer effective against the dominant class of pests that afflict cotton cultivation.

These difficulties combine to limit the reliable flow of information from agricultural research universities to farmers, and may limit their awareness of and willingness to adopt new agricultural technologies. Overcoming these "informational inefficiencies" may therefore dramatically improve agricultural productivity and farmer welfare. The emergence of mobile phone networks and the rapid growth of mobile phone ownership across South Asia and Sub-Saharan Africa has opened up the possibility of using a completely different model in delivering agricultural extension services.

2.2.2 Avaaj Otalo: Mobile Phone-Based Extension

Roughly 52% of the Indian labor force, or 270 million people, are engaged in agriculture. As approximately 36% own a mobile phone, mobile phone-based extension could serve as many as 97 million farmers nationally, including over 9 million in Gujarat alone [9]. Mobile phone access has fundamentally changed the way people communicate with each other, and has increased information flows across the country's diverse geographic areas. As coverage continues to expand in rural areas, mobile phones carry enormous promise as a means for delivering extension to the country's numerous small and marginal farmers (Aker, 2011).

Our intervention utilizes an innovative information technology service, Avaaj Otalo (AO). AO uses an open-source platform to deliver information by phone delivered by phone. Information can be delivered to and shared by farmers. Farmers receive weekly push-content, which includes detailed agricultural information on weather and crop conditions that are delivered through an automated voice message.

Farmers can also call into a toll-free hotline that connects them to the AO platform and ask questions on a variety of agricultural topics of interest to them. Staff agronomists at the Development Support Centre (DSC) – our field partner – with experience in local agricultural practices receive these requests and deliver customized advice to these farmers, via recorded voice messages. Farmers may also listen and respond to the questions their

[9] These figures are calculated using estimates from the 2010-2011 Indian Ministry of Labor and the Annual Report of the Telecom Regulatory Authority of India

peers ask on the AO platform, which is moderated by DSC. The AO interface features a touch-tone navigation system with local language prompts, developed specifically for ease of use by semi-literate farmers. The platform, which has now been deployed in a range of domains, was initially developed as part of a Berkeley-Stanford research project on human-computer interaction, in cooperation with the DSC in rural Gujarat (Patel *et al.*, 2010).

Mobile phone-based extension allows us to tackle many of the aforementioned problems with traditional extension. AO has the capability to reach millions of previously excluded farmers at a virtually negligible marginal cost. Farmers in isolated villages can request and receive information from AO at any point during the agricultural season, something they are typically unable to do under traditional extension. Farmers receive calls with potentially useful agricultural information on their mobile phones, and need not leave their fields to access the information. In case a farmer misses a call, she can even call back and listen to that information on the main line. AO thus largely solves the spatial problems of extension delivery discussed earlier.

A considerable innovation of AO is tackling the temporal problem of extension delivery. The agricultural cycle can be subject to unanticipated shocks such as weather irregularities and pest attacks, both of which require swift responses to minimize damage to a standing crop. Because farmers can call in and ask questions as frequently as they want, they can get updated and timely information on how to deal with these unanticipated shocks. This functionality may indeed increase the risk-bearing capacity of farmers by empowering them with access to consistent and quality advice.

With respect to problems of an institutional nature mentioned earlier, AO facilitates precise and low-cost monitoring. The computer platform allows easy audits of answers DSC agronomists offer, greatly limiting the agency problem. Additionally, the AO system allows for demand-driven extension, increasing the likelihood that the information is relevant and useful to farmers. Push-content is developed by polling a random set of farmers each week to elicit a representative set of concerns. In addition to this polling, the questions asked by calling into AO also provide the information provider a sense of farmers' contemporaneous concerns. This practice of demand-oriented information provision should improve both the allocation and the likelihood of utilization of the information.

However, while AO overcomes many of the challenges of traditional extension, it eliminates in-person demonstrations, which may be a particularly effective way of conveying information about agricultural practices. As discussed in the following section, our study design allows us to estimate the extent to which in-person extension serves as a complement to AO-based extension, by providing a subset of farmers with both traditional extension administered through staff at DSC and toll-free access to AO.

2.3 Experimental Design & Empirical Strategy

Two administrative blocks [10] in Surendranagar district, Chotila and Sayla, were chosen as the site of the study, as our field partner, DSC, had done work in the area. Lists of farmers were enumerated in 40 villages, with the criteria for selection being that they were 1.) interested in participating in the study, 2.) grew cotton, 3.) owned a mobile phone and 4.) were the chief agricultural decision maker of their household.

A sample of 1200 respondents was selected from this pool, with 30 households in each village participating in the study. Figure 2.1 summarizes the experimental design used in this study. Treatments were randomly assigned at the household-level using a scratch-card lottery. The sample was split into three groups of 400 households each. The first treatment group (hereafter, AOE) received toll-free access to AO in addition to traditional extension. The traditional extension component consisted of a single session each year lasting roughly two-and-half hours on DSC premises in Surendranagar. The second treatment group (hereafter, AO) received toll-free access to AO, but no offer of traditional agricultural extension, and a final 400 households served as the control group. In addition, among the two treatment groups (AO and AOE), 500 received bi-weekly reminder calls (hereafter,

[10] A block is an administrative unit below the district level

reminder group) to use the service while the remaining 300 did not.

Figure 2.2 lists the timeline for the study. Baseline data was collected in June and July, 2011, and a phone survey consisting of 798 respondents was completed in November 2011.[11] The midline survey was completed by August 2012, and the endline survey was completed by August 2013.

To gauge balance and describe our first stage, we compute a simple difference specification of the form:

$$y_{iv} = \alpha_v + \beta_1 \; Treat_{iv} + \varepsilon_i \tag{2.1}$$

where, α_v is a village fixed effect, $Treat_{iv}$ is an indicator variable that takes on the value 1 for an individual, i, in village v assigned to a treatment group and 0 for an individual assigned to the control group. We report robust standard errors below the coefficient estimates.

Because of random assignment, the causal effect of the intervention can be gauged by computing a standard difference-in-difference specification:

$$y_{ivt} = \alpha_v + \beta_1 \; Treat_{iv} + \beta_2 \; Post_t + \beta_3 \; (Treat * Post)_{ivt} + \varepsilon_i \tag{2.2}$$

where, α_v and $Treat_{iv}$ are as above, $Post_t$ is an indicator variable that takes on a value of 1 if the observation was collected at the endline (or the midline) 0 otherwise, and $(Treat * Post)_{ivt}$ is the interaction of the preceding two terms. The empirical results largely estimate (3.1) for the outcome variables of interest, using robust standard errors.

In addition, we explore heterogeneity in the treatment effect by interacting the difference-in-difference specification in equation (3.2) with a dummy variable capturing the hetero-

[11]The previous version of this paper (Cole and Fernando, 2014) analyzed treatment effects using results from this phone survey.

FIGURE 2.1: EXPERIMENTAL DESIGN

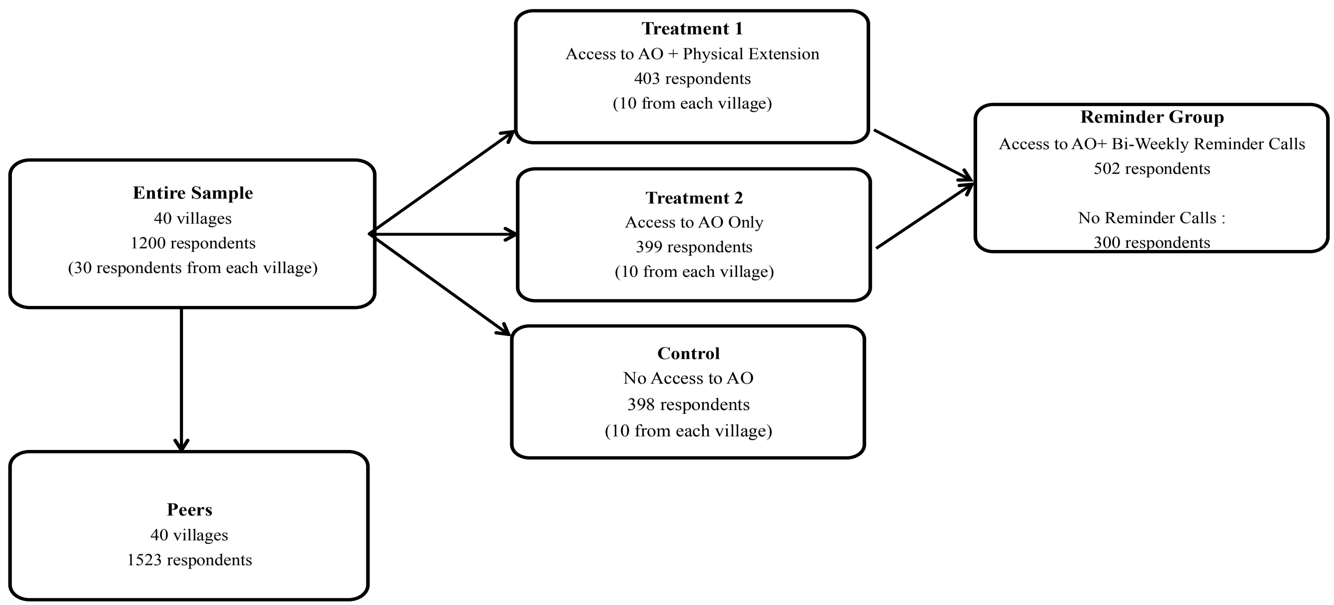


FIGURE 2.2: PROJECT TIMELINE

Date	Event
May/2011	Cotton planting decisions begin
May/2011	Listing for baseline survey
Jul/2011	Baseline (paper) survey
Aug/2011	AO training for treatment respondents
Aug/2011	AO service activated for all treatment respondents
Sep/2011	Reminder calls started
Nov/2011	Physical extension Round 1
Nov/2011	Phone Survey Round 1
Dec/2011	Phone Survey Round 2
Mar/2012	Peer Survey
Jun/2012	Midline (Paper) Survey
Aug/2012	AO training for treatment respondents Round 2
Oct/2012	Field visits to gather information on Rabi planting decisions
Nov/2012	Peer Survey Part 2
Nov/2012	Physical Extension Round 2
Mar/2013	Phone Survey 3
Jul/2013	Endline (Paper) Survey
Jul/2013	Willingness to Pay Study
Jul/2013	Ending push calls/intervention

Notes:

1. Phone surveys were conducting with roughly half the treatment sample i.e. 400 respondents who had access to AO and the 398 control respondents who did not have access to AO

2. Peer surveys reached out to roughly 1000 farmers listed as farmer friends by respondents.

geneity of interest:

$$\begin{aligned}
y_{ivt} &= \alpha_v + \beta_1\ Treat_{iv} + \beta_2\ I(X_{iv} > Median) + \beta_3\ Treat_{iv} * I(X_{iv} > Median) \\
&\quad + \beta_4\ Treat_{iv} * Post_t + \beta_5\ I(X_{iv} > Median) * Post_t \\
&\quad + \beta_6\ Treat_{iv} * I(X_{iv} > Median) * Post_t + \varepsilon_{ivt} \qquad (2.3)
\end{aligned}$$

where, X_{iv} is the variable across which we explore heterogeneity in treatment effects, and $I(X_{iv} > Median)$ a dummy equal to one when the observation is above the median level of X_{iv}. The results presented in the section on heterogeneous treatment effects are virtually identical when X_{iv} is included as a continuous variable.

While dramatically increasing statistical power, the decision to randomize at the household rather than village level raises the possibility that the control group may also have access to information through our treatment group. This suggests that any treatment effects may in fact underestimate the value of the service.[12] We collected information on 1523 peers of study respondents using a phone survey in March 2012 and November 2012, hereafter the 'peer survey'.[13] This data allows us to estimate whether the treatment also influences the outcomes of individuals in our study respondents' social networks. We estimate the extent of such peer effects or information spillovers with the following specification:

$$y_{iv} = \alpha_v + \beta(\frac{\#\ References\ in\ Treatment}{\#\ References})_{iv} + \sum_{i=2}^{7} I(\#\ References = i)_{iv} + \varepsilon_{iv} \qquad (2.4)$$

where, α_v is as above, $\sum_{i=2}^{7} I(\#\ References = i)_{iv}$ is a fixed effect for the number of peers who cite a respondent as a top agricultural contact and $(\frac{\#\ References\ in\ Treatment}{\#\ References})_{iv}$ is the fraction of these respondents who are assigned to treatment.

[12] In order to control for spillovers, we estimated the main difference-in-differences specification with controls for the fraction of a respondent's social network that was also a part of the study and the fraction that was assigned to treatment. These controls leave the main estimates largely unchanged.

[13] At baseline we asked all respondents to list the three contacts with whom they most frequently discussed agricultural information and collected their phone numbers. The 'peer survey' collected information from all these contacts. Note, some of these 1523 peers may themselves be study respondents. The analysis largely focuses on 1114 non-study peers

We did not prepare a pre-analysis plan prior to undertaking the study. This was in part due to the dynamic nature of the treatment: the service responded to farmer questions, and ex-ante, it was not always clear which subjects farmers would inquire about. We address concerns about multiple inference in two ways. First, we use the content generated by farmers, and by our agronomist, as a broad guide for conducting empirical analysis.[14] Second, we aggregate agricultural practices into indices, following, for example, Kling *et al.* (2007).

To construct indices, we do the following. Our agronomist, Tarun Pokiya, characterized each agricultural practice as either positive, negative, or neither (e.g., neutral or situation-dependent). We then aggregate all variables with unambiguous value by calculating a z-score for each component, and then take the average z-score across components. Each component z-score is computed relative to the control group mean and standard deviation at baseline. We have compared this to the method that uses 'seemingly unrelated regression' which gives slightly different standard errors and identical point estimates but is virtually indistinguishable from this method as suggested by Kling *et al.* (2007).

2.3.1 Summary Statistics and Balance

In this section we assess balance between the 'combined treatment' group (AO + AOE) and the control group and the subset of the treatment group that receives reminder calls, referred to as the 'reminder' group, and the control group. We do not find important differences in the separate treatment effects of the AO and AOE groups and the interaction of these treatments with reminder calls, so we instead focus on the 'combined treatment' group and the 'reminder' group.[15]

Table 2.1 contains summary statistics for age, education, income and cultivation patterns for respondents in the study, using data from a baseline paper survey conducted in July

[14]See Appendix Table D1 for details of questions asked by farmers on the AO service and push content provided.

[15]Appendix D2 tests the balance for the AO and AOE group, while Appendix D3 reports treatment effects separately for the AO and the AOE group.

and August of 2011. Column (1) reports the mean and standard deviation for the control group, column (2) tests the initial randomization balance between the combined treatment group and the control group. Finally, column (3) tests the balance between the reminder group and the control group.

We see that respondents are on average 46 years old and have approximately 4 years of education. Columns (2) and (3) show that the randomization was largely successful for both treatment groups across demographic characteristics (Panel B) and indices capturing information sources, crop-specific and general input use (Panel C). However, an imbalance exists in the area of cotton planted between the treatment groups and the control group in 2010 but not in 2011 (both periods are prior to treatment).[16] The combined treatment group is also more likely to grow wheat, but this crop is mostly grown for home consumption in this context.

As cotton is the most important crop in our sample, we take a conservative approach to the possibility that baseline cotton levels affect subsequent outcomes and include as controls the area of cotton cultivated in 2010 and its interaction with the 'Post' term in both the difference-in-difference specification (equation (3.2)), the heterogeneous effects specification (equation (2.3)) and peer effects specification (equation (3.3)).[17]

2.4 Experimental Results

Cole and Fernando, 2014 describe initial differences measured seven months after the implementation of AO. After seven months, take-up among the treated group was high, and we measured several important changes in agricultural behavior: farmers changed

[16]Note, the 2011 figures for wheat and cumin are not reported as they are grown during the Rabi season after the treatment was administered.

[17]Appendix Table D4 provides a more systematic treatment of balance in our sample. We look for significant differences in baseline characteristics between the combined treatment group and control, and the reminder group and control respondents. Among the differences computed using the latter specification (examining all 2,295 baseline variables) we find that 0.7% are significantly difference from zero at the 1% level, 4.4% are different at the 5% level of significance and 9.5% at the 10% level. These results confirm that the randomization was successful, and that the cotton imbalance is a result of chance rather than any systematic mistake in the randomization mechanism.

TABLE 2.1: SUMMARY STATISTICS AND BALANCE

Dependent Variable	Control Mean (Baseline) (1)	Treat-Control ITT (2)	Treat+Reminder- Control ITT (3)
A. Sample Size			
Entire Sample	398	1200	900
B. Individual Characteristics			
Age	46.539	-0.369	-0.249
	(15.161)	(0.915)	(0.999)
Years of Education	4.235	-0.187	-0.177
	(3.836)	(0.230)	(0.253)
Landholdings- Acres	6.077	0.095	-0.013
	(5.596)	(0.332)	(0.356)
Agricultural Income	11.551	-0.006	0.022
(log rupees)	(1.361)	(0.085)	(0.093)
C. Indices (Standard Deviation Units)			
Mobile Phone-Based	0.000	0.002	0.002
Information Usage	(0.289)	(0.018)	(0.019)
Cotton Management	0.000	-0.024	-0.028
	(0.433)	(0.025)	(0.027)
Wheat Management	0.000	-0.005	0.004
	(0.347)	(0.023)	(0.024)
Cumin Management	0.000	-0.007	0.007
	(0.303)	(0.018)	(0.020)
Pesticide Management	0.000	-0.003	-0.013
	(0.306)	(0.021)	(0.021)
Fertilizer Management	0.001	-0.003	-0.014
	(0.311)	(0.021)	(0.021)
D. Agricultural Activity			
Planted Cotton (2010)	0.984	0.002	0.000
	(0.126)	(0.008)	(0.009)
Area Cotton Planted (2010)	4.448	0.422*	0.449*
(Acres)	(3.622)	(0.232)	(0.261)
Area Cotton Planted (2011)	4.990	0.293	0.236
(Acres)	(3.846)	(0.247)	(0.272)
Planted Wheat (2010)	0.776	-0.053**	-0.040
	(0.417)	(0.025)	(0.027)
Area Wheat Planted (2010)	1.171	0.016	0.074
(Acres)	(1.346)	(0.089)	(0.106)
Planted Cumin (2010)	0.425	-0.018	-0.004
	(0.495)	(0.028)	(0.031)
Area Cumin Planted (2010)	0.762	-0.019	-0.010
(Acres)	(1.406)	(0.083)	(0.092)

Notes:

1. This table reports summary statistics and assesses balance across groups using data from the baseline survey, conducted between June 26 and August 11, 2011.

2. Participants were randomized into three groups. AO group received AO access. AOE group received AO access and physical extension. 'Treat' refers to the combined treatment group. The control group received neither treatment.

3. `Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.

4. The indices aggregate information over multiple outcomes for which we expect unidirectional treatment effects. Each index consists of the average of the z-scores for each component of the index, with the control group mean and standard deviation as reference.

5. Mobile phone-based information usage index: Aggregates mobile phone use across crop decision, soil preparation, pest identification, weather, cotton pesticides, cotton fertilizers, wheat fertilizers, cumin pesticides and cumin fertilizers.

6. Management practices indices: seed usage + pesticide purchase + pesticide usage + pesticide quantities + pesticide expenditure + fertilizer purchase + fertilizer usage + fertilizer quantities + fertilizer prices for the three different crops – cotton, wheat and cumin.

7. Pesticide management index: dummy to indicate purchase/use of a pesticide + pesticide expenditure + pesticide quantities across purchase and usage decisions.

8. Fertilizer management index: dummy to indicate purchase/use of a fertilizer + fertilizer expenditure + fertilizer quantities across purchase and usage decisions.

9. Seed management index: dummy to indicate purchase/use of recommended seeds + seed expenditure + seed quantities across purchase and usage decisions.

10. Agricultural income refers to income earned from all crops grown in the past 12 months

11. Column 1 shows the summary statistics (mean and standard deviation) for the control group at baseline.

12. Columns 2-3 report an Intention to Treat (ITT) estimate of the difference in means (and robust standard error) between the treatment groups and control group.

13. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

their information gathering activity, relying less on peers and more on mobile-phone based advice; treatment farmers were more likely to adopt more effective pesticides, and reduce expenditure on hazardous, ineffective pesticides; and treated farmers were more likely to grow cumin. A short-coming of the early evidence was that it was based on interviews, conducted by telephone, of only a sub-sample of study participants. In the sections below, we describe results after treatment households had been offered the service for two full years primarily using the difference-in-differences specification in equation (3.2).

2.4.1 First Stage: Take-Up and Usage of AO

Table 2.2 reports information on take-up and usage (first stage). While control respondents were not barred from AO usage, only four control respondents called into the AO line by the midline and a further 25 had called in after two years. As a result, virtually all AO usage is accounted for by respondents in the treatment group. As of August 2013, two years after commencement of the service, 88% of the treatment group had called into the AO line, making an average of 22 calls. This represents a substantial increase from the midline, where 64% of the combined treatment group called in, making an average of 9 calls. The mean usage for treatment respondents is over 2.5 hours, as compared to 1.3 hours at midline. On average, treatment respondents have listened to 68% of total push call content (83% of total push call content was the average at the midline). By the endline, the average number of questions asked by the treatment group is 1.7, with 9% of the treatment group responding to a question. Further, columns 4 (midline) and 8 (endline) show that the reminder group had used the service almost an hour more on average, but were not statistically more likely to call into the line.

Taken together the results represent substantial induced usage for treatment farmers, although one-fifth of the treatment group did not use the service. Additionally, these average effects also mask important temporal patterns shown in Figure 2.3 which reports average AO use by month. We see that there was substantial usage across treatment arms during the first six months after the intervention was administered. Following this period, usage

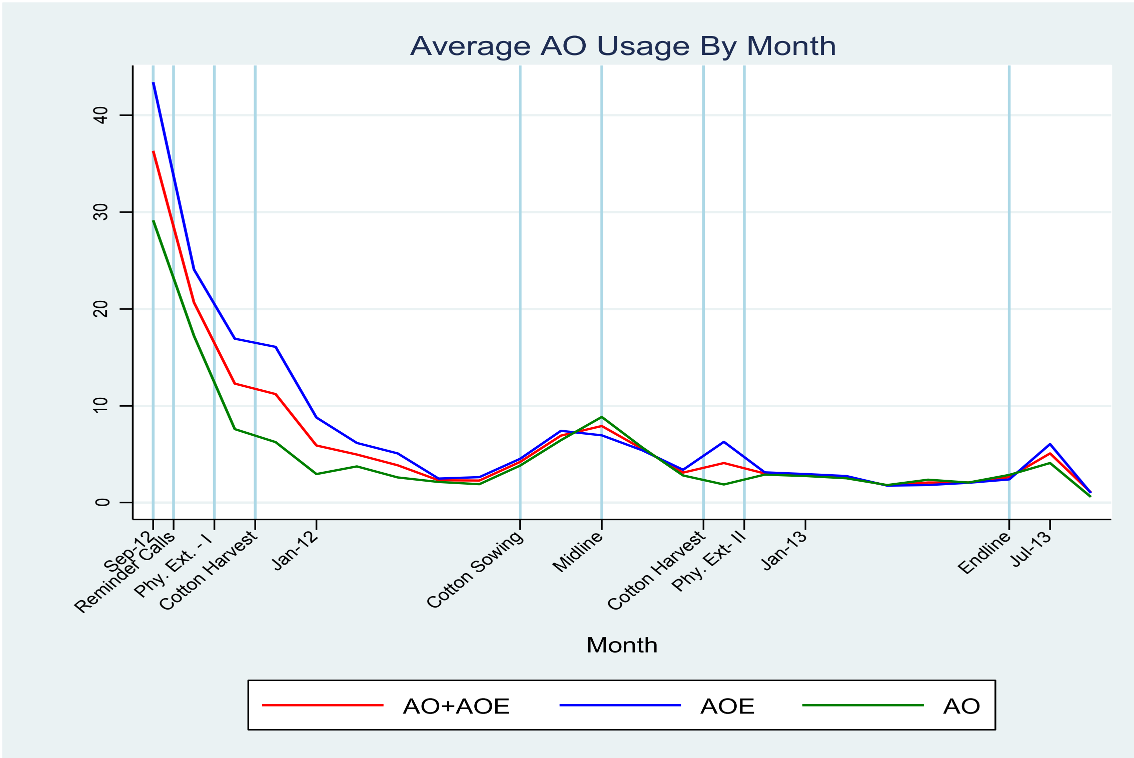
TABLE 2.2: USAGE OF AVAAJ OTALO (AO) INFORMATION SERVICE

Dependent Variable	Control Mean (Midline) (1)	Treat-Control ITT (Midline) (2)	Treat+Reminder-Control ITT (Midline) (3)	Treat+Reminder-Treat ITT (Midline) (4)	Control Mean (Endline) (5)	Treat-Control ITT (Endline) (6)	Treat+Reminder-Control ITT (Endline) (7)	Treat+Reminder-Treat ITT (Endline) (8)
A. Total AO Usage (Incoming)								
Called in to the AO line	0.018	0.622***	0.657***	0.091**	0.073	0.803***	0.818***	0.032
	(0.132)	(0.019)	(0.022)	(0.037)	(0.260)	(0.017)	(0.019)	(0.026)
Total number of calls	0.040	8.401***	9.597***	3.376*	0.221	20.854***	22.668***	4.818
	(0.436)	(1.025)	(1.570)	(1.839)	(1.336)	(2.080)	(2.909)	(3.883)
Total AO usage (minutes)	0.020	80.718***	97.515***	46.879***	0.612	154.539***	174.803***	56.569**
	(0.339)	(9.606)	(14.847)	(17.669)	(4.767)	(16.367)	(23.813)	(28.464)
Avg. call time (minutes)	0.005	4.760***	5.069***	0.683	0.236	4.512***	4.584***	0.065
	(0.054)	(0.266)	(0.341)	(0.595)	(1.457)	(0.224)	(0.254)	(0.484)
Called AO, but did not access any features	0.070	0.244***	0.234***	-0.034	0.023	0.086***	0.080***	-0.008
	(0.256)	(0.021)	(0.024)	(0.035)	(0.149)	(0.013)	(0.016)	(0.024)
Recorded a message on AO	0.000	0.118***	0.129***	0.031	0.015	0.161***	0.165***	-0.004
	(0.000)	(0.012)	(0.015)	(0.024)	(0.122)	(0.015)	(0.018)	(0.028)
Asked a question	0.010	0.304***	0.318***	0.039	0.010	0.390***	0.400***	0.027
	(0.100)	(0.017)	(0.022)	(0.034)	(0.100)	(0.018)	(0.023)	(0.036)
Responded to a question	0.000	0.072***	0.080***	0.021	0.000	0.090***	0.095***	0.018
	(0.000)	(0.009)	(0.012)	(0.019)	(0.000)	(0.010)	(0.013)	(0.021)
B. Push calls (Outgoing)								
Percentage of total push call time listened to	0.000	0.600***	0.619***	0.041**	0.000	0.309***	0.319***	0.021**
	(0.000)	(0.009)	(0.011)	(0.019)	(0.000)	(0.005)	(0.006)	(0.010)
Listened to atleast 10% of total push call time	0.000	0.963***	0.973***	0.021	0.000	0.927***	0.944***	0.042**
	(0.000)	(0.007)	(0.007)	(0.015)	(0.000)	(0.009)	(0.010)	(0.021)
Listened to atleast 50% of total push call time	0.000	0.670***	0.704***	0.076**	0.000	0.053***	0.059***	0.013
	(0.000)	(0.017)	(0.021)	(0.035)	(0.000)	(0.008)	(0.011)	(0.016)
N	398	1200	900	802	398	1200	900	802

Notes

1. This table reports usage statistics collected on the AO server.
2. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
3. `Treat' group refers to the 802 farmers that received access to AO.
4. `Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
5. Column 1 and 5 provides the mean and standard deviation for the control group at the midline and endline respectively.
6. Columns 2-4 and 6-8 report an Intention to Treat (ITT) estimate of the difference in means (and robust standard error) between the treatment groups and control group.
7. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

FIGURE 2.3: AO USAGE BY MONTH



Notes

1. This figure reports average monthly usage of the AO service based on data collected from the AO server.
2. The service was provided to all treatment farmers from September 2011 to July 2013.
3. `AO+AOE' group refers to the 802 farmers that received access to AO. AOE group includes 403 farmers who had access to AO and physical extension. AO group refers to the 399 farmers who only had access to AO.

has been trending down, but with important spikes during sowing times and harvest time. This figure is suggestive of AO users acquiring a stock of knowledge and supplementing thereafter with dynamic information needs throughout the season.

Appendix Table D1 provides a categorization of the questions asked by treatment respondents during the two years of service. (The categories are not mutually exclusive.) Unsurprisingly, columns 3 and 4 show that most questions (50%) relate to cotton, and a majority (54%) focus on pest management and these numbers are relatively stable across both years. Table D1 also reports information on the content of push calls (columns 5-8), which tended to provide more information on cumin and wheat cultivation than incoming questions and were the primary source for weather information.

2.4.2 Impact on Sources of Information for Agricultural Decisions

Panel A of Table 2.3 examines the use of mobile-phone based information in agricultural decision-making, and measured trust (on a scale of 1-10) of information provided by mobile phones. By the endline, treatment farmers are 70 percent more likely to report using mobile phone-based information to make agricultural decisions. The reported level of trust in mobile phone-based information is also dramatically higher in the treatment group: approximately 6.27 points more on a 10-point scale. An index aggregating the importance of mobile-phone based information (analysis of the topics comprising this index follows immediately below) for all subject areas is 1.26 standard deviations higher in the treatment group.

We asked farmers for their most important source of information for a series of agricultural decisions. The survey responses are recorded as free text, without prompting, and coded into categories by our data entry teams. We present results across a variety of subject areas. Panel B of Table 2.3 shows that the treatment group consistently reports using mobile phone-based information across a series of agricultural decisions. By the endline, large effect sizes can be seen in the case of pest management (24.3%) and smaller effects in the case of fertilizer decisions (10%) and crop planning (5.6%).

Other than input-related decisions, mobile-phone information is used increasingly by the treatment group for other topics such as weather (36.8%) and this effect size is somewhat smaller at midline (25%). Importantly, we do not find any effect of our treatment on the use of mobile-phones for price information. The AO service never provided price information. This helps address the concern that social desirability bias may be contributing to our results. Additionally, across virtually all agricultural decisions, we do not observe differences between the combined treatment group and the reminder group.

Appendix D5 provides more disaggregated effects on sources of information. As

TABLE 2.3: EFFECTS ON SOURCES OF AGRICULTURAL INFORMATION

Dependent Variable	Control Mean (Baseline)	Difference-in-Difference Estimates (Treat*Post Coefficient)			
		Treat vs. Control (Midline)	Treat vs. Control (Endline)	Treat+Reminder vs. Control (Midline)	Treat+Reminder vs. Control (Endline)
	(1)	(2)	(3)	(4)	(5)
A. Across all agricultural decisions					
Index of Mobile Phone-Based	0.000	1.829***	1.259***	1.848***	1.236***
Information Usage (standard deviation units)	(0.289)	(0.158)	(0.099)	(0.196)	(0.116)
Used Mobile Phone-Based	0.093	0.616***	0.704***	0.621***	0.698***
Information	(0.291)	(0.031)	(0.030)	(0.032)	(0.032)
Trust in Mobile Phone-Based	0.606	5.353***	6.274***	5.470***	6.310***
Information (on a scale of 1-10)	(2.031)	(0.262)	(0.234)	(0.277)	(0.258)
B. By decision type					
Crop Decision	0.000	0.044***	0.056***	0.038***	0.051***
	(0.000)	(0.008)	(0.010)	(0.009)	(0.011)
Pest Management	0.000	0.093***	0.243***	0.096***	0.258***
	(0.000)	(0.011)	(0.018)	(0.014)	(0.022)
Fertilizer Management	0.003	0.073***	0.100***	0.068***	0.114***
	(0.050)	(0.011)	(0.013)	(0.013)	(0.016)
Weather	0.003	0.234***	0.369***	0.242***	0.362***
	(0.050)	(0.019)	(0.022)	(0.023)	(0.026)
Soil Preparation	0.000	0.047***	0.010	0.032***	0.006
	(0.000)	(0.008)	(0.006)	(0.008)	(0.007)
Prices	0.023	-0.019	0.001	-0.014	0.010
	(0.149)	(0.017)	(0.016)	(0.018)	(0.018)
N	398	2323	2280	1743	1716

Notes

1. This table reports the impact of AO on usage of mobile phone-based information across different agricultural decisions.
2. The indices aggregate information over multiple outcomes for which we expect unidirectional treatment effects. Each index consists of the average of the z-scores for each component of the index, with the control group mean and standard deviation as reference.
3. Mobile phone-based information usage index: Aggregates mobile phone use across crop decision, soil preparation, pest management, weather, cotton pesticides, cotton fertilizers, wheat fertilizers, cumin pesticides and cumin fertilizers.
4. Data on agricultural decision-making and other information sources can be seen in Appendix Table A2.
5.`Treat' group refers to the 802 farmers that received access to AO.
6.`Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
7. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
8. Column 1 provides the mean and standard deviation for the control group at baseline.
9. Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.
10. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

suggested by the index of information sources, we observe across the board increases in the use of mobile phone-based information. The treatment group reports using information from input dealers less often in making pesticide decisions (-7.2% at midline) although, interestingly, they report consulting input dealers *more* often in the case of cotton fertilizer use (5%) and cumin pesticide use (3.7%) at the endline. There are also reported reductions in the use of information from 'other farmers' and 'past experience'. The reduction in reliance on past experience for cumin fertilizers is significant at the midline.

Taken together, these results suggest that AO has been successful in establishing itself as a source of information for treatment respondents in making a variety of important agricultural decisions. These results also suggest that demand exists for agricultural information in rural Gujarat and that this information is not currently being provided via mobile phone. In the next sections we look at whether the provision of information through AO affected input use and agricultural productivity more broadly.[18]

2.4.3 Overall Impact on Input Adoption

A number of input choices influence agricultural productivity. Cotton is the main cash crop grown in our sample – grown by 98.4% of the sample at baseline – and chemical inputs such as pesticides and fertilizers greatly affect cotton yields.[19] In addition, Bt cotton is the dominant variety of cotton grown in this context – although there are literally hundreds of sub-varieties and brands which pose other difficulties – and yields are particularly sensitive to regular irrigation.

Panel A of Table 2.4 shows that total input expenditure is not significantly different between the combined treatment group and the control group at either the midline or the endline. However, we observe that expenditure on irrigation is twice as high for the

[18] Appendix D2 provides even more detail on changes in sources of information. Across a number of agricultural decisions, farmers tend to rely heavily on other farmers, with input shops being particularly important for pesticide decisions. Notably unimportant are government extension services, virtually unmentioned by farmers as a source of information.

[19] In 2006-2007, 87% of all land under cotton in India was treated with pesticide. In contrast, this figure is just 51% for paddy and 12% for wheat. Calculations by author (Agricultural Census of India, 2006).

combined treatment group and the reminder group at the midline (significant at the 1% level). Similarly, by the endline irrigation is 60% higher in the combined treatment group (t-statistic = 1.63) and 80% higher in the reminder group (significant at the 5% level).[20]

Panel B of Table 2.4 shows that the treatment group consistently adopted more cotton-related inputs and practices suggested by the service (0.05-0.07 standard deviation units). These input decisions include recommended seed varieties, pesticides, fertilizers and irrigation practices. While the treatment effects on the overall wheat and cumin indices are not significantly different from zero, the point estimates are qualitatively consistent.[21]

2.4.4 Impact on Seed Selection

The presence of a wide variety of cotton seeds, some counterfeit, makes seed selection a particularly important decision in our context. In Panel C of Table 2.4, we observe that the index of cotton seed-related decisions is consistently higher (0.046-0.050 standard deviation units) in the combined treatment group and the reminder group at midline and is statistically significant at the 10% level

This result may be driven by the purchase of Ganga Kaveri, with treatment farmers purchasing 0.08 kg more of this brand relative to control. An inventory analysis conducted in Sayla and Chotila following the conclusion of the study found that this is one of the four most commonly stocked varieties along with Vikram, Rasi and Ajit.

2.4.5 Pest Management Practices

In Panel D of Table 2.4 we examine the treatment effect on pest management practices. The index which includes all pest management practices is 0.08 standard deviation units

[20] Panel B of Appendix Table D6 reports a detailed breakdown of changes in input costs. In addition to changes in irrigation costs, we observe changes in expenditure on seeds, but these changes are not significant at traditional levels (t-statistic = 1.4).

[21] The standard errors also suggest that the experiment may be underpowered to detect effects for cumin (grown by just 34% of the sample), while wheat cultivation involves substantially fewer chemical inputs and is primarily produced for home consumption.

TABLE 2.4: EFFECTS ON INPUT ADOPTION DECISIONS

Dependent Variable	Control Mean (Baseline) (1)	Difference-in-Difference Estimates (Treat*Post Coefficient) Treat vs. Control (Midline) (2)	Treat vs. Control (Endline) (3)	Treat+Reminder vs. Control (Midline) (4)	Treat+Reminder vs. Control (Endline) (5)
A. Expenditure on Inputs					
Total Input Expenditure	9.682	-0.012	0.082	-0.028	0.266
(log rupees)	(0.766)	(0.164)	(0.210)	(0.181)	(0.223)
Expenditure on Irrigation	4.821	1.009***	0.605	1.118***	0.817**
(log rupees)	(4.469)	(0.376)	(0.369)	(0.414)	(0.404)
B. Index of All Input-Related Decisions (standard deviation units)					
Cotton	0.000	0.050*	0.061**	0.056*	0.074**
	(0.289)	(0.027)	(0.029)	(0.033)	(0.034)
Wheat	0.000	0.098	0.038	0.081	0.056
	(0.433)	(0.070)	(0.037)	(0.076)	(0.041)
Cumin	0.000	0.037	0.064	0.034	0.048
	(0.347)	(0.042)	(0.043)	(0.047)	(0.054)
C. Index of Seed-Related Decisions (standard deviation units)					
Cotton	0.000	0.046*	0.454	0.051*	0.688
	(0.296)	(0.027)	(0.280)	(0.031)	(0.432)
Wheat	0.000	-0.033	-0.032	-0.020	0.017
	(0.505)	(0.070)	(0.071)	(0.066)	(0.066)
Cumin	0.000	-0.020	0.012	-0.018	0.015
	(0.792)	(0.072)	(0.075)	(0.084)	(0.087)
D. Indices of Pesticide-Related Decisions (standard deviation units)					
Cotton	0.000	0.025	0.062	0.034	0.081*
	(0.361)	(0.035)	(0.039)	(0.039)	(0.045)
Cumin	0.000	0.040	0.054	0.023	0.002
	(0.437)	(0.043)	(0.045)	(0.047)	(0.047)
E. Indices of Fertilizer-Related Decisions (standard deviation units)					
Cotton	0.000	0.090*	0.074*	0.093	0.086
	(0.457)	(0.047)	(0.044)	(0.058)	(0.052)
Wheat	0.000	0.087	0.028	0.088	0.036
	(0.553)	(0.071)	(0.046)	(0.078)	(0.051)
Cumin	0.001	0.005	0.072	0.025	0.109
	(0.524)	(0.057)	(0.080)	(0.062)	(0.112)
N	398	2323	2280	1743	1716

Notes

1. This table reports the impact of AO on input decisions for seeds, pesticides and fertilizers.
2. All coefficient estimates are in standard deviation units.
3. The indices aggregate information over multiple outcomes for which we expect unidirectional treatment effects. Each index consists of the average of the z-scores for each component of the index, with the control group mean and standard deviation as reference.
4. Input expenditure includes total money spent on seeds, fertilizers, irrigation and pesticides for all crops in a year.
5. All Input index: seed usage + pesticide purchase + pesticide usage + pesticide quantities + pesticide expenditure + fertilizer purchase + fertilizer usage + fertilizer quantities + fertilizer expenditure for the three different crops – cotton, wheat and cumin + irrigation expenditure
6. Pesticide management index: dummy to indicate purchase/use of a pesticide + pesticide expenditure + pesticide quantities across purchase and usage decisions.
7. Fertilizer management index: dummy to indicate purchase/use of a fertilizer + fertilizer expenditure + fertilizer quantities across purchase and usage decisions.
8. Seed management index: dummy to indicate purchase/use of recommended seeds + seed expenditure + seed quantities across purchase and usage decisions.
9. `Treat' group refers to the 802 farmers that received access to AO.
10. `Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
11. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
12. Column 1 provides the mean and standard deviation for the control group at baseline.
13. Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.
14. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

higher in the reminder group at the endline. This effect was not significant for the combined treatment group but all estimated coefficients move in the same direction.

Examining the sub-components of the index (see Appendix Table D7) , there are no statistically significant results for pesticide purchase and usage for the treatment group, once again in contrast to the simple difference estimates 7 months after the intervention has been administered in Cole and Fernando, 2014.[22]

We do observe a 2.4% increase (2.28% in the midline) in the fraction of treatment respondents using tricoderma, a biological method of pest control, relative to the control group. The AO service provided extensive information in both Kharif and Rabi on the use of Tricoderma, as a means of preventing wilt disease in cotton and cumin.

2.4.6 Fertilizers

In Panel E of Table 2.4 we examine fertilizer practices. The index of cotton fertilizer practices is 0.07 standard deviation units higher among the combined treatment group in the endline, as compared to 0.09 standard deviation units in the midline. The index is similarly higher for the reminder group at both the midline and the endline but the point estimate are not statistically significant.

Dis-aggregating these results, we see a 5% (5.5% in the midline) increase among the treatment group in purchase of ammonium sulfate- a fertilizer we focused on for cotton push call content. Additionally, the service lead to a 5% decrease (2% decrease, not significant in the midline) in those purchasing di-ammonium phosphate among the treatment group. Lastly, we see a 28 rupees decrease in the amount spent on Murate of Potash (significant at the 5 % level) although we did not push any content on this bio-fertilizer through AO.[23]

[22] While total money spent on acetamaprid increases, this number is only significant for the AOE group (an increase of Rs. 80, not reported). Similarly, while total spent on monocrotophos decreases, the only statistically significant result is among the AOE treatment group (a decrease of Rs. 60, not reported).

[23] Appendix Table D7 shows the breakdown of results by input type.

TABLE 2.5: EFFECTS ON SOWING DECISIONS AND AGRICULTURAL PRODUCTIVITY

Dependent Variable	Control Mean (Baseline) (1)	Difference-in-Difference Estimates (Treat*Post Coefficient) Treat vs. Control (Midline) (2)	Treat vs. Control (Endline) (3)	Treat+Reminder vs. Control (Midline) (4)	Treat+Reminder vs. Control (Endline) (5)
A. Sowing Decisions					
Planted Cotton	0.985	0.005	-0.002	0.001	0.001
	(0.122)	(0.011)	(0.016)	(0.012)	(0.017)
Planted Wheat	0.776	0.047	0.069*	0.023	0.061
	(0.417)	(0.038)	(0.039)	(0.042)	(0.042)
Planted Cumin	0.425	-0.015	-0.001	-0.007	0.004
	(0.495)	(0.039)	(0.040)	(0.043)	(0.044)
B. Agricultural Outcomes (Expenditure, Profit, Yield)					
Profit From Agriculture	11.466	0.085	0.086	0.085	0.168
(log rupees)	(1.015)	(0.082)	(0.124)	(0.089)	(0.131)
Cotton Yield	694.819	31.554	19.649	59.935*	44.716
(kg/acre)	(468.752)	(33.844)	(33.480)	(36.005)	(35.466)
Wheat Yield	981.132	-33.693	-32.610	-49.903	-28.317
(kg/acre)	(702.002)	(77.551)	(71.201)	(84.578)	(76.883)
Cumin Yield	172.570	-7.169	48.295**	0.074	54.270**
(kg/acre)	(191.017)	(23.731)	(23.386)	(26.822)	(25.947)
N	398	2323	2280	1743	1716

Notes

1. This table reports the impact of AO on sowing decisions and agricultural outcomes.
2. Total input expenditure refers to money spent on seeds, fertilizers, pesticides and irrigation in the past year.
3. Profit from agriculture refers to the difference between total income from of all crops grown and total input expenditure (on seeds, pesticides, irrigation and fertilizer) in the past year.
4. `Treat' group refers to the 802 farmers that received access to AO.
5. `Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
6. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
7. Column 1 provides the mean and standard deviation for the control group at baseline.

8 Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.

9. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

2.4.7 Sowing and Productivity

In Table 2.5 we examine sowing choices and agricultural productivity. We do not observe any effect of the treatments on the frequency of cultivation or area planted of cotton, cumin or wheat.

Panel B shows that cotton yields are consistently higher for the treatment group and the reminder group at the midline and the endline. However, this effect is only significant for the reminder group at the midline (increase of 60 kg per acre, or 3.5% higher than the control group). Additionally, we see that yield for cumin is about 48 kilograms per acre higher (7 kg/acre lower in the midline) among the treatment group and 54 kilograms per acre higher for the reminder group (same at midline) and statistically significant at the five-percent level. These results are robust to winsorizing (p=0.25).

As in the case of agricultural yields, the detection of treatment effects on profits is greatly complicated by measurement error. At the endline, both the treatment group and the reminder group have profits that are more than $200 higher than the control group (16% higher), although both these effects are imprecisely estimated. In addition, we see an 8% increase in input expenditure by the endline for the combined treatment group (26% higher for the reminder group) but this effect is also imprecisely estimated. Measuring in levels rather than logs, we find that input expenditure is higher for the reminder group at endline by roughly $50, significant at the 10% level (not reported).

2.4.8 Impact on Agricultural Knowledge

It is important to understand the mechanisms by which AO works: does it serve as an education tool, creating durable improvements in knowledge, or does it function as an advisory service, in which farmers follow instructions, without necessarily comprehending why a particular course of action is the right one? In Table 2.6, we examine whether AO improves farmers' ability to answer basic agricultural questions. The questions we ask test the respondents on a wide range of topics, which are generally invariant to their personal circumstances.[24]

Baseline agricultural knowledge is low, with farmers answering only 29% of questions correctly. There are no significant differences between treatment and control for the total at the baseline. Given that these are very basic questions about agriculture, this suggests that there is a substantial lack of information on even basic topics concerning crop cultivation.

As reported in Table **??**, we do not observe differences between the treatment and control groups in agricultural knowledge in the midline or in the endline survey. In part, the types of knowledge that respondents gain reflect their actual demand for information. The

[24] The full text of the questions is available in Appendix D8.

TABLE 2.6: EFFECT ON AGRICULTURAL KNOWLEDGE

Dependent Variable	Control Mean (Baseline) (1)	Difference-in-Difference Estimates (Treat*Post Coefficient) Treat vs. Control (Midline) (2)	Treat vs. Control (Endline) (3)	Treat+Reminder vs. Control (Midline) (4)	Treat+Reminder vs. Control (Endline) (5)
All Questions	14.156	0.492	0.760	0.525	0.619
(44 questions)	(5.279)	(0.438)	(0.496)	(0.480)	(0.547)
Cotton-related	4.847	0.146	0.304	0.290	0.368
(20 questions)	(1.989)	(0.217)	(0.234)	(0.239)	(0.257)
Wheat-related	3.419	0.108	0.063	0.081	-0.043
(7 questions)	(1.629)	(0.180)	(0.149)	(0.199)	(0.164)
Cumin-related	5.164	0.098	-0.019	0.169	-0.078
(9 questions)	(1.791)	(0.241)	(0.212)	(0.263)	(0.235)
Pesticide-related	0.887	0.018	0.048	0.074	0.064
(8 questions)	(0.717)	(0.106)	(0.107)	(0.115)	(0.117)
Fertilizer-related	0.606	0.055	0.072	0.042	0.081
(3 questions)	(0.656)	(0.059)	(0.065)	(0.065)	(0.072)
N	398	2323	2280	1743	1716

Notes

1. This table reports the effect of AO on respondents' agricultural knowledge.
2. Respondents were asked agricultural questions across crop and topic, and a knowledge score was computed based on the proportion of correct answers. The question categories are not mutually exclusive.
3. Treat' group refers to the 802 farmers that received access to AO.
4. Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
5. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
6. Column 1 provides the mean and standard deviation for the control group baseline.
7. Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.
8. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

majority of questions asked on the AO platform relate to pesticides.

2.4.9 Heterogeneous Treatment Effects

While the importance of technological progress to growth is beyond doubt, there are growing concerns about the possibility of a "digital divide," in which the poorest or least educated are less able to take full advantage of the promise of new technologies. We test this hypothesis by comparing AO usage and knowledge gain by education level. We focus on respondent education for at least two reasons: first, while the service is designed to be accessible to illiterate users, it may be easier to use or navigate for a literate population, who can take advantage of instructional material. Second, educated individuals may be in a better position to learn. (We also examined landholdings as a source of heterogeneity in treatment effects, and found virtually no difference between above- and below-median landholders.) The median farmer in our survey reports 4 years of education.

Are AO Usage and Education Complements?

In Table 2.7, we regress measures of AO usage on a treatment dummy, a dummy for having more than the median number of years of formal education (4 years), a time-trend dummy and the corresponding interaction terms as in equation (2.3).

Columns (2) and (3) suggest there may be some complementarities between AO use and education: more educated farmers make more use of the service on average but these differences are not statistically significant. We do not find an effect on the extensive margin; that is, more educated individuals are no more likely to call into the AO line. This table makes use of administrative data for all 1,200 respondents as their calls (and the absence of calls from control) are logged on to the server. We do not observe heterogeneous effects of AO across education for input adoption or agricultural knowledge.

TABLE 2.7: HETEROGENEOUS EFFECTS BY EDUCATION AND INCOME

Dependent Variable	Control Mean	Triple Difference Estimates			
		Education		Income	
		Treat*Post*Educ (Midline)	Treat*Post*Educ (Endline)	Treat*Post*Inc (Midline)	Treat*Post*Inc (Endline)
	(1)	(2)	(3)	(4)	(5)
A. AO Usage					
Called AO line	0.000	0.025	-0.055	0.063*	-0.025
	(0.000)	(0.042)	(0.037)	(0.038)	(0.035)
Total Incoming AO Usage (Minutes)	0.000	22.950	27.042	40.204**	46.656
	(0.000)	(21.194)	(38.043)	(17.319)	(30.028)
B. Indices of Input-related Practices (standard deviation units)					
Cotton Management Practices	0.000	0.003	0.037	0.059	0.104*
	(0.289)	(0.052)	(0.055)	(0.052)	(0.055)
Wheat Management Practices	0.000	-0.441***	0.022	0.237	0.020
	(0.433)	(0.144)	(0.065)	(0.144)	(0.069)
Cumin Management Practices	0.000	-0.233***	-0.052	0.040	0.057
	(0.347)	(0.081)	(0.089)	(0.084)	(0.081)
Pesticide Management Practices	0.000	-0.009	-0.057	0.004	0.082
	(0.303)	(0.056)	(0.062)	(0.056)	(0.061)
Fertilizer Management Practices	0.000	-0.242***	0.054	0.137	0.093
	(0.306)	(0.086)	(0.091)	(0.087)	(0.078)
C. Agricultural Knowledge Score					
Total Correct Answers (44 questions)	14.156	0.210	1.256	-0.371	0.770
	(5.279)	(0.772)	(0.913)	(0.848)	(0.982)
N	398	2323	2280	2323	2280

Notes

1. This table tests for heterogeneity in the treatment effect across education and income level.

2. Management practices indices: seed usage + pesticide purchase + pesticide usage + pesticide quantities + pesticide expenditure + fertilizer purchase + fertilizer usage + fertilizer quantities + fertilizer prices for the three different crops – cotton, wheat and cumin.

3. Pesticide management index: dummy to indicate purchase/use of a pesticide + pesticide expenditure + pesticide quantities across purchase and usage decisions.

4. Fertilizer management index: dummy to indicate purchase/use of a fertilizer + fertilizer expenditure + fertilizer quantities across purchase and usage decisions.

5. Seed management index: dummy to indicate purchase/use of recommended seeds + seed expenditure + seed quantities across purchase and usage decisions.

6. Education and income measures are all collected during the baseline survey which took place in June 2011. Income refers to agricultural income for the past 12 months.

7. 'Treat' group refers to the 802 farmers that received access to AO.

8. The estimates are from the endline survey that took place between 23rd July and 30th August 2013.

9. Column 1 provides the mean and standard deviation for the control group at baseline.

10. Columns 2 and 3 report the triple interactions from a triple difference specification with interactions with above-median education, columns 4 and 5 report the same for income. Columns 2 and 4 are midline comparisons while columns 3 and 5 are endline comparisons.

11. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Income and AO

Treatment respondents with above median incomes are 6% more likely to call into the AO line at the midline, although by the endline there is no significant difference. Mean usage for farmers with above median incomes is also higher by approximately 40 minutes (46 minutes in the endline). Farmers with higher incomes also show differential effects in the cotton practices index (about 0.1 standard deviation units higher, 0.06 at midline but not significant).

2.4.10 Peer Effects

Given randomization at the household-level it is possible that access to AO indirectly influenced the outcomes of people not a part of the study through information spillovers. For instance, treatment respondents may have discussed advice they received or even asked questions on behalf of their peers. Alternatively, peers may follow suit after directly observing changes in agricultural practices of their neighbors.

Table 2.8 estimates whether access to AO influenced the outcomes of individuals in the social networks of the treatment group.[25] Columns (1) and (2) refer to individuals in the treatment group, allowing us to estimate whether there are complementarities between peers assigned to AO. Columns (3) and (4) refer to non-study respondents. In each case, we estimate whether the fraction of peers assigned to treatment influenced respondent outcomes as in equation (3.3).

In Panel A we see that treatment respondents with peers who are also assigned to the treatment group use AO even more, although this estimate is not statistically significant. In Panel B, we see that having more treated peers in a social network also increases the likelihood of reporting 'NGO' as a source of information for agricultural decision-making. In the overwhelming majority of cases the name of the NGO is our field partner. This effect

[25] Appendix D9 assesses whether the fraction of treated peers in a social network is independent of other observable characteristics. The only characteristic that shows an imbalance is cotton acreage. We control for baseline cotton acreage in all peer regressions.

TABLE 2.8: PEER EFFECTS ON AO USAGE, SOURCES OF INFORMATION AND SOWING

Dependent Variable	Treatment Group		Non-Study Peers	
	Control Peer Group	Fraction of Peers Treated *Post	Control Peer Group	Fraction of Peers Treated
	(Baseline)	(Midline)	(Baseline)	(Midline)
	(1)	(2)	(3)	(4)
A. AO Usage				
Called AO line	0.000	-0.061	-	-
	(0.000)	(0.067)		
Total AO usage (Minutes)	0.000	18.572	-	-
	(0.000)	(22.722)		
B. Sources of Information				
Past Experience	0.998	0.025	0.217	0.012
	(0.039)	(0.020)	(0.412)	(0.031)
Input Dealerships	0.850	-0.066	0.383	0.004
	(0.357)	(0.089)	(0.487)	(0.034)
Cell Phone	0.099	-0.097	0.040	-0.001
	(0.299)	(0.069)	(0.197)	(0.009)
NGO	0.272	0.190***	0.024	0.021
	(0.445)	(0.072)	(0.153)	(0.014)
C. Sowing Decisions				
Planted Cumin	0.422	0.151*	0.237	0.059*
	(0.494)	(0.090)	(0.425)	(0.030)
Area of Cumin Planted (Acres)	0.792	0.454	0.525	0.255*
	(1.499)	(0.305)	(1.695)	(0.133)
Planted Wheat	0.722	0.029	0.253	-0.011
	(0.448)	(0.092)	(0.435)	(0.031)
Area of Wheat Planted (Acres)	1.181	-0.086	0.328	-0.050
	(1.816)	(0.222)	(1.153)	(0.077)
Proportion of Cotton Lost to Pest Attacks (%)	-	-	0.142	-0.039***
			(0.224)	(0.015)
N	654	1604	545	1114

Notes

1. This tables assesses whether the fraction of one's peers assigned to the treatment group is independent of observable characteristics preceeding the treatment.

2. The midline survey took place between 4th June and 8th July 2012.

3. Column 1 reports the mean and standard deviation for all treated respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment.

4. Column 2 report the coefficient on the interaction of the number of peers who were assigned to the treatment group and a dummy for whether the observation is from the endline. The regression specification includes both of these variables separately and their interaction.

5. Column 3 reports the mean and standard deviation for peers who were not respondents in the main study and who were

6. Column 4 reports the coefficient on the number of peers who were assigned to the treatment group, from a regression of

7. The regresson specifications in column 2 & 4 include dummies for the number of peers referenced, the amount of cotton grown at baseline, and village fixed effects.

8. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

is suggestive of complementarities between multiple treated peers in the same network acting collectively to obtain advice from our NGO partner. In Panel C, we also observe that treated individuals with more treated peers are more likely to plant cumin.

Non-study respondents with more treated peers are more likely to grow cumin (6%) and plant a large amount of it (.25 acres more). Those with more treated peers in their networks also report 4% less cotton crop loss as a result of pest attacks, suggesting that pest management practices provided by the AO service may have been shared.

2.4.11 Willingness to Pay

After the endline survey we conducted a series of experiments to assess willingness to pay for the AO service among the original 1200 study respondents and additional 457 non-study respondents. The first method we used were 'Take it or Leave it' (TIOLI) offers which randomized the price of a nine-month subscription to AO. [26] The second method used the Becker-DeGroot-Marschak (BDM) method as an incentive compatible price elicitation mechanism. In this method the respondent first indicates their willingness to purchase at a series of price points. They then record a specific bid, after which the respondent is shown a randomly generated offer price.[27] If the respondent's bid is greater than the offer price they can buy it at the offer price and if not they cannot purchase the product. The TIOLI method was randomized to a quarter of the sample, while the BDM method was randomized to the remaining three-quarters.

The two methods of eliciting willingness to pay deliver similar results. Of the 390 respondents that were offered AO through the TIOLI method, 150 respondents (38.4%) bought a subscription at an average price of Rs. 107 ($1.78). Similarly, of the 1043 respondents that were offered AO through the BDM method, 370 (33%) purchased a subscription at an average price of Rs.108 ($1.8).

---

[26] The prices offered were Rs. 40 ($0.67), Rs. 90 ($1.5), Rs. 140 ($ 2.3), Rs. 190 ($3.2) and Rs. 240 ($4).

[27] The respondent is asked to indicate their willingness to purchase the policy for Rs. 40 ($0.67), Rs. 90 ($1.5), Rs. 140 ($ 2.3), Rs. 190 ($3.2), Rs. 240 ($4), Rs. 290 ($4.8), Rs. 390 ($6.5), Rs.490 ($8.1)

Table 2.9 investigates correlates of the decision to purchase AO. Surprisingly, we do not find that treatment status is an important predictor of purchasing AO. Rather we find that education positively predicts the decision to purchase AO while the offer price does the opposite.

Figure 2.4 shows the elicited demand curves for AO for both methods. The methods yield comparable estimates of willingness to pay which we estimate at Rs. 108 ($1.78) for a nine month subscription. AO costs little, requiring just $0.83 to service one farmer per month, inclusive of airtime costs, staff time and technology fees. In contrast, a single round of traditional extension (educational demonstration by a government extension worker to a gathering of farmers) $ 8.5 per farmer (based on extension provided to the AOE group).

In our study, airtime was provided freely for farmers to encourage take-up (costing approximately $ 0.31). If farmers paid airtime, the per-farmer operating cost of the AO service could be as low as $0.52 per month. However, even at this rate AO would require a subsidy of roughly $0.35 per month per farmer given the elicited willingness to pay. It is important to note that the per-farmer cost of providing AO is likely to drop considerably as the service scales up, as labor costs need not scale linearly if pre-recorded answers can be directed to commonly asked questions.

2.4.12 Cost-Benefit Analysis

To compute the return to investing in an AO subscription we weigh measured increases in yield against increases inputs costs. A 3.5% increase cotton yields for the treatment group implies an average revenue increase of nearly $200 while a 26.3% increase in cumin yields implies an average return of $65.[28] This $265 average increase in revenue must be weighed against an increase in input costs of $50.[29] This implies a profit of $215 on the basis of a $20,

[28]These calculations are based on average values of crop acreage and crop selling prices for the entire sample. On average, respondents grew 4.4 acres of cotton (0.55 acres of cumin) and sold cotton at a price of $0.74 per kg ($2.18 per kg for cumin) at the time of the endline survey. We observe an increase of 60 kg per acre in cotton yields and 54 kg per acre for cumin yields for the treatment group.

[29]Input costs include the costs of seeds, fertilizers, pesticides, hired and household labor. Household labor is priced at the mean of the hired wage. This effect is precisely estimated for the reminder group at the endline.

TABLE 2.9: WILLINGNESS TO PAY FOR AVAAZ OTALO (AO) INFORMATION SERVICE

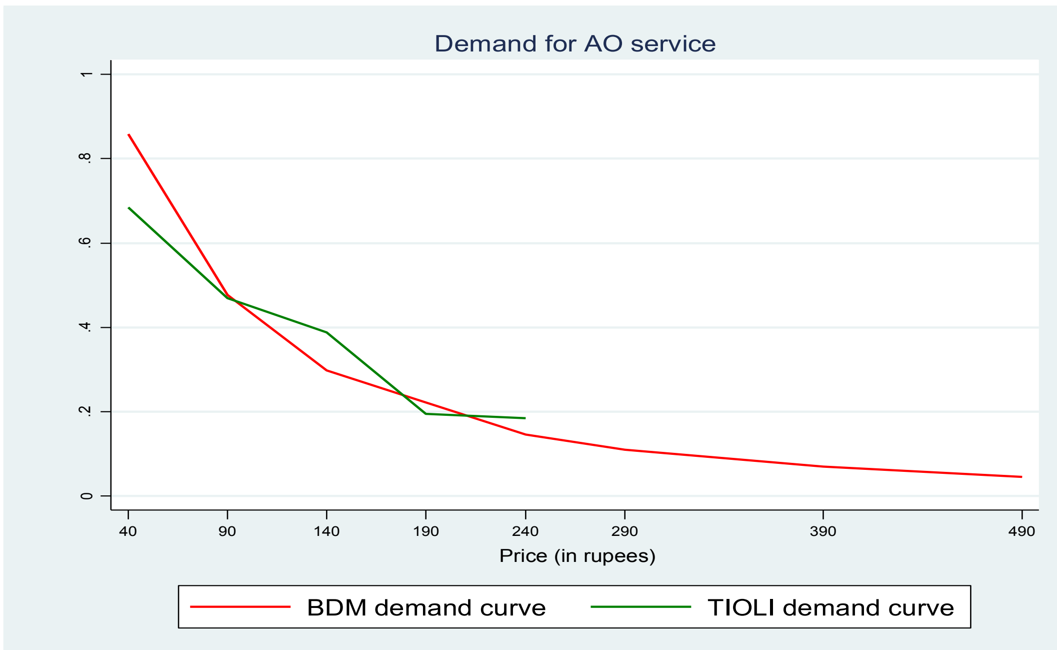
Dependent Variable	Entire Sample (1)	Study (2)	Treatment (3)	Control (4)	Non-Study (5)
Mean willingness to pay (Rs.)	108.993	121.246	123.858	115.307	93.282
N	1043	586	407	179	457

Dependent Variable	Bought AO (1)	Bought AO (2)	Bought AO (3)	Bought AO (4)	Bought AO (5)	Bought AO (6)	Bought AO (7)	Bought AO (8)
Offer Price	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Treatment status	-0.015	-0.016	-0.016	-0.015	-0.015	-0.015	-0.015	-0.016
	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Total duration of calling in time	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Education		0.009***						0.008***
		(0.002)						(0.002)
Total Correct Answers to Knowledge Questions			0.002					0.001
			(0.002)					(0.002)
Total Area of Cotton Planted (Acres)				0.001				0.001
				(0.002)				(0.002)
Referrals in the Treatment Group					-0.003			-0.002
					(0.020)			(0.019)
Skepticism toward Technology						-0.020		-0.015
						(0.020)		(0.020)
Age of Household Head							-0.001	-0.001*
							(0.001)	(0.001)
N	7885	7885	7885	7885	7885	7885	7885	7885
N(Clusters)	1200	1200	1200	1200	1200	1200	1200	1200

Notes

1. This table reports results from the willingness to pay study that was carried out between 23rd July and 30th August 2013 along with the endline paper survey.
2. Respondents took part in two types of willingness to pay exercises – 75% participated in bidding game based on the Becker-Degroot-Marschak (BDM) method and 25% participated in a simpler take-it-or-leave-it (TIOLI) exercise.
3. Table 9A reports the mean maximum willingness to pay for the AO service by treatment status for all those respondents that participated in the BDM exercise.
4. Column 1 refers to the entire sample include study farmers (both with and without access to AO) as well as their farmer friends (peers) who were not part of the sample. Columns 2, 3 and 4 refer to only the study sample, only farmers with access to AO and farmers without access to AO, respectively. Column 5 refers to peer farmers i.e. friends who discuss
5. Table 9B reports predictors of willingness to pay for the service based on characteristics from the baseline survey and the bids placed during the willingness to pay exercise. This table includes only those farmers that were part of the treatment or control group.
6. Referrals in the treatment group refers to those respondents who referenced peers - a maximum of 3 were elicited - who were themselves assigned to treatment.
7. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

FIGURE 2.4: DEMAND CURVE FROM WILLINGNESS TO PAY EXPERIMENTS



Notes

1. This figure looks at the demand for the AO service as determined by a willingness to pay exercise that was carried out between 23rd July and 30th August 2013 along with the endline paper survey.

2. Respondents took part in two types of willingness to pay exercises – 75% participated in bidding game based on the Becker-Degroot-Marschak (BDM) method and 25% participated in a simpler take-it-or-leave-it (TIOLI) exercise.

3. Respondents who participated in the BDM exercise were offered the service at decreasing price points ranging from Rs. 490 to Rs. 40.

4. Respondents participating in TIOLI were offered the chance to buy the service at a particular offer price ranging from Rs. 40 to 240.

2-year subscription to AO. This implies a return of more than $10 for each dollar invested in AO, net of cost.

2.5 Threats to Validity

2.5.1 Attrition

In the endline survey, we had 120 attritees, of which 39 were control farmers, 43 from the AOE group and 38 from the AO group. In comparison, we had 77 attritees in the midline, of which 23 were control farmers, 22 were from the AOE group and 32 were from the AO group. We do not observe any significant differences between the treatment and control group for the attritees, as measured by baseline characteristics. These results are reported in Appendix Table D10.

2.5.2 Experimenter Demand Effects

A second obvious concern is that respondents in the treatment group may offer answers that they believe the research team seeks, perhaps in the hopes of prolonging the research project, or due to a sense of reciprocity. While it is difficult to rule this out entirely, the fact that we find no effect on sources of price information in Table 2.3 – which the AO service does not provide – in spite of finding large differences for sources of other information provides some comfort. We also note that we can observe some outcomes perfectly: the AO platform records precisely how many times respondents call in. Respondents provide remarkably unbiased answers to the question "did you call into the AO line with a question," with 55.5% self-reported call-in rate vs. a 53.5% call-in rate using administrative data (results not reported in tables).

The values at midline and for the combined treatment group imply a smaller increase in input costs but are not precisely estimated.

2.6 Conclusion

This paper presents the results from a randomized experiment studying the impact of providing toll-free access to AO, a mobile phone-based technology that allows farmers to receive timely agricultural information from expert agronomists and their peers.

Firstly, we show that the intervention was successful in generating a substantial amount of AO usage, with roughly 60% of the treatment group calling into listen to content or ask a question within 7 months of beginning the intervention, and 80% using it after two years. We then showed that AO had a large impact on reported sources of information used in agricultural decisions, reducing the reliance of treatment respondents on input dealers and past experience for advice.

Having established AO as a reliable source of information, we then show that advice provided through AO resulted in farmers changing a wide variety of input decisions that ultimately lead to increases in crop yields. In addition, we find evidence that treated respondents had a limited influence on the information sources and cropping decisions of peers not in the study. Richer respondents are more likely to use AO and adopt inputs, suggesting that richer farmers may be differentially well-positioned to take advantage of technological change.

We estimate that a $1 investment in AO generates a return of more than $10. Elicited willingness to pay for a $7.5 subscription is only $1.7, but implied subsidy is more than justified by the returns generated by AO. A two-year subscription generates a profit of more than $200 on average. In addition, while the cost of this intervention is quite low: we estimate a monthly cost of approximately USD $0.83 per farmer (including all airtime costs, staff time, and technology fees) if the project were implemented at scale, the costs may drop dramatically, as pre-recorded answers to specific questions dramatically reduce the amount of time the agronomists must spend on each question. In contrast, the "all-in" costs for physical extension were about $8.50 per farmer. In addition to this high cost, we do not find any evidence to suggest that outcomes between respondents provided with AO and physical extension and those only provided with AO were different.

These results represent the beginning of a research agenda seeking to understand the importance of information and management in small farmer agriculture. Many important questions remain unanswered. Going forward, the individual nature of delivery and information access (each farmer can potentially receive a different push call message, and each can choose which other reported experiences to listen to) will allow us to test the importance of top-down vs. bottom-up information.

One of the features of the current intervention is that the NGO providing the service, DSC, has established trust by providing services to farmers for many years. While certain aspects of observed input adoption like pesticide use allow for sequential learning, for large investments where the downside risk could be potentially devastating, as in the case of cumin sowing, trust would appear to be a lot more important. AO comes across as a service without a vested interest (impartial) in addition to being experts, which may well serve to both encourage farmers to switch away from other sources and act on AO information. We hope to experimentally vary the source of information (if only to present it as a peer instead of an expert) in order to understand the importance of this aspect for technology adoption.

To understand the exact mechanism through which AO affects behavior, it is also important to understand whether the treatment effect is working through acquired knowledge or "merely" persuasion. One definition of cognitive persuasion that has been adopted in the literature is that it consists of "tapping into already prevailing mental models and beliefs" through associations rather than teaching or inculcating the subject with new information. From qualitative work we have conducted, many farmers claim to distrust input dealerships but still adopt their advice for lack of a better source. While this is not something that is emphasized in the AO service itself, the presentation of information that seems to conflict with the advice given by input merchants may well serve to reinforce this distrust. We hope to be able to test these hypotheses using pre- and post- subjective evaluations of the trustworthiness of information sources. However, a more elaborate treatment play may be necessary to clearly distinguish between the two models of how information affects behavior.

Finally, we stress the practical importance of this technology. Climate change and the mono-cropping of new varieties of cotton may significantly alter both the types and frequency of pests, and the effectiveness of pesticides in the near future. Farmers in isolated rural areas have little recourse to scientific information that might allow them to adapt to these contingencies. We believe mobile phone-based agricultural extension presents a cost-effective and salient conduit through which to relay such information.

Chapter 3

Social Interactions, Technology Adoption and Information Exchange: Evidence from a Field Experiment

3.1 Introduction

The diffusion of information plays a critical role in facilitating the adoption of modern technologies. Differences in technology adoption, in turn, are thought in part to account for the dramatic differences in the productivity of farms and firms within countries (Banerjee and Duflo, 2005). In remote areas of the developing world, individuals may become aware of new technologies through a number of different channels including government programs, non-profit organizations and market-oriented advisory services. Yet, given limited state-capacity and noted problems in establishing markets for information (Samuelson, 1954; Arrow, 1969) individuals in rural areas may instead learn valuable information about technologies through social interactions with their peers. This may occur through direct observation of production processes or through the social exchange of information. In an agricultural setting, prior literature has shown that such 'social learning' can result in the adoption of profitable technologies (Foster and Rosenzweig, 1995 ; Conley and Udry, 2010).

In spite of these estimated gains, however, other studies suggest that information may not flow freely through network and frictions may inhibit beneficial exchanges of information (Conley and Udry, 2001 ; Duflo *et al.*, 2008).

This paper asks how modern information and communication technologies (ICTs) influence the structure of social interactions and potential for social learning in a rural agricultural context. Specifically, mobile phone-based agricultural extension systems are increasingly viewed as a cost-effective conduit to deliver information to remote areas of the developing world (Aker, 2011). Such programs have been shown to dramatically alter farmers' sources of information and induce technology adoption (Cole and Fernando, 2015). Through providing a reliable conduit for information, such programs may act as a substitute for advice from other farmers and in so doing crowd out the need for traditional social interactions. On the other hand, in a model of multiple equilibria where previously there was no valuable information to be gained from farmers interacting with each other, the provision of valuable information may increase the gains to exchanges of information and induce social interactions (Duflo, 2006).

This paper estimates the effect of a mobile phone-based agricultural extension service, Avaaz Otalo (AO), on the structure and content of social interactions and how these changes influence agricultural outcomes. AO allows farmers to call in to a hotline, ask questions and receive responses from agronomists. Working with a field partner, the research team randomly assigned toll-free access to AO to the chief agricultural decision maker of 800 cotton-growing households in Gujarat, India. A further 400 households served as a control group. In addition to a baseline survey, roughly half of the treatment group and the entire control group were surveyed by phone by phone after 5 months and in person after one year (800 households, hereafter, the 'phone survey' group). To estimate peer effects, the top 3 agricultural contacts of individuals in the phone survey group were surveyed by phone (1523 respondents, hereafter, the 'peer survey').

After 7 months, nearly 60% of the treatment group used the service for almost an hour on average. After 5 months, treatment respondents were no more likely to share agricultural information with their peers than the control group, but after one year they are 7% more likely to have shared information and 7% more likely to report recommending an input to a peer. The types of information shared suggest that this lag may be a result of waiting to observe the effects of inputs before sharing information about them. In addition, treated respondents are 8% less likely to have received information from their peers and 11% less likely to gather information by directly observing their peer's fields. Both the content and sources of shared information change: treated respondents are 48% more likely to report that they shared mobile phone-based information. In contrast, they are substantially less likely to report other farmers, input dealers and their past experiences as sources of shared information.

These changes in the behavior of treated respondents also indirectly influenced their peers. Treated respondents with peers who are also assigned to the treatment group are more likely to visit their homes to discuss agricultural topics and to report speaking to peers who they are not related to. Similarly control group respondents who have peers assigned to the treatment are less likely to consult input dealers for agricultural information and substantially more likely to cite our partner NGO as a source of agricultural information.[1] Finally, these changes in social interactions also have effects on the agricultural outcomes of peers. Non-study respondents with treated peers in their network report 4% lower losses in their cotton crop due to pest attacks – the most frequently requested and shared type of information from the AO system. In addition, they are more likely to grow (6%) and cultivate a larger area of cumin (0.25 acres), a risky but profitable cash crop.

Prior literature has demonstrated the potential for social learning in facilitating technology adoption in agriculture. Foster and Rosenzweig (1995) examine the period of Green Revolution in India. Using a panel dataset spanning 1968-71, they find that their neighbours experience with the adoption of high-yield varieties (HYV) influences profitability and probability of adoption. In the same context, Munshi (2004) shows that wheat growers

[1] In a related audit study of input dealers conduct in the region by the research team, we find that input dealers often give inappropriate recommendations for pesticide use in the context of cultivation.

are more responsive to neighbors behavior relative to rice growers. He argues that rice cultivation is more sensitive to local conditions thereby making social learning more difficult. Conley and Udry (2010) show evidence of social learning among pineapple growers in Ghana. Their empirical strategy relies in establishing whether farmers temporally adjust fertilizer inputs in response to information from peers.

In spite of these gains to social learning a number of other studies identify constraints to the exchange of information within villages and, in particular, observe variation in the adoption of profitable technologies within villages (Conley and Udry, 2001;Duflo *et al.*, 2008). The results of this paper suggest that the use of ICT's to provide valuable agricultural information can stimulate social interactions that support the sharing of information, particularly between treated respondents. However, ICT's may also change the ways in which people learn about agriculture, in this case, by reducing the frequency of direct observation of peer's field.

Information flow between peers may also depend on the nature of reference groups that relate peers. Van den Broeck and Dercon (2011) look at the adoption of pest-resistant technologies by banana farmers in Tanzania. The authors find differential effects on adoption that vary by whether other adopters were part of a kinship group, geographical neighbors or part of an informal insurance group. This paper provides evidence that ICT's may induce interactions between treated respondents outside of their traditional networks serving to both democratize access to and exchanges of valuable information.[2]

Section 2 describes the context of the study, Section 3 provides a conceptual framework for the study and Section 4 describes the data sources used and the empirical strategy. Section 5 discusses the results while Section 6 provides a discussion of the mechanisms underlying these results. Section 7 concludes.

[2]In this setting the vast majority of farmers cite peers as belonging to their *Jati* or sub-caste.

3.2 Context: Mobile Phone-Based Agricultural Extension in Rural India

Mobile phone-based agricultural extension systems are becoming increasingly popular in the developing world (Aker, 2011). Traditional 'Training & Visit' in-person extension typically involves extension agents either visiting farmers in person or inviting them to a central location. Mobile phone-based extension addresses many of the challenges presented by traditional systems of extension. First, it provides farmers with a dynamic source of information that can help farmers effectively respond to unanticipated shocks such as changing weather patterns and pest attacks. Mobile phone-based agricultural extension can also address agency problems in working with extension agents in remote areas. Depending on the type of system in place, the information provided is often publicly observable and it may be possible to confirm receipt as well. Importantly, as in the context of our technology, such systems may allow for demand-driven extension: by allowing farmers to request information, agencies are better able to respond with specific and customized advice that is of relevance to their agriculture.

The intervention studied in this paper is a mobile phone-based platform called Avaaj Otalo (AO). AO is an open-source platform that utilizes mobile phone networks to allow information to be delivered to and shared by farmers. Farmers receive weekly 'push-content, which includes detailed and complete agricultural information on weather and crop conditions that are delivered through an automated voice message. A previous paper (Cole and Fernando, 2015) estimated the impact of this service and found that demand for agricultural advice is substantial and farmers offered the service turn less often to traditional sources of agricultural advice. Treated respondents are nearly 40% more likely to report mobile phone-based information as an important source of agricultural information than control respondents. Farmers also invest more in recommended agricultural inputs resulting in large increases in yield for cumin (26.3%), and cotton yield (3.5%) for a sub-group that received frequent reminders to use the service. Importantly, for the present study, the authors do *not* find gains in agricultural knowledge among the treatment group.

3.3 Conceptual Framework

Several studies suggest that social learning can facilitate technology adoption in an agricultural context [3]. However, a stylized feature of agriculture in developing countries is variation in agricultural productivity both within regions and villages. These differences may in part result from differences in technology adoption or the appropriateness of technologies to underlying production conditions. Experimental work suggests that constraints in information sharing may also impose restrictions on the extent to which technologies spread within villages. Duflo *et al.* (2008) suggests that such conditions prevail in the context of Western Kenya. They suggest that less than 15% of farmers use fertilizer on their maize crop in spite of returns to doing so being greater than 100%. An experiment they conducted induced 10% of treatment farmers to adopt fertilizer, but the peers of treated farmers were no more likely to adopt than the peers of control farmers.

A subsequent experiment conducted in the same context by the authors randomly invited the peers of treatment farmers to a demonstration of the use of fertilizer by their field partner. Such peers were 17.8% more likely to adopt the fertilizer than the peers of farmers in the control group, suggesting the potential for gains through social learning. Yet, information collected by the authors suggest that farmers do not organically share information about agricultural production outside the context of their experiment and appear to know little about the production of their neighbors (Duflo *et al.*, 2008).[4] The authors speculate that the external provision of information by their partner NGO may have improved the reliability of information and the gains to sharing it.

These findings suggest that the propensity of farmers to share information and reduce barriers to technology adoption may in part be determined by farmer perceptions of

[3]For a review, see Foster and Rosenzweig (2010b)

[4]In the context of Ghana, Conley and Udry (2001) similarly describes agricultural information as not 'free-flowing' and networks for information as being 'sparse' .

the availability of valuable information. Social interactions that support the exchange of agricultural information may be characterized by multiple steady states: when there is valuable information, this increases the return to social interactions and induces technology adoption. In the absence of valuable information farmers may not have a reason to discuss their agriculture. In the context of our experiment, the external provision of AO may induce social interactions, induce social learning and support technology adoption, particularly between members of the treatment group. Alternatively, if farmers previously received information from their peers and these sources were replaced by the AO service, they may instead reduce the need for social interactions and reduce the potential for social learning. An even more exotic possibility maybe that mobile phone-based networks facilitate social learning through a 'virtual network' thereby reducing the need for in-person social interactions.

Even having identified the existence of such peer effects, there is considerable debate about the precise mechanisms which underlie them.Young (2001) identifies three mechanisms through which social interactions may influence adoption. 'Pure conformity' refers to situations in which fashion dictates behavior and individuals adopt a technology because they receive a benefit from 'fitting in'. 'Instrumental conformity' suggests a role for coordination in the adoption of technologies. Sometimes referred to as 'network effects' such conformity may occur when the return to adoption increases in the fraction of one's network that adopts a technology. Finally, 'informational conformity' emphasizes the role of peers in being a source of information and facilitating adoption. This latter mechanism has typically been described as 'social learning'.

Although conformity can be interpreted through the lens of economic rationality by appropriately defined preferences, most economic modeling of peer effects has focused on social learning and informational constraints. In these situations, social learning is modeled as a Bayesian learning process (Jovanovic and Nyarko, 1994). As such, social learning may be described as the process through which agents learn by observing or sharing information about the behavior of others.

Empirically, it has been extremely difficult to establish the presence of peer effects since individuals with similar characteristics or tastes may associate in the same social groups or might be affected by common shocks. Regressions of own peer outcomes on own outcomes may therefore be compromised by such concerns, which has been termed the 'reflection problem' in the literature(Manski, 1993). In order to convincingly estimate peer effects, one requires finding or creating situations in which an individual is 'randomly' assigned into a shared setting, or finding a setting in which existing networks are differentially exposed to treatments that are orthogonal to respondent and group characteristics. The section that follows makes the case for the latter case towards the estimation of peer effects in this paper.[5]

In this context as well, disentangling between imitation, learning and coordination effects is difficult because they have very similar empirical predictions (a discussion we return to in Section 6). Measuring intermediate outcomes such as knowledge may be useful in supporting a learning channel, or, alternatively, showing that information that was apriori unproductive was followed may improve the case for imitation.[6]

3.4 Data and Empirical Strategy

The results in this paper build off a previous experiment conducted to evaluate the impact of mobile phone-based agricultural extension (Cole and Fernando, 2015). The households that are a part of the experiment are in Surendranagar district in Gujarat, India. Lists of farmers were enumerated in cooperation with a field partner, the Development Support Center (DSC) in 40 villages, with the criteria for selection being that they were 1.) interested in participating in the study, 2.) grew cotton, 3.) owned a mobile phone and 4.) are the chief agricultural decision maker of their household.

A sample of 1200 respondents was selected from this pool, with 30 households in each

---

[5]In particular, Appendix E6 shows that the fraction of one's peer group exposed to the treatment are largely independent of respondent characteristics.

[6]Providing unproductive information would incur significant reputational effects for the service.

village participating in the study. Treatments were then randomly assigned at the household-level using a scratch-card lottery. The sample was split into three groups (T1, T2 and Control) of 400 households each. Treatment group 1 (hereafter, T1) receives toll-free access to AO in addition to traditional extension. The traditional extension component consisted of a single session lasting roughly two and half hours on DSC premises in Surendranagar. Treatment group 2 (hereafter, T2) only received toll-free access to AO, and a final 400 households served as the control group. The results presented in this paper are limited to a subset of this treatment group – roughly half of T1 and T2, referred to as the 'AO group' in this paper – and the entire control group, that were surveyed by phone. This 'phone survey' group consists of 797 households.

Baseline data was collected from June-July 2011, the 'phone survey' was completed in December 2011 and a midline survey was completed in June-July 2012. In addition, at baseline all respondents reported up to three peers with whom they most frequently discussed agricultural topics (referred to as 'peers' or 'top contacts' hereafter). These peers (1523 respondents) were then surveyed by phone in March, 2012 (referred to as the 'peer survey').[7]

Because of random assignment the causal effect of the intervention can be gauged by computing simple differences of the form:

$$y_{ij} = \alpha_j + \beta_1 \; Treat_{ij} + \varepsilon_{ij} \tag{3.1}$$

Where for household i in village j $Treat_i$ is an indicator variable that takes on the value 1 for an individual assigned to the AO group and α_j is a village fixed effect. The empirical results largely estimate (3.1) for a variety of outcome variables, using robust standard errors.

In addition, I compute a difference-in-differences estimator where both baseline data and midline data are available:

[7] Note, some of these 1523 peers may themselves be study respondents. An effort is made to look at the heterogeneous effects between peers of varying initial treatment assignments in Table 3.5 and Table 3.6.

$$y_{ijt} = \alpha_j + \beta_1 \ Treat_{ij} + \beta_2 \ Post_t + \beta_3 \ (Treat * Post)_{ijt} + \varepsilon_{ijt} \tag{3.2}$$

where, α_j and $Treat_{ijt}$ are as above, $Post_t$ is an indicator variable that takes on a value of 1 if the observation was collected at the midline) 0 otherwise, and $(Treat * Post)_{ijt}$ is the interaction of the preceding two terms.

Finally, I use the peer data to estimate whether the treatment also influenced the outcomes of peers in study respondents' social networks. The extent of such peer effects or information spillovers can be estimated with the following specification:

$$y_{ij} = \alpha_j + \beta(\frac{\# \ References \ in \ Treatment}{\# \ References})_{ij} + \sum_{i=2}^{7} I(\# \ References = i)_{ij} + \varepsilon_{ij} \tag{3.3}$$

where, α_j is as above, $\sum_{i=2}^{7} I(\# \ References = i)_{ij}$ is a fixed effect for the number of peers who cite a respondent as a top agricultural contact and $(\frac{\# \ References \ in \ Treatment}{\# \ References})_{ij}$ is the fraction of these respondents who are assigned to treatment. Appendix Table E6 assess whether respondents characteristics are balanced across the fraction of their peers assigned to the treatment.

3.5 Experimental Results

3.5.1 Summary Statistics and Balance

Demographic Variables

Table 3.1 contains summary statistics for age, education, landholdings, income and crop cultivation of respondents in the study and assesses balance across treatment and control for these variables. The first column reports summary statistics (mean and standard deviation) for the control group. The column that follows reports the mean and standard deviation for the AO group (i.e. all treatment respondents) and column (3) reports the simple difference between these two groups. The statistical significance of the differences is indicated to the right of the point estimate, with asterisks, at the 10% (*), 5% (**), and 1% (***) level.

TABLE 3.1 - SUMMARY STATISTICS AND BALANCE

Cell contents:	Control Group Mean (1)	AO Only Mean (2)	AO-Control ITT (3)
How old are you now?	35.768	36.092	0.324
	(10.221)	(10.794)	(0.738)
Years of Education	4.214	3.890	-0.315
	(3.838)	(3.943)	(0.267)
Landholdings- Acres	6.693	6.108	-0.586
	(16.218)	(5.438)	(0.836)
Imputed Agricultural Income from past 12 months	1.63e+05	1.77e+05	14380.126
	(1.47e+05)	(1.74e+05)	(10923.697)
Did you plant cotton in K'10?	0.985	0.980	-0.005
	(0.122)	(0.140)	(0.009)
Area of cotton planted in K'10 - Acres	4.448	5.186	0.735***
	(3.622)	(4.260)	(0.275)
Did you plant wheat in R'10-11?	0.776	0.745	-0.031
	(0.417)	(0.436)	(0.029)
Area of wheat planted in Rabi 10-11 - Acres	1.171	1.374	0.201
	(1.346)	(2.271)	(0.128)
Did you plant cumin in R'10-11?	0.425	0.405	-0.019
	(0.495)	(0.492)	(0.032)
Area of cumin planted in Rabi 10-11 - Acres	0.762	0.762	-0.002
	(1.406)	(1.414)	(0.096)
N	398	399	797

Notes: This table reports summary statistics and balance tests for phone survey sample. 'AO' refers to 399 respondents who were assigned access to the treatment and surveyed by phone. Column 1 provides the mean and standard deviation of the control group. Column 2 provides the mean and standard deviation for the AO group. Column 3 provides an Intention to Treat (ITT) estimate of the difference in means (and the robust standard error) between the AO group and the control group. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Respondents in the control group are 36 years old, have approximately 4 years of education, own roughly 6.7 acres of land, and earn roughly \$300 a month on average. Overall, column (3) shows that that the treatment and control are balanced on covariates. The one exception is the area of cotton cultivated which has a significant and large difference. In a previous paper (Cole and Fernando, 2015) we show that the randomization was successful across a wider set of outcome variables and that the main results are robust to controls for baseline cotton. The results that follow are similarly robust to controls for baseline cotton and, where applicable, it's interaction with a time-trend.

Sources of Information and Social Interactions

Table 3.2 assesses balance across treatment and control for a set of variables related to sources of information for agricultural decisions and social interactions.

Nearly all farmers cite past experience as an important source of information while roughly 85% cite input dealers. In contrast just 10% of farmers cite mobile phone-based information and government extension agents as important sources of information. Roughly 20-30% of farmers report NGO's, commission agents (who may either function as middlemen, buyers or both) and TV & print media as important sources of information.

Farmers report 'usually' speaking to their peers quite often about agricultural issues (nearly 5 times a month) with the vast majority of these interactions being in person (84%) rather than on the phone (3%). In addition, nearly 72% report that they have used an input recommended by a peer, suggesting that social networks are an important conduit for technology adoption. In contrast just 11% of respondents report receiving information from 10 farmers picked at random in their village in a study on social networks in Ghana (Conley and Udry, 2001). While not directly comparable to our setting, the authors suggest that the networks in the previous study can be characterized as 'sparse' and that information is not 'free-flowing'. In this context, peers live roughly 1.7 km from respondents on average and are assessed as being quite knowledgeable (nearly 4/5 scale on average) about agriculture, once again suggesting they are an important source of information about agriculture.

TABLE 3.2 - SUMMARY STATISTICS AND BALANCE - SOURCES OF INFORMATION AND SOCIAL INTERACTIONS

Cell contents:	Control Group Mean (1)	AO Only Mean (2)	AO-Control ITT (3)
Sources of information for agricultural decisions			
Past experience	0.997	0.998	0.000
	(0.050)	(0.050)	(0.004)
TV programmes	0.284	0.240	-0.043
	(0.451)	(0.428)	(0.030)
Cellphone-based information	0.093	0.108	0.015
	(0.291)	(0.310)	(0.021)
Newspaper/Magazines	0.231	0.188	-0.043
	(0.422)	(0.391)	(0.028)
Government extension work	0.093	0.085	-0.008
	(0.291)	(0.279)	(0.020)
NGO	0.264	0.262	-0.001
	(0.441)	(0.441)	(0.030)
Input dealer	0.857	0.863	0.007
	(0.351)	(0.345)	(0.023)
Commision agents	0.349	0.350	0.002
	(0.477)	(0.478)	(0.031)
Social Interactions with Peers			
No. of times discussed agriculture with peers in a usual month?	4.720	4.750	0.027
	(1.318)	(1.361)	(0.092)
Ever used an input recommendation from peers?	0.724	0.728	0.004
	(0.448)	(0.446)	(0.031)
Usually discuss agriculture in person with peers?	0.842	0.825	-0.017
	(0.365)	(0.380)	(0.026)
Usually discuss agriculture over phone with peers?	0.033	0.030	-0.003
	(0.178)	(0.171)	(0.012)
Avg. Distance from peers	1.737	1.753	0.015
	(0.461)	(0.510)	(0.034)
Avg. Subjective Knowledge Rating of peers	3.872	3.898	0.025
(Scale 1-5)	(0.704)	(0.663)	(0.047)
N	398	399	797

Notes: This table reports summary statistics and balance tests for sources of information and social interactions. 'AO' refers to 399 respondents who were assigned access to the treatment and surveyed by phone. Column 1 provides the mean and standard deviation of the control group. Column 2 provides the mean and standard deviation for the AO group. Column 3 provide an Intention to Treat (ITT) estimate of the difference in means (and the robust standard error) between the AO group and the control group. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

3.5.2 Usage of Mobile Phone-Based Extension Service

Panel A of Table 3.3 looks at whether providing toll-free access to AO was successful in inducing usage. After 7 months just four control respondents called in to the AO line. As a result, virtually all AO usage is accounted for by treated respondents. The mean usage for the AO group is roughly 65 minutes, although this reflects heavy usage by a small group of respondents. Winsorizing 5% of the observations brings down the overall mean to 31 minutes of usage.[8] Nearly 60% of treatment respondents called into the AO line, and 32% have asked questions. In addition, respondents listened to nearly 65% of all push call content on average.

Panel B of Table 3.3 shows questions asked on the AO system dis-aggregated by the crop to which the question is related. Cotton accounts for the majority (67%) of questions asked on the AO system while cumin (6%) and wheat (2%) are the next most popular crops. Splitting up requested information by topic, questions about pest management account for 63% of all questions asked, while fertilizer questions account for 8% and seeds account for 6%. The content of push calls, in comparison, focused more heavily on fertilizer use (53%) and on advice related to cumin cultivation (38%).

3.5.3 Social Interactions and Information Exchange

Table 3.4 estimates the impact of AO on social interactions centered about the exchange of agricultural information.

Panel A reports estimates using phone survey data roughly 5 months after the treatment had been administered. Initially, treated farmers are no more likely to share information with peers than their counterparts in the control group. However, Panel B shows that by the midline (post-harvest) they are 7% more likely to have shared information with their peers and 7% more likely to have recommended an input. This delay is perhaps indicative of the need for treated respondents to observe the realizations of inputs prior to sharing it. On the

[8] The analysis of AO usage in part builds of a mid-term evaluation of AO which in a previous version of Cole and Fernando (2015).

TABLE 3.3 - AO USAGE AND CONTENT

Cell contents:	Control Group Mean (1)	AO Only Mean (2)	AO-Control ITT (3)
Panel A: AO Usage			
Duration of Usage	0.003	65.983	66.635***
(Minutes)	(0.036	(211.147)	(11.375)
Called AO	0.011	0.568	0.556***
	(0.104	(0.496)	(0.026)
Purchased ag items as a result of AO advice -Self-reported	0.000	0.707	0.705***
	(0.000	(0.456)	(0.023)
Asked a question on the AO service	0.000	0.321	0.320***
	(0.000)	(0.467)	(0.024)
Percentage of Total Push Call time Listened to	0.000	0.655	0.654***
	(0.000)	(0.010)	(0.015)
N	398	399	797

Panel B: AO Content

Cell contents:	No. of Questions asked (1)	% of Total Asked (2)	No. of Push Calls (3)	% of Push Calls (4)
Question Count by Crop				
Cotton	739	0.67	12	0.38
Cumin	62	0.06	15	0.47
Wheat	19	0.02	9	0.28
Question Count by Theme				
Pest mangagement	690	0.63	22	0.69
Cotton pest management	542	0.49	8	0.25
Fertilizers	90	0.08	17	0.53
Seeds	67	0.06	4	0.13
Other	301	0.27	-	-
N	1104	1	32	1

Notes: Panel A reports usage of AO across the treatment and control groups. Usage statistics were collected on the AO server. 'AO' refers to 399 respondents who were assigned access to the treatment and surveyed by phone. Column 1 provides the mean and standard deviation of the control group. Column 2 provides the mean and standard deviation for the AO group. Column 3 provide an Intention to Treat (ITT) estimate of the difference in means (and the robust standard error) between the AO group and the control group. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level. Panel B reports questions asked on the AO system and are categorized by their related crops and themes. All Push calls contain information on multiple themes. The numbers include those questions asked by respondents who were assigned to receive AO access but were not included in the phone survey sample. A total of 32 push calls were sent during August 2011 - March 2012, with the average length of a push call at approximately 5 minutes.

TABLE 3.4 - EFFECT OF AO ON SOCIAL INTERACTIONS AND EXCHANGE OF INFORMATION

Cell contents:	Control Group Mean (1)	AO Only Mean (2)	AO-Control ITT (3)
A. Phone Survey Data			
Shared info with peers?	0.693	0.668	-0.019
	(0.462)	(0.471)	(0.034)
Received info from peers?	0.562	0.482	-0.076**
	(0.497)	(0.500)	(0.037)
Gathered info from observing peer's fields?	0.239	0.129	-0.107***
	(0.427)	(0.335)	(0.028)
N	398	399	797
B. Midline Data			
Shared info with peers?	0.617	0.683	0.070**
	(0.487)	(0.466)	(0.035)
Received info from peers?	0.766	0.754	-0.009
	(0.424)	(0.432)	(0.032)
Gathered info from observing peer's fields?	0.328	0.300	-0.033
	(0.470)	(0.459)	(0.034)
Recommended input to peers?	0.485	0.552	0.072*
	(0.500)	(0.498)	(0.037)
N	363	363	720
C. Sources of Shared Information (Phone Survey Data)			
Source of shared info: Past experience	0.329	0.156	-0.175***
	(0.472)	(0.364)	(0.031)
Source of shared info: TV program	0.060	0.027	-0.033**
	(0.237)	(0.163)	(0.015)
Source of shared info: Cell phone based info	0.008	0.471	0.468***
	(0.090)	(0.500)	(0.026)
Source of shared info: Other farmers	0.185	0.121	-0.064**
	(0.389)	(0.326)	(0.026)
Source of shared info: Input Dealers	0.196	0.110	-0.085***
	(0.397)	(0.313)	(0.025)
N	398	399	797

Notes: This table reports the effect of AO on social interactions and the exchange of information. Panel A uses data from the 'phone survey' conducted in Nov, 2011, while Panel B uses data from the midline survey conducted in June, 2012. 'AO' refers to 399 respondents who were assigned access to the treatment and surveyed by phone. Column 1 provides the mean and standard deviation of the control group. Column 2 provides the mean and standard deviation for the AO group. Column 3 provide an Intention to Treat (ITT) estimate of the difference in means (and the robust standard error) between the AO group and the control group. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level face of it, this evidence suggests that the availability of high quality agricultural information appears to induce more information sharing through networks by treated respondents and increases the potential for social learning.[9]

However, even by the time of the phone survey, treated respondents report that they are 7.6% less likely to receive information from their peers and report that they are 10% less likely to *learn* new information by observing their peers' fields. However, Appendix Table E4 shows that this is not because treated respondents interact less frequently with their peers about agricultural topics. Appendix E3 further disaggregates the types of information which are received less frequently by the treatment group, and those they less likely to learn about from observing fields of their peers. The considerable overlap in the prior categories suggest that AO may act as a substitute for information from peers and in so doing influence the propensity of respondents to gather information about their peers' agriculture through direct observation. However, these dynamics leave the overall frequency of social interactions about agriculture unchanged.[10]

Panel C shows that mobile phone-based information is the dominant source of information shared by treated respondents. Treated respondents are 47% more likely to report mobile phone-based information as the source of information they share. In contrast, they are far less likely to report traditional media as the source of shared information. Treated respondents are 18% less likely to report past experiences, 6% less likely to report other farmers and 9% less likely to report input dealers as the source of information they share relative to the control group.[11]

---

[9] Appendix E3 shows that information on pesticides (45%), fertilizers (31%) and seeds (28%) were the most frequently shared types of information. In the case of pesticides the 'phone survey' (5 months later) may have been too early to observe the effects of pesticides on plant growth, while the effects of seeds and fertilizers are likely observed post-harvest.

[10] It is worth noting here that direct observation of peers' fields maybe independent of social interactions.

[11] In general, appendix Table E5 shows an absence of heterogeneous effects of AO on information exchange across education, income and age. The one important exception is that more educated respondents are more likely to share mobile phone-based information.

3.5.4 Peer Effects: Structure of Social Interactions and Agricultural Outcomes

Balance for Peer Regressions

Random assignment of the treatment may also result variation in the proportion of one's peers who are treated, provided there is sufficient overlap in the social networks of study respondents. If the proportion of respondent social networks treated is independent of potential outcomes, such variation would allow for the identification of peer effects as comparisons can be made between individual's with differing fractions of peers who are treated. Appendix Table E6 provides support for this claim by assessing whether the fraction of peers who are treated in one's network varies with observable respondent characteristics. For both the the balance regressions that follow and peers effects that are report (Table 3.5 and Table 3.6) the universe of peers refers to the 'top agricultural contacts' elicited at baseline from study respondents. Respondents were requested to state up to three peers with whom they frequently discussed agriculture. As such, it is possible that there are peers with whom respondents interact less frequently who either were, or were not exposed to the treatment and this may result in attenuation bias.

Columns (1) and (2) assess this claim for respondents who were assigned to the treatment group, while columns (3) and (4) do so for respondents who are not a part of the study (hereafter, non-study respondents). Columns (1) and (3) report the mean and standard deviation for treatment and non-study respondents who do not have treated peers in their social network. Columns (2) and (4) report the respective coefficients for these groups from a regression of respondents characteristics on the fraction of their peers who are assigned to the treatment group. We observe that in each case, that the fraction of one's peer group assigned to the treatment appears uncorrelated with characteristics such as age, education and landholdings. The exception is once again the area of cotton grown, where non-study respondents with a higher fraction of peers who are treated grow larger amounts of cotton. The results that follow are robust to baseline controls for cotton.

Peer Effects: Social Interactions and Information Exchange

Table 3.5 uses the peer survey to estimate whether the fraction of one's peers assigned to the treatment influenced social interactions. Columns (1), (3) and (5) report the mean and standard deviation for control respondents, treatment respondents and non-study respondents, respectively, who did not have any peers assigned to the treatment group.

Column (2) reports the coefficient on a regression of control respondent outcomes on the fraction of their peer group assigned to the treatment. In general, we observe that the fraction of a control group respondent's peer group assigned to the treatment did not influence their own outcomes. The one exception is that such respondents were less likely to report gathering information by observing peer's fields (significant at the 10% level). In addition, control group respondents with more treated peers are less likely to share information with their peers and more likely to receive information from their peers, but these effects are not significant at traditional levels.

Column (4) provides analogous estimates for treatment respondents. In this case, we observe that treatment respondents with peers who are also treated are substantially more likely to visit their peer's houses to discuss agricultural topics (significant at 10% level). Such respondents are also more likely to report speaking to peers who are not their relatives (significant at the 5% level). They are also more likely to share information with their peers, receive information from their peers and rate their peers as having a higher subjective level of agricultural knowledge, however in these cases the coefficients are not significant at traditional levels of statistical significance. Column (6) presents estimates of the effect of treated peers on non-study respondent outcomes. In this case, we do not find that having a higher proportion of treated peers influences outcomes related to social interactions.

Taken together, we observe some evidence to suggest that the provision of AO influences social interactions between treated respondents and their peers. In particular, these results suggest that AO may have externalities on physical interactions for the control group (i.e. whether they observe the fields of treated peers) and that there may be complementarities between treated respondents. As such, there is some support for the proposition that the

TABLE 3.5: PEER EFFECTS : SOCIAL INTERACTIONS AND EXCHANGE OF INFORMATION

Dependent Variable	Control Group		Treatment Group		Non-Study Peers	
	Control Peer Group (Baseline) (1)	Fraction of Peers Treated *Post (Midline) (2)	Control Peer Group (Baseline) (3)	Fraction of Peers Treated *Post (Midline) (4)	Control Peer Group (Baseline) (5)	Fraction of Peers Treated (Midline) (6)
Shared info with peers?	0.633	-0.144	0.682	0.061	0.504	0.007
	(0.483)	(0.106)	(0.467)	(0.096)	(0.500)	(0.030)
Received info from peers?	0.767	0.029	0.747	0.086	0.413	0.023
	(0.423)	(0.089)	(0.436)	(0.092)	(0.493)	(0.030)
Gathered info from observing peer's fields?	0.335	-0.160*	0.282	0.040	0.177	-0.007
	(0.473)	(0.088)	(0.451)	(0.099)	(0.382)	(0.023)
Recommended input to peers?	0.482	0.068	0.531	0.096	-	-
	(0.501)	(0.106)	(0.500)	(0.099)		
Went to peer's house to discuss ag topics	0.522	0.022	0.449	0.171*	0.441	0.011
	(0.501)	(0.100)	(0.498)	(0.100)	(0.497)	(0.029)
Subjective rating of peer's agricultural knowledge	3.856	-0.025	3.911	0.128	3.683	0.058
	(0.714)	(0.190)	(0.676)	(0.191)	(0.962)	(0.061)
N	245	363	245	357	466	773
Spoke to a more than once a month about ag issues	0.617	-0.016	0.576	0.033	0.102	0.006
	(0.487)	(0.115)	(0.495)	(0.112)	(0.302)	(0.019)
Spoke to a peer who is not a relative	1.000	-0.128	1.000	0.196**	-	-
	(0.000)	(0.089)	(0.000)	(0.084)		
Spoke to a peer on the phone	0.030	0.004	0.025	-0.061	-	-
	(0.171)	(0.050)	(0.158)	(0.052)		
N	266	791	276	787	466	773

Notes: This table assesses whether the fraction of one's peers assigned to the treatment group influences the structure of social interactions. The midline survey took place between 4th June and 8th July 2012. Column 1 reports the mean and standard deviation for all control respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment. Column 2 reports the coefficient on the interaction of the number of peers who were assigned to the treatment group and a dummy for whether the observation is from the midline. The regression specification includes both of these variables separately and their interaction. Column 3 reports the analogous mean and standard deviation for all treatment respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment. Column 4 likewise reports the coefficient on the interaction of the number of peers who were assigned to the treatment group and a dummy for whether the observation is from the midline. Column 5 reports the mean and standard deviation for peers who were not respondents in the main study and who were not referenced by a treatment respondent. Column 6 reports the coefficient on the number of peers who were assigned to the treatment group, from a regression of the characteristic in question on this variable. The regresson specifications in column 2 ,4 and 6 include dummies for the number of peers referenced, the amount of cotton grown at baseline, and village fixed effects. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

provision of valuable external information may result in more frequent visits to discuss agricultural topics among treated respondents as well as a broadening of their peer base. The next section asks whether these changes also influenced agricultural outcomes.

3.5.5 Peer Effects: Agricultural Outcomes

Table 3.6 estimates analogous peer regressions to see if the presence of treated peers in one's network also influenced agricultural outcomes. Columns (1), (3) and (5) once again report the mean and standard deviation for control respondents, treatment respondents and non-study respondents, respectively, who did not have any peers assigned to the treatment group.

Column (2) reports the effect of having a higher fraction of treated peers on outcomes for the control respondents. We observe that control respondents with more treated peers are less likely to report input dealerships as a source of information for agricultural decisions (significant at the 10% level) but are more likely to report the NGO we partnered with as a source of information (not significant at traditional levels).[12]

In column (4) we see that treatment respondents with peers who are also assigned to the treatment group use AO even more, although this estimate is not statistically significant. We also observe that having more treated peers in a social network also increases the likelihood of reporting NGO as a source of information for agricultural decision-making. In the overwhelming majority of cases the name of the NGO is our field partner. This effect is suggestive of complementarities between treated peers in the same network acting collectively to obtain advice from our NGO partner. Treatment respondents with more treated peers are also more likely to cultivate cumin (significant at 10% level) and are less likely to report input dealers as a source of information (not significant).

Column (6) shows that non-study respondents with more treated peers are also more likely to grow cumin (6%) and plant a larger amount of it (.25 acres more). Those with more

[12]From qualitative interviews, the name of our partner NGO was also commonly used to refer to the AO service.

TABLE 3.6: PEER EFFECTS : AO USAGE, SOURCES OF INFORMATION AND SOWING

Dependent Variable	Control Group		Treatment Group		Non-Study Peers	
	Control Peer Group	Fraction of Peers Treated *Post	Control Peer Group	Fraction of Peers Treated *Post	Control Peer Group	Fraction of Peers Treated
	(Baseline)	(Midline)	(Baseline)	(Midline)	(Baseline)	(Midline)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: AO Usage						
Called AO line	-	-	0.000	-0.061	-	-
			(0.000)	(0.067)		
Total AO usage (Minutes)	-	-	0.000	18.572	-	-
			(0.000)	(22.722)		
Panel B: Sources of Information						
Past Experience	0.996	0.000	0.998	0.025	0.217	0.012
	(0.061)	(0.009)	(0.039)	(0.020)	(0.412)	(0.031)
Input Dealerships	0.872	-0.193*	0.850	-0.066	0.383	0.004
	(0.335)	(0.112)	(0.357)	(0.089)	(0.487)	(0.034)
Cell Phone	0.075	-0.100	0.099	-0.097	0.040	-0.001
	(0.264)	(0.092)	(0.299)	(0.069)	(0.197)	(0.009)
NGO	0.259	0.079	0.272	0.190***	0.024	0.021
	(0.439)	(0.094)	(0.445)	(0.072)	(0.153)	(0.014)
Panel C: Sowing Decisions						
Planted Cumin	0.436	-0.030	0.422	0.151*	0.237	0.059*
	(0.497)	(0.108)	(0.494)	(0.090)	(0.425)	(0.030)
Area of Cumin Planted	0.735	-0.116	0.792	0.454	0.525	0.255*
(Acres)	(1.186)	(0.280)	(1.499)	(0.305)	(1.695)	(0.133)
Planted Wheat	0.756	-0.146	0.722	0.029	0.253	-0.011
	(0.431)	(0.108)	(0.448)	(0.092)	(0.435)	(0.031)
Area of Wheat Planted	1.152	-0.023	1.181	-0.086	0.328	-0.050
(Acres)	(1.307)	(0.223)	(1.816)	(0.222)	(1.153)	(0.077)
Proportion of Cotton Lost	-	-	-	-	0.142	-0.039***
to Pest Attacks (%)					(0.224)	(0.015)
N	266	791	276	787	540	870

Notes

This table assesses whether the fraction of one's peers assigned to the treatment group influences own agricultural outcomes. The midline survey took place between 4th June and 8th July 2012. Column 1 reports the mean and standard deviation for all control respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment. Column 2 reports the coefficient on the interaction of the number of peers who were assigned to the treatment group and a dummy for whether the observation is from the midline. The regression specification includes both of these variables separately and their interaction. Column 3 reports the analogous mean and standard deviation for all treatment respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment. Column 4 likewise reports the coefficient on the interaction of the number of peers who were assigned to the treatment group and a dummy for whether the observation is from the midline. Column 5 reports the mean and standard deviation for peers who were not respondents in the main study and who were not referenced by a treatment respondent. Column 6 reports the coefficient on the number of peers who were assigned to the treatment group, from a regression of the characteristic in question on this variable. The regresson specifications in column 2 ,4 and 6 include dummies for the number of peers referenced, the amount of cotton grown at baseline, and village fixed effects. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

treated peers in their networks also report 4% less cotton crop loss as a result of pest attacks. It is worth noting that cotton pest management practices are both the most requested type of information from the AO service and the type of information most likely to be shared between peers.

3.6 Discussion of Mechanisms

These results show that the provision of mobile phone-based extension induced treatment respondents to (1) share information received from the service with their peers, (2) alter the site of interactions with their peers and indirectly influence (3) the agricultural outcomes of their peers.

While the results show that having treated peers in one's social network influences own agricultural outcomes, it is unclear which mechanisms underlie these effects. One possibility is that individuals may conform with the practices of their peers who have gained valuable information from AO. Given prior research which shows that AO has a causal effect on technology adoption and yields, conforming in this case may also be productive. In the absence of information provided that is clearly unproductive, it is difficult to rule out this mechanism.

Similarly, it may be the case that individuals internalize the information provided by their peers and update their priors on the productivity of recommended inputs. In this case, the social interactions can be said to facilitate social learning. While the particular technologies under consideration do not lend themselves to 'instrumental conformity' (i.e. my return to cultivating cumin is not a direct function of my peer's decision to cultivate it) it may well be the case that providing more valuable information into a network through AO increases the returns to social interactions, particularly given evidence for complementarities between treatment respondents.

Surprisingly, Cole and Fernando (2015) find that AO does not influence the agricultural outcomes of treated respondents. As such, I do not find evidence that having treated peers in one's social network influences own agricultural knowledge either (not reported). On the surface, the lack of effects on knowledge is consistent with a story of conformity.

3.7 Threats to Validity

3.7.1 Attrition

As Appendix Table E2 shows, while 56 respondents were no longer a part of the respondent group at the time of thephone survey, these numbers were spread fairly evenly across treatment (29) and control (27). In addition, the treatment and control attritees appear to be statistically indistinguishable on the basis of a standard set of covariates, including age, area of land owned, years of education and amount of cotton planted.

3.8 Conclusion

This paper investigates the effects of a mobile phone-based agricultural extension service on information exchange and the structure of social interactions in village India.

An initial evaluation of the intervention shows that roughly 60% of the treatment group used the service for almost an hour on average. Previous research shows that this intervention dramatically altered the reported sources of information that treated respondents used in making agricultural decisions. This paper finds that treated respondents were initially no more likely to share this agricultural information with their peers than the control group, but after one year they are 7% more likely to have shared information and recommended an input to their peers. One interpretation of delayed sharing may be that treated respondents waited to observe the effects of inputs recommended by the service before sharing these recommendations with their peers. A large shift towards using mobile phoned-based information as the source of shared information and the types of input information shared (pesticides and fertilizers) also lend support to this interpretation.

The intervention also had important effects on the nature of exchanges of information within study villages. Treated respondents are 8% less likely to have received information from their peers and 11% less likely to gather information by directly observing their peer's fields. There is considerable overlap between the types of information treated respondents are less likely to receive and those they are less likely to receive through direct observations of their peer's fields. This overlap suggests that the reduction in information received is passive in nature and may occurs through a decreased return to physically observing their peer's fields.

Through the collection of data on the peer's of study respondents, I find that the intervention also indirectly influenced the outcomes of peers. Treated respondents with a higher fraction of treated peers in their social network were more likely to visit their homes to discuss agricultural topics and report speaking to peers who they are not related to. Furthermore, control group respondents with a higher fraction of treated peers are less likely to consult input dealers for agricultural information and substantially more likely to cite our partner NGO as a source of agricultural information. Taken together, these effects suggest that the intervention indirectly influenced the structure of social interactions in villages. In addition, they suggest that the provision of external information through ICT's may increase the returns to social interactions, particularly among the treated, but reduce learning through direct observation.

These changes in social interactions also influenced agricultural outcomes. Peers surveyed who were not a part of the original study with a higher fraction of treated peers report 4% lower losses in their cotton crop due to pest attacks. Information of pest management for cotton was the most frequently requested and shared type of information. Both treatment respondents and non-study respondents with more treated peers in their network are 6% more likely to have cultivated cumin, a lucrative cash-crop.

The precise mechanisms underlying these effects is difficult to establish. The lack of changes in agricultural knowledge – both directly through the intervention and indirectly through exposure to treated peers – suggests a role for imitation rather than social learning. Future research may attempt to manipulate the productivity of the information provided – the adoption of unproductive information by peers would not be consistent with social learning – to more clearly distinguish between these channels.

References

Agarwal, B. (1994). *A field of one's own: Gender and land rights in South Asia*, vol. 58. Cambridge University Press.

Aker, J. (2011). Dial "a" for agriculture: Using ict's for agricultural extension in development countries. *Agricultural Economics*.

Anderson, J. and Birner, R. (2007). How to make agricultural extension demand-driven? : The case of india's agricultural extension policy. *IFPRI Discussion Paper 00729*.

Anderson, K. and Feder, G. (2007). Chapter 44: Agricultural extension. *Handbook of Agricultural Economics*.

Angrist, J. and Evans, W. (1998). Children and their parents labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, **88** (3), 450–477.

—, Lavy, V. and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, **28** (4), 773–824.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, **90** (430), 431–442.

Arnold, F., Kishor, S. and Roy, T. (2002). Sex-selective abortions in india. *Population and Development Review*, **28** (4), 759–785.

Arrow, K. J. (1969). Classificatory notes on the production and transmission of technological knowledge. *The American Economic Review*, pp. 29–35.

Auty, R. (1993). Sustaining development in mineral economies: the resource curse thesis.

Bandiera, O., Burgess, R., Gulesci, S., Rasul, I. and Sulaiman, M. (2011). Can entrepreneurship programs transform the economic lives of the poor?

Banerjee, A. V. and Duflo, E. (2005). Growth theory through the lens of development economics. *Handbook of economic growth*, **1**, 473–552.

— and Newman, A. F. (1993). Occupational choice and the process of development. *Journal of political economy*, pp. 274–298.

Bardhan, P. (2012). *Awakening giants, feet of clay: Assessing the economic rise of China and India*. Princeton University Press.

— and Mookherjee, D. (2011). Subsidized farm input programs and agricultural performance: A farm-level analysis of west bengal's green revolution, 1982-1995. *American Economic Journal: Applied Economics*, **3** (4).

Beegle, K., De Weerdt, J. and Dercon, S. (2011). Migration and economic mobility in tanzania: Evidence from a tracking survey. *Review of Economics and Statistics*, **93** (3), 1010–1033.

Besley, T. and Burgess, R. (2000). Land reform, poverty reduction, and growth: evidence from india. *Quarterly Journal of Economics*, pp. 389–430.

Bianchi, M. and Bobba, M. (2012). Liquidity, risk, and occupational choices. *The Review of Economic Studies*, p. rds031.

Birkhaeuser, D., Evenson, R. and Feder, G. (1991). The economic impact of agricultural extension: A review. *Economic Development and Cultural Change*.

Blattman, C., Fiala, N. and Martinez, S. (2013). Credit constraints, occupational choice, and the process of development: long run evidence from cash transfers in uganda. *SSRN Scholarly Paper*.

Bleakley, H. and Ferrie, J. (2010). Shocking behavior: Land lotteries in 1832 georgia and 1901 oklahoma and later life outcomes.

— and Ferrie, J. P. (2013). *Up from Poverty? The 1832 Cherokee Land Lottery and the Long-run Distribution of Wealth*. Tech. rep., National Bureau of Economic Research.

Bryan, G., Chowdhury, S. and Mobarak, A. M. (2011). Seasonal migration and risk aversion.

Chernina, E., Castañeda Dower, P. and Markevich, A. (2013). Property rights, land liquidity, and internal migration. *Journal of Development Economics*.

Cole, S. A. and Fernando, A. N. (2014). The value of advice: Evidence from mobile phone-based agricultural extension.

— and — (2015). The value of advice: evidence from mobile phone-based agricultural extension. *Harvard Business School Finance Working Paper*, (13-047).

Conley, T. and Udry, C. (2001). Social learning through networks: The adoption of new agricultural technologies in ghana. *American Journal of Agricultural Economics*, pp. 668–673.

Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, **94** (1), 260–272.

— and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review*, pp. 35–69.

Das, N., Yasmin, R., Ara, J., Kamruzzaman, M., Davis, P., Behrman, J., Roy, S. and Quisumbing, A. R. (2013). How do intrahousehold dynamics change when assets are transferred to women? evidence from bracs challenging the frontiers of poverty reductiontargeting the ultra poor program in bangladesh.

De Janvry, A., Emerick, K., Gonzalez-Navarro, M. and Sadoulet, E. (2012). Certified to migrate: Property rights and migration in rural mexico.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications.

— and Dreze, J. (2002). Poverty and inequality in india: a re-examination. *Economic and Political Weekly*, pp. 3729–3748.

Deininger, K. and Feder, G. (2001). Land institutions and land markets. *Handbook of agricultural economics*, **1**, 288–331.

— and Goyal, A. (2012). Going digital: Credit effects of land registry computerization in india. *Journal of Development Economics*, **99** (2), 236–243.

—, — and Nagarajan, H. (2013). Women's inheritance rights and intergenerational transmission of resources in india. *Journal of Human Resources*, **48** (1), 114–141.

—, Jin, S. and Nagarajan, H. K. (2008). Efficiency and equity impacts of rural land rental restrictions: Evidence from india. *European Economic Review*, **52** (5), 892–918.

—, — and — (2009). Determinants and consequences of land sales market participation: Panel evidence from india. *World Development*, **37** (2), 410–421.

Duflo, E. (2006). Field experiments in development economics. *Prepared for the World Congress of the Econometric Society*.

—, Kremer, M. and Robinson, J. (2008). How high are rates of return to fertilizer? evidence from field experiments in kenya. *The American economic review*, pp. 482–488.

—, — and — (2011). How high are rates of return to fertilizer? evidence from field experiments in kenya. *Nudging Farmers to USe Fertilizer: Theory and Experimental Evidence in Kenya*.

Feder, G. (2005). The challenges facing agricultural extension - and a new opportunity. *New Agriculturalist*.

—, Lau, L. and Slade, R. (1987). Does agricultural extension pay? the training and visit system in northwest india. *American Journal of Agricultural Economics*, **69** (3).

Field, E. (2007). Entitled to work: Urban property rights and labor supply in peru. *The Quarterly Journal of Economics*, pp. 1561–1602.

Foster, A. and Rosenzweig, M. (2010a). Microeconomics of technology adoption. *Annual Reviews of Economics*.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of political Economy*, pp. 1176–1209.

— and — (1996). Technical change and human-capital returns and investments: evidence from the green revolution. *The American economic review*, pp. 931–953.

— and — (2002). Household division and rural economic growth. *Review of Economic Studies*, pp. 839–869.

— and — (2007). Economic development and the decline of agricultural employment. *Handbook of development economics*, **4**, 3051–3083.

— and — (2010b). Microeconomics of technology adoption. *Annual Review of Economics*, **2**.

Galor, O. and Zeira, J. (1993). Income distribution and macroeconomics. *The review of economic studies*, **60** (1), 35–52.

Gandhi, R., Veeraraghavan, R., Toyama, K. and Ramprasad, V. (2009). Digital green: Participatory video and instruction for agricultural extension. *Information Technologies and International Development*, **5** (1).

Glendenning, C., Babu, S. and Asenso-Okyere, K. (2010). Review of agricultural extension in india: Are farmer′ information needs being met? *IFPRI Discussion Paper 01048*.

Goldstein, M. and Udry, C. (2008). The profits of power: Land rights and agricultural investment in ghana. *Journal of Political Economy*, **116** (6), 981–1022.

Gollin, D. (2014). The lewis model: A 60-year retrospective. *Journal of Economic Perspectives*, **28** (3), 71–88.

—, Lagakos, D. and Waugh, M. E. (2014a). Agricultural productivity differences across countries. *American Economic Review, Papers and Proceedings*.

—, — and — (2014b). The agricultural productivity gap. *Quarterly Journal of Economics*.

— and Rogerson, R. (2010). Agriculture, roads, and economic development in uganda. *NBER Working Paper*.

Gulesci, S. (2012). *Labor-tying and poverty in a rural economy: evidence from bangladesh*. Tech. rep., IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Gupta, M. D. (2005). Explaining asia's missing women: a new look at the data. *Population and development review*, **31** (3), 529–535.

Harris, J. R. and Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American Economic Review*, pp. 126–142.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, **124** (4), 1403–1448.

IBRD (2014). *World Development Indicators 2014*. World Bank Publications.

Jack, K. (2011). Market inefficiences and the adoption of agricultural technologies in developing countries. *ATAI*.

Jayachandran, S. and Pande, R. (2013). Why are indian children shorter than african children? *Department of Economics, Northwestern University, Mimeo.*

Jensen, R. and Miller, N. (2011). Keepin' em down on the farm: Migration and strategic investment in childrens schooling. *Unpublished manuscript*.

Jensen, R. T. (2003). *Equal Treatment, Unequal Outcomes?: Generating Sex Inequality Through Fertility Behaviour*.

Jodhka, S. S. (2006). Beyond'crises': rethinking contemporary punjab agriculture. *Economic and Political Weekly*, pp. 1530–1537.

Jovanovic, B. and Nyarko, Y. (1994). *Learning By Doing and the Choice of Technology.* Tech. rep., National Bureau of Economic Research.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, **75** (1), 83–119.

Kumari Bhat, A. and Dhruvarajan, R. (2001). Ageing in india: drifting intergenerational relations, challenges and options. *Ageing and Society*, **21** (05), 621–640.

Kuran, T. (2012). *The long divergence: How Islamic law held back the Middle East*. Princeton University Press.

Laband, D. N. and Lentz, B. F. (1998). The effects of sexual harassment on job satisfaction, earnings, and turnover among female lawyers. *Industrial and Labor Relations Review*, pp. 594–607.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The manchester school*, **22** (2), 139–191.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, **60** (3), 531–542.

McKenzie, D. and Woodruff, C. (2012). What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Policy Research Working Paper*, (6202).

Morris, S. and Pandey, A. (2007). Towards reform of land acquisition framework in india. *Economic and Political Weekly*, pp. 2083–2090.

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of development Economics*, **73** (1), 185–213.

— and Rosenzweig, M. (2007). Why is mobility in india so low? social insurance, inequality, and growth.

Patel, N., Chittamuru, D., Jain, A., Dave, P. and Parikh, T. (2010). Avaaj otalo - a field study of an interactive voice forum for small farmers in rural india. *Proceedings of ACM Conference on Human Factors in Computing Systems*.

Piketty, T. (2014). *Capital in the Twenty-first Century*. Harvard University Press.

Pörtner, C. C. (2010). Sex selective abortions, fertility and birth spacing. *University of Washington, Department of Economics, Working Paper*.

Rao, N. (2007). Custom and the courts: ensuring women's rights to land, jharkhand, india. *Development and change*, **38** (2), 299–319.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Saito, K. and Weidemann, C. (1990). Agricultural extension for women farmers in africa. *World Bank*.

Samuelson, P. A. (1954). The pure theory of public expenditure. *The review of economics and statistics*, pp. 387–389.

Sen, A. (1990). More than 100 million women are missing. *The New York Review of Books*.

Sharma, A. (2007). The changing agricultural demography of india evidence from a rural youth perception survey. *International Journal of Rural Management*, **3** (1), 27–41.

Skoufias, E. (1995). Household resources, transaction costs, and adjustment through land tenancy. *Land Economics*, pp. 42–56.

Van den Broeck, K. and Dercon, S. (2011). Information flows and social externalities in a tanzanian banana growing village. *The journal of development studies*, **47** (2), 231–252.

Vogl, T. S. (2013). Marriage institutions and sibling competition: Evidence from south asia*. *The Quarterly journal of economics*, **128** (3), 1017–1072.

Wang, S.-Y. (2012). Credit constraints, job mobility, and entrepreneurship: Evidence from a property reform in china. *Review of Economics and Statistics*, **94** (2), 532–551.

Young, H. P. (2001). *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press.

Appendix A

Appendix to Chapter 1

A.1 Data Appendix and Summary Statistics

Appendix A1: Variable Definitions

Birth Order: This variable is constructed using the date of birth of all siblings of the head of the household and their own date of birth. The variable is ordered by year of birth and any siblings born in the same year as assigned the same birth order.

Changes in Landholdings : This variable is constructed for all siblings of the household head including himself. It is the difference between the reported amount of current land owned by the sibling and the land inherited.

Hired Agricultural Labor : This variable aggregates self-reported information on the number of man days and the wage rate paid to for agricultural labor by task. The tasks include preparatory tillage, sowing & transplanting, manuring & fertilizer, weeding & interculture, irrigation, harvesting, threshing & winnowing and 'other operations'. The total value of hired labor for each of these tasks is then summed.

Household Consumption : This variable is an aggregate of yearly expenditure on cereals (rice, wheat, maize, bajra, jowar, ragi and other cereals), pulses (tur, gram, urd, moong and other pulses) and values home production of these crops at village specific market prices.Other food items include gur/khandsari, edible oils, spices, milk, milk products, eggs, meat, fish, fruits, vegetables, bread, biscuits, confectionaery, processed food, beverages, cooked meals as wages and any other items. Any home production of these items is also valued at village-specific market prices.

This measure of consumption also includes yearly expenditure on durables including radios, transistors, fans, torch, lantern, petromax, metal utensils, water boilers, buckets, korosen stoves, bicycles, tricyles, motor cycle/scooter, car/jeep/van, sewing machines, wooden furniture, cots, wooden boxes, Almirahs, steel trunk/boxes, steel furniture, watches, clock/time piece, camera, television, VCR, cassette recorder, washing machines, pressure cookers, mixed/grinder, electric iron, geysers, refrigerators, cassette players and walkmans.

Expenditure on clothing was elicited separately and includes expenditure on readymade garments, dhoties, sarees, cloth for garments, shawls/pullovers, hosiery, footwear, tailoring charges. Expenditure on fuel – and any home production using for consumption – was also elicited separately and is include in the measure of consumption. This includes firewood, kerosene, charcoal, soft coke, gas, electricity and other fuels.

The final set of expenditure categories include toiletry and cosmetics, bedding charges, towels/linen, pan/beedis/cigarettes, intoxicants, newspapers/periodicals, medical expenses for all household members, education expenses, entertainment expenses, expenses in hotels/restaurants, house rent paid, repairs to house rented-in, repairs to consumer durables, payments to domestic servants, payments to barber/laundry/priest/sweeper, travel expenses other regular expenses, expenditure on marriage ceremonies (including gifts), expenditure on other social ceremonies and expenditure on religious ceremonies.

Inherited Land : This variable is directly asked from respondents, distinct from land currently owned, and is measured in acres. The question asked is '[What is the amount of] land inherited prior to recent period'.

Family Land: This variable is directly asked from respondents and is measured in acres. The question asked is '[What is the amount of] land owned by head's parents?'.

Land Improvement: This variable sums the costs of hired labor, the imputed value of family labor and any material costs – valued at market rates – incurred in improving the quality of land through terracing, bunding, leveling, fencing and reclamation in the last 10 years.

Land Increase : This variable is constructed for all siblings of the household head including himself. It is coded as '1' if the respondent's current landholdings are greater than their inherited landholdings and '0' otherwise.

Net Dowry : The household head is asked the value of dowry payments paid and received by his parents for each of his siblings. Net dowry is the difference between the value of all payments received minus all payments paid.

Non-agricultural Occupation: The question asked in the survey used to define this variable is the household head's response to their 'Primary Activity Status'. The options for this question include 1. self-employed farming, 2. self employed non-farming, 3. salaried, 4.agricultural wages, 5.non-agricultural wages, 6. agricultural family worker, 7.non-agricultural family worker, 8.pensioner, 9. other. 'Non-agricultural occupation' was coded as '0' if the head responded with 1,4,6 or 7 and '1' otherwise.

Predicted Share : This instrument is constructed using the total number of brothers reported by a respondent who reached the age of 10. This data is contained in a section enumerating all siblings of the head of the household

Rural to Urban Migration : The household head is asked where each sibling and child 'lives now', where the options are: 1. same village, 2. town in same district, 3. village in same district, 4. town in other district of same state, 5. village in other district of same state, 6. town of other state, 7. village of other state or 8. village/town of other country. The variable is coded as '1' if head responded with 2,4,6 or 8, and '0' otherwise.

Total Man Days of Agricultural Labor : This variable aggregates self-reported information on the number of days in the past year the household head worked in agricultural labor by aggregating the number of man days reported for preparatory tillage, sowing & transplanting, manuring & fertilizer, weeding & interculture, irrigation, harvesting, threshing & winnowing and 'other operations'.

Total Man Days of Non-Agricultural Labor : This variable aggregates self-reported information on the number of days in the past year the household head worked in a self-employment activity, as a salary earner, and as a wage earner.

Took out Loan : This variable is coded as '1' if the respondent reported taking out a loan in the last five years. The loans could be taken out for any purpose including agricultural investment, investment in self-employment enterprises, social ceremonies, purchasing of consumer durables, education of children or 'other'.

Total Value of Loans : This variable sums the total amount repaid and the total amount outstanding for loans taken out in the last five years. This includes both cash loans and the rupees value of in-kind loans.

APPENDIX A2: UNDERSTANDING OF INHERITANCE RULES FROM GUJARAT DATA

Table A2.1 Understanding of Inheritance Rule

	Freq.	Percent
Equal Shares to all sibs	152	14.66
Equal Shares to brothers	852	82.16
Oldest Brother gets more	2	0.19
No Standard Rule	19	1.83
888	4	0.39
999	8	0.77
Total	1,037	100

Table A2.2 Do Some Brothers Inherit Better Land than Others?

	Freq.	Percent
Yes	302	29.12
No	733	70.68
888	1	0.1
999	1	0.1
Total	1,037	100

Table A2.3 If one brother has more education or a better job, will he inherit less land?

	Freq.	Percent
Yes	70	6.75
No	965	93.06
888	1	0.1
999	1	0.1
Total	1,037	100

Notes:

These tables present summary statistics on perceptions of land inheritance rules and adherence to them. The data was collected by the author and is from a random sample of 1,037 respondents enganged in agriculture in Gujarat, India. Data Source: Collected by Author

APPENDIX A3: DETAILS OF NON-AGRICULTURAL OCCUPATIONS

Table A3.1 Distribution of Non-Agricultural Occupations

	Freq.	Percent
Non-farm business	259	18.05
Salaried position	458	31.92
Non-agricultural wage work	315	21.95
Other	403	28.08
Total	1435	100.00

Table A3.2: Details of Salaried Work

	Freq.	Percent
Teachers	101	22.49
Service Workers (guides, undertakers and embalmers, peons, helper, priest)	79	17.59
Clerical and related workers	60	13.36
Other	209	46.55
Total	449	31.29

Table A3.3: Details of Non-Farm Businesses

	Freq.	Percent
General Merchant	45	17.44
Tea Shop, Restaurant, Hotel	14	5.43
Tailoring	13	5.04
Artisan	12	4.65
Commission Agent	11	4.26
Other	186	72.09
Total	258	100.00

Notes:
These tables present summary statistics on the breakdown of non-agricultural occupations. The occupational classifications are based on what users report as their primary activity status in the data. The sample is restricted to all children of Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. Data Source: ARIS-REDS Dataset.

APPENDIX A4: WAGE GAPS BY OCCUPATION IN RURAL INDIA (WITHIN DISTRICT OLS ESTIMATES)

	Occupational Wage Gaps by Farmer Landholdings		
	(1)	(2)	(3)
Comparison Group	All Farmers	<= 3 acres	> 3 acres
A. Primary Occupation: Non-Farm Business			
Wage Gap	-1.219	18.885***	-10.669***
	(1.603)	(1.827)	(1.834)
N	30707	13621	22266
B. Primary Occupation: Salaried Job			
Wage Gap	21.610***	37.265***	11.290***
	(2.051)	(2.347)	(2.258)
N	31637	14551	23196
C. Primary Occupation: Non-Agricultural Labor			
Wage Gap	-15.062***	4.695***	-26.601***
	(1.113)	(1.236)	(1.398)
N	35019	17933	26578
Daily Wage for Comparison Farmers	81.902	50.933	97.171
Sex FE	Y	Y	Y
Age FE	Y	Y	Y
Education FE	Y	Y	Y
District FE	Y	Y	Y

Notes:

This table computes differences in the average daily wage within districts by occupation in rural India. The sample is restricted to all individuals in rural India in the Indian Human Development Survey. Column 1 in Panel A reports the estimated OLS coefficient from a regression of the daily wage on a dummy variable coded as 1 if the main source of income is business and 0 if it is farming. Column 2 reports the same coefficient but restricting the comparison group to farmers with less than or equal to 3 acres of land. Column 3 restricts the comparison group to farmers with more than 3 acres of land. Panel B does similary where the main ocupation is instead a salaried job, and Panel C considers Non-agricultural wage work. All specifications include district fixed effects, non-parametric controls for sex, age (15 dummies, 0-80, 5 year intervals) and years of education (15 dummies, 0-15 years, 1 year intervals). Primary Source of Income defined as source of income with highest proportion relative to total income for an indvidual. A 'farmer' is defined as an individual whose highest proportion of income is from own agricultural cultivation or agricultural labor. The daily wage for farming is calculated as total farm profit divided by the number of days spent in agricultural labor, or the agricultural wage income in the case of agricultural labor. Data Source: Indian Human Development Survey.

APPENDIX A5: DEMOGRAPHIC CHARACTERISTICS

Table A5.1: Distribution of Number of Siblings

Siblings (1)	Frequency (2)	Percent (3)
0	274	5.7
1	473	9.84
2	729	15.16
3	811	16.86
4	777	16.16
5	726	15.1
6	506	10.52
7	265	5.51
8	142	2.95
9	64	1.33
10	32	0.67
11	5	0.1
12	3	0.06
14	2	0.04
Total	4809	100
Mean	3.77	-
Median	4.00	-

Table A5.2: Distribution of Number of Brothers

Brothers (1)	Frequency (2)	Percent (3)
0	782	16.26
1	1,288	26.78
2	1,270	26.41
3	795	16.53
4	421	8.75
5	181	3.76
6	56	1.16
7	10	0.21
8	6	0.12
Total	4,809	100
Mean	1.92	-
Median	2.00	-

Table A5.3: Distribution of Birth Order

Birth Order (1)	Frequency (2)	Percent (3)
1	1,814	37.72
2	1,035	21.52
3	715	14.87
4	544	11.31
5	304	6.32
6	211	4.39
7	99	2.06
8	52	1.08
9	21	0.44
10	10	0.21
11	4	0.08
Total	4,809	100

Notes:
These tables report summary statistics on the distribution of siblings, brothers and birth order. Siblings born in the same year were assigned the same birth order since it is not possible to distinguish between twins and those born in the same calendar year. The sample is restricted to all children of Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. Data Source: ARIS-REDS Dataset.

APPENDIX A6: DETAIL ON BORROWING

Table A6.1 : Purpose of Loan

	Freq.	Percent
Agricultural Investment	332	27.35
Non-Agricultural Investment	384	31.63
Self Employment	67	5.52
Repayment of Earlier Loans	7	0.58
Social Ceremonies	78	6.43
Purchasing Consumer Durables	42	3.46
Education of Children	10	0.82
Other	294	24.22
Total	1214	100.01

Table A6.2 : Collateral Required for Loan?

	Freq.	Percent
Yes	285	23.48
No	929	76.52
Total	1,214	100

Table A6.3 : Type of Collateral

	Freq.	Percent
Land	236	83.1
Gold and Jewellery	22	7.75
Agricultural Asset	1	0.35
Consumer Durable	1	0.35
Others	24	8.45
Total	284	100

Notes:

These figures correspond to all loans taken out in the last 5 years by male Hindu household heads who parents owned land in the 1999 ARIS-REDS survey. Data source: ARIS-REDS Dataset.

Appendix B

Appendix to Chapter 1

B.1 Additional Results

APPENDIX B1: EFFECTS OF EXPECTED INHERITANCE OF LAND ON OCCUPATIONAL CHOICE, CONSUMPTION AND EDUCATION

Dependent Variable	Non-Ag Occupation Binary Variable		Household Consumption Log(Rs.)		Education (Years)	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Inherited Land	-0.253**	-0.140*	0.056	0.031	-0.000	-0.000
(Acres)	(0.125)	(0.081)	(0.121)	(0.061)	(1.264)	(0.650)
No. of Siblings FE	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y
Mean of Dep. Var.	0.33	0.33	10.288	10.288	6.688	6.688
First Stage F-statistic	-	4.898	-	4.898	-	4.898
N	898	898	898	898	898	898

Notes:

This table reports estimates of the long-term effect of inherited land occupational choice, household consumption and education for the subsample of heads whose father is still alive. The sample in columns 1,2,5 and 6 is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. The sample in column 3 and 4 are all male siblings of these household heads (including the heads) aged above 10 years. Note, this data is reported for all siblings not just siblings residing in the household at the time of the survey. Panel A includes all households, while Panel B limits the analysis to households whose family had less than 3 acres. The dependent variable in col 1 and 2 is Non-Ag occupation and is defined as the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable in column 3 and 4 is dummy variable for whether or not the sibling migrated to an urban area in the same district or outside of it. The dependent variable col 5 and 6 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The instrument specfication used is Predicted Share = 1 / (1+ Brothers). The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Standard errors are clustered at the family level for sibling-level regressions. Data Source: ARIS-REDS Dataset.

Appendix B2: The Effects of Birth Order on Land Ownership (Sibling-Level Data)

Appendix B2 reports the coefficients on the birth order dummies from the following within-family regression:

$$
\begin{aligned}
Y_{ij} \quad = \quad & \alpha_j + \gamma_z \sum_{z=1}^{6} I(Birthorder_{ij} = z) + \mu_1 Age\_Dummies_{ij} \\
& + \mu_2 Education\_Dummies_{ij} + \eta_{ij} \qquad \text{(B.1)}
\end{aligned}
$$

Where Y is a dummy coded as 1 if current landholdings are greater than inherited landholdings and 0 otherwise, for sibling i of head j, and α_j is a family fixed effect. Column (2) shows that latter-borns are less likely to experience an increase in their current landholdings over their inherited landholdings. In the main sample, 26% of household heads report that they experienced an increase in their landholdings over the prior two decades. Of these respondents, nearly 40% report receiving 'gifts' of land, a category distinct to inheriting, leasing or purchasing land. While no further details are given about these gifts in the 1999 wave, in the 2006 wave of the REDS survey, 80% of land leased in is from family members. The majority of these contracts are oral rather than written, they do not involve a fee and have no specified term. Taken together, these facts support the interpretation that latter-born siblings, unbound by social obligations, 'lease' their land to first-born siblings, and are more likely to specialize in non-agricultural occupations to the benefit of their family's future consumption.

APPENDIX B2: THE EFFECTS OF BIRTH ORDER ON LAND OWNERSHIP
(SIBLING-LEVEL DATA)

Dependent Variable	Inherited Land (Binary) (1)	Land Increase (Binary) (2)	Change in Landholdings (Acres) (3)
2nd Born	-0.009	-0.016*	-0.052
	(0.008)	(0.008)	(0.051)
3rd Born	0.007	-0.032***	-0.083
	(0.011)	(0.011)	(0.069)
4th Born	0.010	-0.049***	-0.269***
	(0.014)	(0.015)	(0.092)
5th Born +	-0.001	-0.051***	-0.321***
	(0.018)	(0.019)	(0.118)
Constant	0.542***	0.275**	1.382*
	(0.112)	(0.123)	(0.751)
Family FE	Y	Y	Y
Age FE	Y	Y	Y
Depvar Mean	0.674	0.256	0.731
N	14773	14773	14773

Notes:
inheriting land and changes in landholdings over time. The sample is restricted to all male siblings who reached the age of 10 years prior to death. In each family, one of the brothers is a household head in the main analysis. The data is at the sibling-level. Note, this data is reported for all siblings not just siblings residing in the household at the time of the survey. The dependent variable in Col 1-3 is the total number of man days spent in agriculture during the prior season. Columns 1-5 report the coefficient on a dummy for being the 2nd born sibling, 3rd born sibling, 4th born sibling and the 5th born or later sibling. The dependent variable in column 1 is a dummy variable coded as 1 if the sibling inherited land. The dependent variable in column 2 is a dummy coded as 1 if the sibling's current landholdings are greater than his inherited landholdings, and in column 3 it is the diference between current and inherited landholdings in acres. All specifications include family fixed effects, age fixed effects (0-100 years, 5 year intervals, 19 dummies), and education fixed effects (0-14 years, 1 year intervals, 13 dummies). The excluded group are first born siblings between the ages of 0-5 with less than an year of education. Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

APPENDIX B3: HETEROGENEOUS EFFECTS OF INHERITED LAND BY TIMING OF HEADSHIP (2SLS ESTIMATES)

Dependent Variable	Non-Ag Occupation (Binary)	HH Consumption Log(Rs.)	Non-Ag Occupation (Binary)	HH Consumption Log(Rs.)	Non-Ag Occupation (Binary)	HH Consumption Log(Rs.)	Non-Ag Occupation (Binary)	HH Consumption Log(Rs.)
Sample	All		Only First Borns		Only Latter Borns		All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Land	-0.024***	0.031***	-0.041***	0.035**	-0.011	0.035***	-0.009	0.040***
	(0.008)	(0.010)	(0.012)	(0.015)	(0.010)	(0.013)	(0.011)	(0.014)
Below Median Age	-0.046	-0.075	0.096	0.232*	-0.029	-0.102	0.008	-0.077
(BMA)	(0.048)	(0.055)	(0.114)	(0.134)	(0.058)	(0.070)	(0.013)	(0.075)
Land*BMA	0.011	-0.007	-0.023	-0.072**	0.007	-0.003	-0.027	-0.005
	(0.011)	(0.013)	(0.027)	(0.032)	(0.013)	(0.016)	(0.032)	(0.017)
Land*BMA*Firstborn	-	-	-	-	-	-	-0.027	-0.058
							(0.032)	(0.038)
No. of Siblings FE	Y	Y	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y
Depvar Mean	0.298	10.442	0.298	10.442	0.298	10.442	0.298	10.442
First Stage F-Stat	44.733	44.733	9.182	9.182	19.944	19.944	3.752	3.752
N								

Notes:
This table tests whether the timing of becoming a household head influences occupational choice and household consumption. All coefficients reported are 2SLS estimates. The sample is restricted to first born Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. The sample in co1 1-2 includes all respondents, in col 3-4 the sample includes only first born respondents, in col 5-6 it includes only latter borns, and in col 7-8 it includes all respondents. Columns 1-6 reports the 2SLS coefficients on inherited land, a dummy variable (Below Median Age) coded as 1 if the respondent became the head of the household at an age that was below the median for the sample (32 years) and their interaction. The two endogenous variables are instrumented with two instruments : Predicted Share = (1/1+Brothers) and the interaction between Predicted Share and the dummy for Below Median Age. In columns 7-8 the specification includes two additional endogenous variables: the interaction between inherited land and a dummy for first born and the triple interaction between inherited land, a dummy for below median age and a dummy for firstborn. The latter two variables are instrumented with the interaction predicted share and first born and the triple interaction between predicted share, first born and a dummy for below median age. The specification also includes controls for first born and its interact with below median age but these are not reported. The dependent variable in columns 1,3,5 and 7 is Non-Ag occupation, and is defined as the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable in columns 2,4,6 and 8 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children, from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

APPENDIX B4: HETEROGENEOUS EFFECTS OF INHERITED LAND BY CROPPING PATTERNS (2SLS ESTIMATES)

Dependent Variable	Non-Ag Occupation (Binary) (1)	HH Consumption Log(Rs.) (2)
Land	-0.007	0.015**
	(0.006)	(0.007)
Paddy	-0.144***	-0.018
	(0.051)	(0.059)
Land*Paddy	-0.022*	0.030*
	(0.013)	(0.015)
No. of Siblings FE	Y	Y
Family Land FE	Y	Y
District FE	Y	Y
Depvar Mean	0.298	10.442
First Stage F-Stat	14.396	14.396
N	4809	4809

Notes:

This table tests whether the types of crops grown leads to differential impacts of inherited land on occupational choice and household consumption. All coefficients reported are 2SLS estimates. The sample is restricted to first-born Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Each column reports the 2SLS coefficients on inherited land, a dummy variable (Paddy) coded as 1 if the respondent's major kharif crop is rice/paddy (i.e. greatest share of land sown) and their interaction. The two endogenous variables are instrumented with two instruments: Predicted Share = (1/1+Brothers) and the interaction between Predicted Share and Paddy. The dependent variable in Column 1, Non-Ag occupation, is defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable Column 2 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children, from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Data Source: ARIS-REDS Dataset.

APPENDIX B5 : THE EFFECT OF INHERITED LAND ON NON-FARM BUSINESS OWNERSHIP, SALARIED WORK AND NON-AGRICULTURAL WAGE WORK

Dependent Variable	Non-Farm Business Binary Variable		Salaried Job Binary Variable		Non-Agricultural Wage Work Log(Rs.)	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Inherited Land	-0.001*	-0.007***	-0.001	-0.000	-0.002***	-0.004
(Acres)	(0.001)	(0.002)	(0.001)	(0.003)	(0.000)	(0.003)
Mean of Dep. Var.	0.054	0.054	0.095	0.095	0.062	0.062
First Stage F-statistic	-	125.952	-	125.952	-	125.952
N	4809	4809	4809	4809	4809	4809
No. of Siblings FE	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y

Notes:

This table reports estimates of the long-term effect of inherited land on non-farm business ownership, holding a salaried position and non-agricultural wage work. Columns 1, 3 and 5 report OLS coeffiient estimates while columns 2, 4 and 6 report 2SLS estimates. The sample are Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head-level. The dependent variable in all columns correspond to the primary status reported by the respondent in the REDS survey. In column 1 and 2 if the primary status is non-farm business then it is coded as 1, in cols 3 and 4 if it is a salaried position and if cols 5 and 6 if it is non-agriculural wage work. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The instrument specfication used is Predicted Share = 1 / (1+ Brothers). The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Standard errors are clustered at the family level for sibling-level regressions. Data Source: ARIS-REDS Dataset.

Appendix C

Appendix to Chapter 1

C.1 Additional Robustness Tests

APPENDIX C1: ROBUSTNESS OF 2SLS ESTIMATES TO ALTERNATIVE INSTRUMENT SPECIFICATIONS

Instrument Specification	Linear (No. of Brothers)	Predicted Share	Predicted Land	Log(Predicted Land)	Non-Parametric (Brother Dummies)
	(1)	(2)	(3)	(4)	(5)
Panel A. Dependent Variable: Non-Ag Occupation (Binary)					
Inherited Land	-0.020***	-0.018***	-0.002	-0.042***	-0.017***
(Acres)	(0.005)	(0.005)	(0.002)	(0.007)	(0.005)
Mean of Dep. Var.	0.298	0.298	0.298	0.298	0.298
First Stage F-Statistic	128.137	125.952	101.458	141.607	22.433
N	4809	4809	4809	4809	4809
Panel B. Dependent Variable: Yearly Household Consumption, Log (Rs.)					
Inherited Land	0.032***	0.027***	0.022***	0.052***	0.031***
(Acres)	(0.006)	(0.006)	(0.003)	(0.007)	(0.006)
Mean of Dep. Var.	10.442	10.442	10.442	10.442	10.442
First Stage F-Statistic	128.137	125.952	101.458	141.607	22.433
N	4809	4809	4809	4809	4809
No. of Siblings FE	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y

Notes:
This table tests the robustness of 2SLS estimates of the long-term effect of inherited land on occupational choice and household consumption to alternative specifications of the instrument. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Column 1 reports 2SLS estimates with the instrument specified as the (linear) number of brothers, in column 2 it is 'Predicted Share' = 1/(1+Brothers), in column 3 it is 'Predicted Land' = Family Land/(1+Brothers), in column 4 it is Log(Predicted Land), and in column 5 it is the a set of dummies for the number of brothers (8 dummies in total, I report the coefficients for up to 5 brothers which account for 98.11% of sample). The dependent variable in Panel A is Non-Ag occupation and is defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable Panel B is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

Appendix C2: Reduced Form Simulations of Selective Migration

A potential selection concern arises from the fact that all surveyed households in the REDS dataset reside in rural areas. As such, a subset of individuals who migrate to urban areas after inheriting land are not sampled and may compromise the estimated relationships. For example, if these 'missing migrants' inherited small amounts of land and subsequently took up non-agricultural jobs in urban areas, I would underestimate the negative effect of land. Conversely, if the migrants inherited large amounts of land and then took up a non-agricultural occupation, I would overestimate the negative effect of land. Given the nature of the REDS data these migrants would need to result from the movement of entire *families* to urban areas; household heads report the location of their siblings irrespective of where they reside. While nationally representative estimates of the extent of permanent rural-urban *family* migration are not available, studies suggest this form of migration is extremely rare in India (Munshi and Rosenzweig, 2007). In the REDS sibling data, just 1.1% of 16,130 male siblings have migrated to urban areas. Foster and Rosenzweig (2007) estimate the *individual* rural to urban migration rate for males aged 15-24 for each decade between 1961 and 2001 using the corresponding Indian censuses.[1] They find that migration rates vary from 3% to 5% for each of the decades between 1961-2001, suggesting very limited migration even when considering the movement of individuals rather than entire families.

The estimated negative relationships between inherited land and both migration and entering non-agricultural work suggest that migrants would need to have large landholdings in order to overturn the estimates. It is worth noting that this is a hypothetical at odds with the estimated negative causal effect of inherited land on urban migration. Additionally, these landholdings cannot be so large that they have little influence on the 2SLS estimates. The latter restriction is a result of the nonlinearity of the estimated relationship between inherited land and both occupational choice and migration as suggested by Panel B of Table 1.3 and Panel A & B of Figure 1.4 and the weighting structure of 2SLS with covariates. The

[1] They assume that mortality does not vary differentially between urban and rural areas and suggest that if anything local amenities may be better in urban areas leading to an overestimate of the out-migration rate.

estimated slopes are especially negative and precise for inheritances of up to 4 acres of land. However, for those inheriting more than 4 acres, the qualitative nature of the relationship is unclear and the estimates are imprecise. As such, migrants with very large landholdings would be included in covariate-specific LATE's (i.e. the 2SLS estimate computed for subsets of the sample covariates) that are qualitatively different in sign from the overall LATE and have little variation in the instrument as a consequence of having few observations and are therefore not heavily weighted in the overall LATE.

These restrictions suggest that migrants whose parents owned intermediate amounts of land – i.e. covariate values that occur frequently in the data and drive the negative estimated relationships – would be the most likely to overturn the reduced form estimates. Having specified family landholdings, the sibling sex composition of migrants would determine their inherited landholdings. Appendix C2 models the sibling sex composition of migrants as resulting from a series of draws from a binomial distribution and estimates the reduced form for occupational choice – all urban migrants are assumed to hold non-agricultural occupations – under varying probabilities of success (i.e. the probability of drawing a male sibling). These simulations quantify how skewed the sibling sex composition of migrants would need to be to overturn the reduced form estimate.[2]

The simulations add observations to mimic a 10% rural to urban migration rate: the census-based *individual* urban migration rate for the three decades preceding the REDS survey. In both panels the reduced form estimate from the main specification is indicated by the horizontal red line, while the grey area shows the 95% confidence interval for the simulated reduced form coefficients. Panel A shows the estimated reduced form coefficients when migrants are assumed to to have the most frequent sibling and family land combination: 3 siblings and parents who own 5 acres of land, while in Panel B they are assumed to have parents who own 40 acres land (95th percentile for family landholdings).[3]

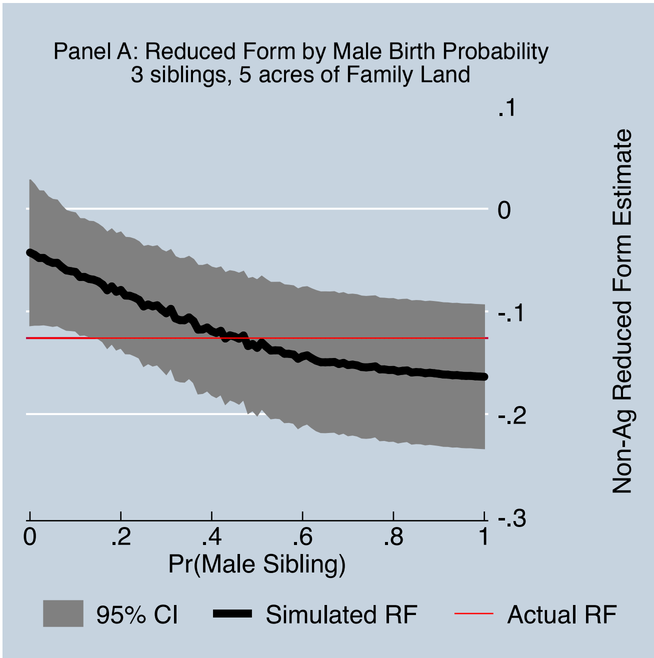
[2]If the probability of drawing a male sibling is zero, all the respondents siblings will be sisters and he will inherit all his parents land.

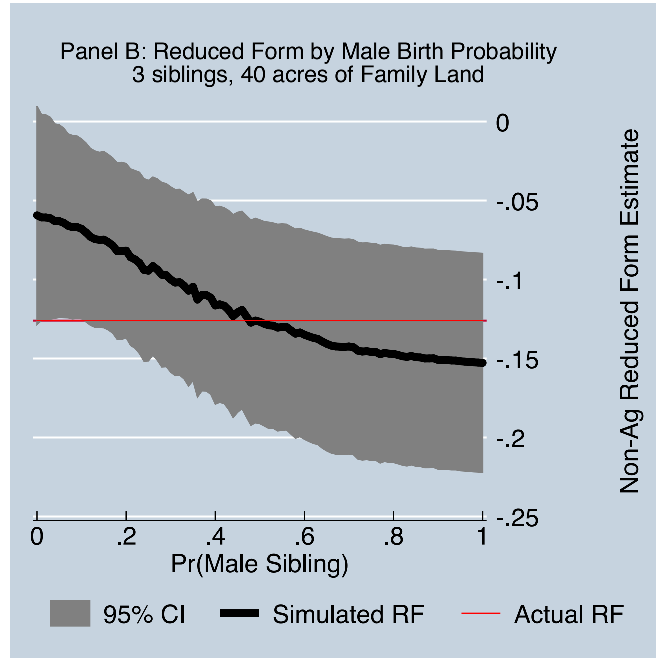
[3]While respondents with an extremely small number of siblings may be the most likely to be missing from the REDS survey, these covariate values occur very infrequently in the data and are given little weight in the

As suggested by the discussion above, migrants with smaller family landholdings (Panel A) have a greater influence on the reduced form estimates than those with large landholdings (Panel B). However, even in Panel B the probability of a male sibling occurring would need to be less than 0.17 – the point at which the red line leaves the confidence interval – in order to overturn the reduced form relationship. This is substantially lower than the biological probability of a male and the observed ratio of brothers to siblings for rural to urban migrants in the IHDS data: 0.53.

2SLS estimates because they also permit little variation in the first stage fitted values. Only children make up just 5.7% of the sample.

APPENDIX C2: SIMULATIONS OF SELECTIVE MIGRATION AND THE REDUCED FORM EFFECT ON OCCUPATIONAL CHOICE





Notes:

These figures plot the results of simulations intended to test the robustness of the reduced form relationship for non-agricultural occupation to selective family migration. Both panels add the simulated data to the main sample of Hindu male household heads whose parents owned land in the 1999 wave of the ARIS-REDS dataset. Both panels assume a family migration rate of 10%. Panel A assumes the migrants have 3 siblings and their parents own 5 acres of land, while Panel B assumes that the migrants have 3 siblings and their parents own 40 acres of family land. Where N is the number of siblings assigned to the migrants, the program takes N draws from a binomial distribution with a success K for each of the migrants. K is varied from 0 to 1 in intervals of 0.01 and the reduced form relationship is estimated for each of these values and plotted with the 95% confidence interval using robust standard errors. The red line in each panel shows the estimated reduced form estimate with the main specification: -0.126. In each case the missing migrants are assumed to come from Allahabad District in Uttar Pradesh. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children, from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. Data Source: ARIS-REDS Dataset.

APPENDIX C3: THE EFFECT OF INHERITED LAND ON BY SIBLING COHORT SIZE (2SLS ESTIMATES)

		2SLS Estimates Estimates				
	Mean/ S.D.	Effect of Inherited Land on Outomces by Sibling Cohort Size†				
Dependent Variable	Full Sample	2 siblings	3 siblings	4 siblings	5 siblings	Full Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Dependent Variable: Non-Agricultural Occupation						
Inherited Land	0.298	-0.007	-0.043***	-0.010	-0.021*	-0.018***
(Acres)	(0.458)	(0.013)	(0.014)	(0.012)	(0.011)	(0.005)
N	4809	729	811	777	726	4809
Panel B. Dependent Variable: Log(Household Consumption)						
Inherited Land	10.442	0.032**	0.005	0.006	0.053***	0.027***
(Acres)	(0.614)	(0.016)	(0.017)	(0.014)	(0.011)	(0.006)
N	4809	729	811	777	726	4809
No. of Siblings FE	-	N	N	N	N	Y
Family Land FE	-	Y	Y	Y	Y	Y
District FE	-	Y	Y	Y	Y	Y

Notes:
This table presents 2SLS estimates for the effect of inherited land on occupational choice and household consumption by sibling cohort size. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. †Columns 2-6 report the 2SLS coefficient on inherited landholdings. Columns 2-5 report the coefficient for household heads with varying numbers of siblings, while Column 6 includes all household heads. Non-Ag occupation is defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. Log(Household Consumption) is the natural logarithm of yearly household consumption which includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children, from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. Brothers are defined as male siblings who grew up to at least the age of 10. Results are robust to alternative definitions and using ever born siblings. Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

Appendix C4: Robustness to Differential Stopping Rules

Another threat to the conditional independence of the instrument stems from son-preferring, differential stopping behavior (SP-DSB). On average, women in India are more likely to belong to families with a larger number of siblings and have less education on average (Jensen, 2003. While the inclusion of sibling fixed effects takes care of some of these concerns, it may still be the case that families with different fertility constraints or preferences end up with a similar number of children, resulting in an apples to oranges comparison in the regressions of interest.

The use of an alternative instrument, which only uses variation resulting from the sex composition of siblings *older* than the respondent, partially addresses this concern. This variation is by definition unaffected by differential stopping behavior as these siblings are born prior to the respondent. When used together with birth order fixed effects for the respondent, the estimates in column (4) of appendix C4 are very similar to those in the main specification. In the case of occupational choice the coefficient is not precisely estimated and this may once again be due to a much weaker first stage (two-thirds of F-statistic from main specification).

In order to further alleviate this concern, column (5) reports the estimates for the subset of respondents whose youngest sibling is female. By definition, such families could not have engaged in son-preferring differential stopping behavior. These families are, therefore, more likely to have exhausted a fertility constraint. When fertility constraints are binding it is more likely that sibling sex composition reflects the biological probability that children are born to a specific sex. Once again, the estimated coefficients are comparable to those obtained in the main specification in column (1) and in this case they are also precisely estimated.

APPENDIX C4: ROBUSTNESS OF 2SLS ESTIMATES TO SEX SELECTIVE FERTILITY PREFERENCES

Specification	Main	Sex Selection Tests: Exact Permuation	Sex Selection Tests: Sibling Spacing Controls	Differential Stopping Tests: Older Siblings	Differential Stopping Tests: Youngest Sibling is Female
	(1)	(2)	(3)	(4)	(5)
Panel A. Dependent Variable: Non-Ag Occupation (Binary)					
Inherited Land	-0.018***	-0.029***	-0.018***	-0.008	-0.018***
(Acres)	(0.005)	(0.007)	(0.005)	(0.008)	(0.006)
Mean of Dep. Var.	0.298	0.298	0.298	0.298	0.292
N	4809	4809	4809	4809	1947
Panel B. Dependent Variable: Log(Household Consumption)					
Inherited Land	0.027***	0.016**	0.027***	0.049***	0.035***
(Acres)	(0.006)	(0.008)	(0.006)	(0.010)	(0.008)
Mean of Dep. Var.	10.442	10.442	10.442	10.442	10.455
F-stat (First Stage)	125.952	89.579	132.955	83.447	59.106
N	4809	4809	4809	4809	1947
No. of Siblings FE	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y
Sibling Sex Permuation FE	N	Y	N	N	N
Sibling Spacing FE	N	N	Y	N	N
Instruments	Pred Share	Pred Share	Pred Share	Prior Bros	Pred Share
Sample	All	All	All	All	Last Born Sis

Notes:
This table tests the robustness of 2SLS estimates of the long-term effect of inherited land on occupational choice and household consumption to sex selection and son-preferring differential stopping behavior, both of which stand to violate the conditional independence assumption of the instrument. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Column 1 reports 2SLS estimates for the main specification used in Table 3. Column 2 includes fixed effects for the exact permutation of the sex of siblings born prior to the head of the household (i.e. MMF, FFM etc..) and includes 223 dummy variables. Column 3 includes fixed effects for the average birth spacing (in years) between siblings (0-10+ years, 6 month intervals, 18 dummies. Column 4 uses the sex composition of siblings born prior to the head as an instrument for inherited landholdings (i.e. number of brothers). This specification also includes a set of 10 dummy variables for the birth order of the head of the household. Column 5 limits the sample to the subset of heads whose youngest sibling is female. This is under the assumption that those families who stop on a girl are more likely to have satisfied a resource constraint than stopped because of son-preferring differential stopping behavior. The dependent variable in Panel A is Non-Ag occupation and is Non-Ag occupation is defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable in Panel B is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

APPENDIX C5: ROBUSTNESS OF 2SLS ESTIMATES TO CONTROLS FOR PARENT'S DOWRY EXPENDITURE AND HEAD'S EDUCATION

	First Stage		Main	Education Controls	Dowry Controls	Main	Education Controls	Dowry Controls
Dependent Variable	Education (Years)	Net Dowry (Binary)	Non-Ag Occupation (Binary)	Non-Ag Occupation (Binary)	Non-Ag Occupation (Binary)	HH Consumption Log(Rs.)	HH Consumption Log(Rs.)	HH Consumption Log(Rs.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A : First Stage Regressions								
Predicted Share	1.397***	-0.087***	-	-	-	-	-	-
	(0.386)	(0.031)						
Mean of Dep. Var.	5.847	0.698	-	-	-	-	-	-
F-stat (First Stage)	13.119	8.091						
N	4809	4809	-	-	-	-	-	-
Panel B. 2SLS Estimates								
Inherited Land	-	-	-0.018***	-0.015***	-0.021***	0.027***	0.028***	0.024***
(Acres)			(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Mean of Dep. Var.	-	-	0.298	0.298	0.298	10.442	10.442	10.442
F-stat (First Stage)			125.952	124.544	118.234	125.952	118.234	124.544
N	-	-	4809	4809	4809	4809	4809	4809
No. of Siblings FE	Y	Y	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y
Education FE	N	N	N	Y	N	N	Y	N
Net Dowry FE	N	N	N	N	Y	N	N	Y

Notes:
This table tests the robustness of 2SLS estimates of the long-term effect of inherited land on occupational choice and household consumption to controls for net dowry receipts and education controls. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Column 1 in Panel A reports the first stage for years of education using the 'predicted share' instrument. Column 2 in Panel A reports the first stage for net dowry. This variable coded as 1 if the net dowry receipts of the household are above median. The latter is calculated as the net sum of all dowry payments and receipts for the parents which are reported for each sibling of the head of the household. 37% of the sample do not report paying or receiving dowry. Columns 3-5 in Panel B report the 2SLS estimates of inherited land on non-agricultural occupation. Non-Ag occupation is defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. Columns 6-8 in Panel B report the 2SLS estimate of inherited land on the log of household consumption. This includes food and non-food items, and values home production at village-specific market prices. Columns 4 and 7 include fixed effects for years of education of the household head, 12 dummies, 1 year intervals. Columns 5 and 8 include fixed effects for the net dowry receipts. This calculates the net difference between dowry received and spent by the head's parents for all siblings, and then creates 19 dummies (Rs. -50,000 - Rs. 50,000+, Rs. 5000 intervals). All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic) Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

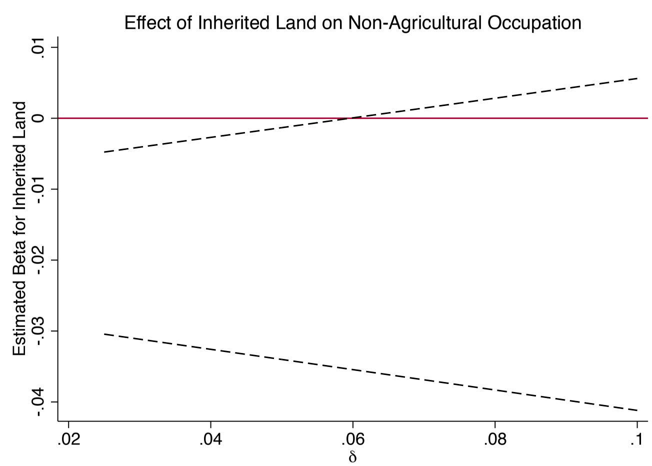
APPENDIX C6: ROBUSTNESS OF 2SLS ESTIMATES TO ALTERNATIVE DEFINITIONS OF OCCUPATIONAL CHOICE AND BIRTH ORDERS CONTROLS

Specification	Main	Definition Change	Main	Eldest Son Control	Main	Eldest Son Control
Dependent Variable	Non-Ag Occupation Binary	Non-Ag Occupation Binary	Non-Ag Occupation Binary	Non-Ag Occupation Binary	Household Consumption Log(Rs.)	Household Consumption Log(Rs.)
	(1)	(2)	(3)	(4)	(5)	(6)
Land	-0.018***	-0.016***	-0.004	-0.012	0.037***	0.055***
(Acres)	(0.005)	(0.004)	(0.006)	(0.007)	(0.007)	(0.014)
First Born	-	-	0.195***	0.173***	0.216***	0.296***
			(0.058)	(0.059)	(0.065)	(0.084)
Land*First Born	-	-	-0.045***	-0.036**	-0.033**	-0.052**
			(0.014)	(0.015)	(0.016)	(0.022)
Age FE	N	N	Y	Y	Y	Y
No. of Siblings FE	Y	Y	Y	Y	Y	Y
Family Land FE	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y
Eldest Son Dummy	N	N	N	Y	N	Y
Mean of Dep. Var.	0.298	0.265	0.298	0.265	10.442	10.442
F-stat (First Stage)	128.137	125.952	26.667	26.667	26.667	29.852
N	4809	4809	4809	4809	4809	4809
Non-Ag Definition	Primary Status	Majority Income	Primary Status	Primary Status	-	-
Eldest Son Dummy	No	No	No	Yes	No	Yes

Notes:
This table tests the robustness of 2SLS estimates of the effect of inherited land on occupational choice to an alternative definition of occupational choice (Columns 1 and 2). This table also tests the robustness of heterogeneous effects by birth order to additional controls. The sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey. The data is at the household head level. Column 1 reports the 2SLS estimates using the main specification with the instrument as predicted share. In column 1, Non-Ag occupation is defined as the 'primary status' reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. Column 2 reports estimates using the definition of Non-Ag occupation as 'majority income'. This is defined as whether the majority of the respondent's income comes from activities not related to self-cultivation or agricultural labor using income data. Columns 3-6 report the 2SLS coefficients on inherited land, first-born - a dummy coded as 1 if the household head was the first-born child in his family - and their interaction. The two endogenous variables are instrumented with two instruments : Predicted Share =(1/1+Brothers) and the interaction between Predicted Share and First Born. These coefficients are from the main specification. Columns 4 and 6 include a dummy variable coded as 1 if the respondent is the eldest son but not the first born in the family. The dependent variable in columns 3 and 4 is Non-Ag occupation and is defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. The dependent variable columns 5 and 6 is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. All specifications include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies) and the number of siblings (14 dummies). The excluded group are heads who are only children from West Godavari district in Andhra Pradesh with family landholdings between 0-5 acres. The F-statistic reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Robust standard errors are given in parentheses, asterisks denote significance: * p<0.10, ** p<0.05, *** p<0.01. Data Source: ARIS-REDS Dataset.

APPENDIX C8: SENSITIVITY OF REDUCED FORM ESTIMATES TO VIOLATION OF EXCLUSION RESTRICTION

Panel A



Panel B

Notes:

These figures show how large the exclusion restriction violation would need to be in order to invalidate the reduced form results. Both panels use the 'union of confidence intervals' (Conley et al. 2012) approach to estimate the size of the exclusion restriction needed to violate the reduced form for non-agricultural occupation (Panel A) and household consumption (Panel B). The dashed lines plot the union of confidence intervals where the parameter δ varies the possible values of γ the size of the exclusion restriction violation such that $\gamma \in [-2\delta, 2\delta]$. These figures were produced using the 'plausexog' code produced by Damian Clarke (2014). Data Source: ARIS-REDS Dataset.

Appendix C7: Post Reform Areas and Urban Areas

Reforms to the Hindu Succession Act

Reforms to inheritance laws giving women equal inheritance rights were passed in Kerala (1976), Andhra Pradesh (1986), Tamil Nadu (1989), Maharashtra (1994) and Karnataka (1994). In the main sample, there are 1,506 respondents residing in these states, 473 of whom inherited land after the reforms. Column (1) in Appendix C7 shows that the reforms significantly reduce the influence of sibling sex composition on inherited landholdings, consistent with Deininger *et al.* (2013). However, sibling sex composition still has an appreciable effect on the inheritance of land, consistent with the primacy of customs over law. The point estimate for consumption in column (7) is qualitatively different for post-reform respondents, but imprecisely estimated. In contrast, the effects for non-agricultural occupation does not vary substantially, but the null cannot be rejected in the post-reform sample.

Urban versus Rural Households

The absence of urban agricultural land among urban households means that the effect of sibling sex composition cannot, by definition, operate through this channel in urban areas. In addition to the absence of agricultural land, urban households are more likely to have recourse to professional legal services, rather than have property disputes settled by male-biased village councils (Rao, 2007). Taken together, this suggests that sibling sex composition may have a smaller role in influencing the size of one's inheritance. The IHDS contains a sample of 16,205 households across rural and urban India for whom sibling sex composition data is available. Column (4) in Appendix C7 tests whether the reduced form effects for household consumption vary across urban and rural areas. The point estimate for consumption in urban areas is very different in magnitude to rural areas and the null hypothesis cannot be rejected in the urban sample.

APPENDIX C7: THE FIRST STAGE AND REDUCED FORM EFFECTS FOR REFORM AREAS AND URBAN AREAS

	Before vs. After Reform			Rural vs. Urban
Dependent Variable	Inherited Land (Acres)	Non-Ag Occupation (Binary)	HH Consumption Log (Rs.)	HH Consumption Log (Rs.)
	(1)	(2)	(3)	(4)
Predicted Share	7.674***	-0.067	0.350***	0.082***
	(1.327)	(0.062)	(0.082)	(0.028)
Dummy for Restricted Sample (Post-Reform/Urban)	0.789	0.044	0.011	0.278***
	(0.559)	(0.047)	(0.053)	(0.024)
Predicted Share*Dummy	-2.054*	0.007	-0.124	-0.060
	(1.229)	(0.090)	(0.095)	(0.037)
Family Land FE	Y	Y	Y	N
No. of Siblings FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Father's Occupation FE	N	N	N	Y
Age FE	N	N	N	Y
Education FE	N	N	N	Y
Can Reject Null Hypothesis in	Yes	No	Yes	No
Depvar Mean	3.746	0.253	10.399	10.575
F-statistic	22.404	-	-	-
N	1506	1506	1506	16205
Data Source	ARIS-REDS, 1999 Wave			IHDS, 2004-05

Notes:
This table tests whether the first stage and reduced form effects for non-agricultural occupation and consumption vary differentially within states, before and after progressive reforms to patrilineal laws were instituted. It also tests whether the reduced form effect on consumption varies among urban and rural areas. These tests are intended to provide support for the exclusion restriction assumption. The data in Columns 1-3 the sample is restricted to Hindu male household heads whose parents owned land in the 1999 ARIS-REDS's survey in states that experienced reforms for the Hindu Succession Act. In Column 4 the data used is from the Indian Human Development Survey and includes all Hindu household heads in rural and urban areas. In Columns 1-3 'Dummy for Restricted Sample' is a dummy variable coded as 1 if the respondent became the head of the household after reforms to inheritance laws and 0 if before. In column 4, it is a dummy coded as 1 if the head resides in an urban area and 0 if the head resides in a rural area. Reforms to the Hindu Succession Act occurred in Kerala (1976), Andhra Pradesh (1986), Tamil Nadu (1989), Maharashtra (1994) and Karnataka (1994), where the parentheses indicate the date of the reform The dependent variable in Column 1 is inherited land. The dependent variable in Column 2 is whether the head held a non-agricultural occupation, defined by the primary status reported by the respondent in the REDS survey. The variable is coded as 0 if this is self-cultivation or agricultural labor and 1 otherwise. In Column 3 and 4 the dependent variable is the natural logarithm of yearly household consumption. This includes food and non-food items, and values home production at village-specific market prices. Specifications 1-3 include district fixed effects (99 dummies), fixed effects for family landholdings (0-80+ acres, 5 acre intervals, 15 dummies and the number of siblings (14 dummies). Specification 4 includes district fixed effects (99 dummies), number of head's siblings (19 dummies), fixed effects for head's education (0-15 years, 1 year intervals, 14 dummies), age of the household head (20-100, 5 year intervals, 15 dummies), father's occupation (89 dummies), and father's education (0-15, years, 1 year intervals, 14 dummies). Parent's landholdings are not reported in the IHDS data. The F-stat reported is the partial F-statistic for the instrument(s) (Cragg-Donald Wald F-statistic). Data Source: ARIS-REDS Dataset.

Appendix D

Appendix to Chapter 2

D.1 Supplementary Tables and Figures

APPENDIX TABLE D1: TOPICS OF QUESTION ASKED AND PUSH CALLS

	No. of Questions		% of Total Questions		No. of Push Calls		% of Total Push Calls	
Cell Contents	Midline	Endline	Midline	Endline	Midline	Endline	Midline	Endline
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. By Crop								
Cotton	679	960	0.50	0.46	30	59	0.68	0.62
Cumin	80	151	0.06	0.07	15	36	0.34	0.38
Wheat	26	43	0.02	0.02	11	27	0.25	0.28
B. By Theme								
Pest Management	739	1126	0.54	0.54	23	73	0.52	0.77
Crop Planning	197	363	0.14	0.17	30	64	0.68	0.67
Fertilizer	106	154	0.08	0.07	13	32	0.30	0.34
Weather	66	88	0.05	0.04	10	26	0.23	0.27
Irrigation	12	21	0.01	0.01	2	5	0.05	0.05
N	1370	2079			44	95		

Notes
1. This table reports information on push calls and questions asked on the AO server, categorized by crop and theme.
2. All push calls contain information on multiple themes.
3. A total of 95 push calls were sent out during September 2011- August 2013, with an average length of approximately 5 minutes.
4. The midline survey took place between 4th June and 8th July 2012.
5. The Endline survey took place between 23rd July and 30th August 2013.

APPENDIX TABLE D2: EFFECTS ON SOURCES OF INFORMATION BY SOURCE AND DECISION TYPE

Source of Information	Past Experience			Mobile Phone-Based Information			Other Farmers			Input Dealers		
Dependent Variable	Control Mean (Baseline) (1)	Treat-Control ITT (Midline) (2)	Treat-Control ITT (Endline) (3)	Control Mean (Baseline) (4)	Treat-Control ITT (Midline) (5)	Treat-Control ITT (Endline) (6)	Control Mean (Baseline) (7)	Treat-Control ITT (Midline) (8)	Treat-Control ITT (Endline) (9)	Control Mean (Baseline) (10)	Treat-Control ITT (Midline) (11)	Treat-Control ITT (Endline) (12)
Decision Type												
Cotton Planting	0.020	0.008	0.008	0.000	0.074***	0.086***	0.296	0.017	0.034	0.080	-0.017	-0.010
	(0.141)	(0.015)	(0.012)	(0.000)	(0.010)	(0.011)	(0.457)	(0.039)	(0.038)	(0.272)	(0.023)	(0.023)
Cotton Fertilizers	0.020	0.011	0.003	0.003	0.062***	0.071***	0.227	-0.001	0.035	0.099	0.015	0.050**
	(0.142)	(0.010)	(0.012)	(0.051)	(0.009)	(0.010)	(0.419)	(0.032)	(0.032)	(0.300)	(0.021)	(0.022)
Cotton Pesticides	0.023	-0.001	0.011	0.000	0.152***	0.162***	0.399	-0.031	0.007	0.440	-0.072*	-0.037
	(0.149)	(0.013)	(0.012)	(0.000)	(0.013)	(0.014)	(0.490)	(0.039)	(0.037)	(0.497)	(0.041)	(0.037)
Wheat Planting	0.008	-0.002	0.003	0.000	0.112***	0.022***	0.116	-0.016	0.004	0.023	-0.003	0.005
	(0.087)	(0.008)	(0.007)	(0.000)	(0.012)	(0.007)	(0.320)	(0.023)	(0.023)	(0.149)	(0.011)	(0.011)
Wheat Fertilizers	0.005	-0.003	0.003	0.000	0.014***	0.009**	0.111	0.009	0.004	0.055	0.014	0.020
	(0.071)	(0.006)	(0.006)	(0.000)	(0.004)	(0.004)	(0.314)	(0.021)	(0.021)	(0.229)	(0.015)	(0.014)
Wheat Pesticides	0.005	0.001	0.004	0.000	0.012***	0.003	0.023	0.004	-0.006	0.013	0.004	0.003
	(0.071)	(0.004)	(0.005)	(0.000)	(0.004)	(0.004)	(0.149)	(0.011)	(0.010)	(0.112)	(0.010)	(0.008)
Cumin Planting	0.003	-0.002	-0.005	0.000	0.153***	0.049***	0.093	-0.019	-0.027	0.065	0.019	0.037**
	(0.050)	(0.008)	(0.007)	(0.000)	(0.013)	(0.009)	(0.291)	(0.024)	(0.023)	(0.247)	(0.018)	(0.017)
Cumin Fertilizers	0.000	-0.008*	-0.001	0.000	0.015***	0.016***	0.068	-0.005	0.020	0.025	-0.010	-0.000
	(0.000)	(0.004)	(0.002)	(0.000)	(0.004)	(0.006)	(0.252)	(0.018)	(0.017)	(0.157)	(0.011)	(0.011)
Cumin Pesticides	0.005	0.000	0.002	0.000	0.029***	0.043***	0.126	0.001	0.018	0.133	-0.026	-0.003
	(0.071)	(0.006)	(0.005)	(0.000)	(0.006)	(0.009)	(0.332)	(0.024)	(0.023)	(0.340)	(0.027)	(0.024)
N	398	2323	2280	398	2323	2280	398	2323	2280	398	2323	2280

Notes

1. This table reports the impact of AO on usage of different information sources for agricultural decision-making over time.
2. First, we ask survey participants if they received information for a particular decision making category. Then, participants were asked to name their most important source for this category.
3. 'Treat' group refers to the 802 farmers that received access to AO.
4. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
5. Column 1, 4, 7 and 10 provides the mean and standard deviation for the control group by information source at baseline.
6. Columns 2-3, 5-6, 8-9 and 11-12 report the Intention to Treat (ITT) estimate of the difference in means (and robust standard error) between the treatment groups and control group by information source.
7. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

APPENDIX TABLE D3: CHARACTERISTICS OF ATTRITORS BY TREATMENT STATUS

Dependent Variable	Control Mean (Midline) (1)	Treat-Control (Midline) (2)	Control Mean (Endline) (3)	Treat-Control (Endline) (4)
Age of Household Head	44.174	1.151	47.090	-1.467
	(11.116)	(3.791)	(13.173)	(2.819)
Years of Education	2.696	0.865	4.077	-0.086
	(3.470)	(1.243)	(4.138)	(0.989)
Agricultural Income (log rupees)	10.745	1.269	11.628	0.198
	(2.677)	(0.880)	(1.033)	(0.235)
Planted Cotton	1.000	-0.045	0.974	0.014
	(0.000)	(0.056)	(0.160)	(0.034)
Total Area, Cotton (Acres)	4.304	0.663	4.859	1.216
	(4.085)	(0.824)	(4.454)	(0.914)
Planted Wheat	0.826	-0.285	0.744	-0.054
	(0.388)	(0.184)	(0.442)	(0.109)
Total Area, Wheat (Acres)	1.617	-0.350	1.121	-0.278
	(1.892)	(0.655)	(1.555)	(0.291)
Planted Cumin	0.391	-0.024	0.308	0.114
	(0.499)	(0.172)	(0.468)	(0.115)
Total Area, Cumin (Acres)	1.449	-0.886	0.559	0.082
	(3.307)	(1.123)	(1.388)	(0.310)
N	23	77	39	120

Notes

1. This table compares baseline characteristics of respondents of attritors from the midline and endline.
2. Agricultural income refers to income earned from all crops from the past 12 months.
3. Columns 1-2 compare baseline characteristics (from 2010) for the 23 control group respondents and 54 treatment group respondents were not reached during the midline survey.
4. Columns 3-4 compare baseline characteristics for the 39 control group respondents and 81 respondents were not reached during the endline survey.
5. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
6. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

APPENDIX TABLE D4: BALANCE FOR PEER REGRESSIONS

Dependent Variable	Treatment Group: Control Peer Group	Treatment Group: Fraction of Peers Treated	Non-Study Peers: Control Peer Group	Non-Study Peers: Fraction of Peers Treated
	(1)	(2)	(3)	(4)
A. Sample Size				
Entire Sample	654	802	393	1114
B. Individual Characteristics				
Age	36.245	-0.111	33.232	0.226
	(10.608)	(1.596)	(9.706)	(0.715)
Years of Education	4.064	-0.732	5.321	-0.017
	(3.948)	(0.515)	(4.217)	(0.317)
Landholdings- Acres	6.331	-0.855	6.681	-0.004
	(6.153)	(0.680)	(10.534)	(0.007)
C. Historical Agricultural Activity, 2010				
Planted Cotton	0.983	-0.007	0.781	0.026
	(0.129)	(0.015)	(0.414)	(0.026)
Area Cotton Planted - Acres	4.960	-0.528	4.111	0.846**
	(4.380)	(0.564)	(6.002)	(0.401)

Notes:

1. This tables assesses whether the fraction of one's peers assigned to the treatment group is independent of observable characteristics preceeding the treatment.

2. Column 1 reports the mean and standard deviation for all treated respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment.

3. Column 3 reports the mean and standard deviation for peers who were not respondents in the main study and who were not referenced by a treatment respondent.

4. Column 2 & 4 report the coefficient on the number of peers who were assigned to the treatment group, from a regression of the characteristic in question on this variable.

5. The regresson specifications in column 2 & 4 include dummies for the number of peers referenced, the amount of cotton grown at baseline, and village fixed effects.

6. Asterisks denote statistical significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$

APPENDIX TABLE D5: SUMMARY STATISTICS AND BALANCE BY TREATMENT GROUP (AO AND AOE)

Dependent Variable	Control Mean (1)	AOE Mean (2)	AOE-C ITT (3)	AO Mean (4)	AO-C ITT (5)
A. Sample Size					
Entire Sample	398	403	801	399	797
B. Individual Characteristics					
Age	46.539	45.692	-0.832	46.610	0.098
	(15.161)	(14.687)	(1.049)	(14.995)	(1.061)
Years of Education	4.235	4.089	-0.118	3.967	-0.252
	(3.836)	(3.897)	(0.263)	(3.970)	(0.272)
Landholdings- Acres	6.077	6.332	0.232	6.017	-0.044
	(5.596)	(5.874)	(0.385)	(6.179)	(0.396)
Agricultural Income	11.551	11.528	-0.022	11.558	0.011
(log rupees)	(1.361)	(1.470)	(0.098)	(1.465)	(0.099)
C. Indices (standard deviation units)					
Mobile Phone-based	0.000	-0.001	0.000	0.050	0.050
Information Usage	(0.704)	(0.702)	(0.050)	(0.995)	(0.061)
Cotton Management	0.000	-0.005	-0.004	0.007	0.007
	(0.289)	(0.281)	(0.020)	(0.335)	(0.022)
Wheat Management	0.000	-0.043	-0.042	-0.005	-0.006
	(0.433)	(0.329)	(0.027)	(0.449)	(0.030)
Cumin Management	0.000	0.007	0.011	-0.019	-0.019
	(0.347)	(0.577)	(0.032)	(0.317)	(0.022)
Pesticide Management	0.000	-0.012	-0.011	-0.002	-0.002
	(0.303)	(0.309)	(0.021)	(0.324)	(0.021)
Fertilizer Management	0.000	0.006	0.009	-0.014	-0.015
	(0.306)	(0.500)	(0.028)	(0.354)	(0.022)
Seed Management	0.000	-0.002	-0.002	0.038	0.037*
	(0.244)	(0.440)	(0.025)	(0.359)	(0.021)
D. Historical Agricultural Activity					
Planted Cotton (2010)	0.985	0.988	0.002	0.977	-0.008
	(0.122)	(0.111)	(0.008)	(0.149)	(0.009)
Area Cotton Planted - Acres (2010)	4.448	4.736	0.277	5.014	0.573**
	(3.622)	(4.426)	(0.281)	(4.045)	(0.269)
Planted Cotton (2011)	0.984	0.984	0.000	0.986	0.003
	-0.126	(0.125)	(0.009)	(0.116)	(0.009)
Area Cotton Planted - Acres (2011)	4.990	5.417	0.382	5.204	0.221
	(3.846)	(4.852)	(0.308)	(4.006)	(0.281)
Planted Wheat (2010)	0.776	0.722	-0.053*	0.724	-0.053*
	(0.417)	(0.449)	(0.029)	(0.447)	(0.029)
Area Wheat Planted- Acres (2010)	1.171	1.067	-0.100	1.314	0.135
	(1.346)	(1.248)	(0.088)	(2.180)	(0.123)
Planted Cumin (2010)	0.425	0.412	-0.014	0.401	-0.023
	(0.495)	(0.493)	(0.033)	(0.491)	(0.032)
Area Cumin Planted - Acres (2010)	0.762	0.705	-0.057	0.789	0.018
	(1.406)	(1.343)	(0.095)	(1.499)	(0.097)

Notes

1. This table reports summary statistics and balance by treatment group using data from the baseline survey, conducted between June 26 and August 11, 2011.

2. Participants were randomized into three groups. AO group received AO access. AOE group received AO access and physical extension. The control group received neither treatment.

3. The indices aggregate information over multiple outcomes for which we expect unidirectional treatment effects. Each index consists of the average of the z-scores for each component of the index, with the control group mean and standard deviation as reference.

4. Mobile phone-based information usage index: Aggregates mobile phone use across crop decision, soil preparation, pest identification, weather, cotton pesticides, cotton fertilizers, wheat fertilizers, cumin pesticides and cumin fertilizers.

5. Management practices indices: seed usage + pesticide purchase + pesticide usage + pesticide quantities + pesticide expenditure + fertilizer purchase + fertilizer usage + fertilizer quantities + fertilizer expenditure for the three different crops – cotton, wheat and cumin.

6. Pesticide management index: dummy to indicate purchase/use of a pesticide + pesticide expenditure + pesticide quantities across purchase and usage decisions.

7. Fertilizer management index: dummy to indicate purchase/use of a fertilizer + fertilizer expenditure + fertilizer quantities across purchase and usage decisions.

8. Seed management index: dummy to indicate purchase/use of recommended seeds + seed expenditure + seed quantities across purchase and usage decisions.

9. Agricultural income refers to income earned from all crops over the past 12 months

10. Column 1, 2 and 4 show the summary statistics (mean and standard deviation) for the control group at baseline.

11. Columns 3 and 5 report an Intention to Treat (ITT) estimate of the difference in means (and robust standard error) between the treatment groups (AO and AOE) and the control group.

12. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

APPENDIX TABLE D6: MAIN EFFECTS BY TREATMENT GROUP (AO and AOE)

Cells Contents:	Control Mean (Baseline) (1)	AOE vs. Control (Midline) (2)	AOE vs. Control (Endline) (3)	AO vs. Control (Midline) (4)	AO vs. Control (Endline) (5)
		Difference-in-Difference Estimates (Treat*Post Coefficient)			
A. AO Usage					
Called in to the AO line	0.000	0.610***	0.685***	0.589***	0.664***
	(0.000)	(0.025)	(0.024)	(0.025)	(0.024)
Total duration of calling in time (Minutes)	0.000	114.890***	143.016***	76.009***	98.149***
	(0.000)	(17.665)	(22.442)	(9.941)	(11.739)
B. Indices (standard deviation units)					
Mobile Phone-Based Information Usage	-0.000	2.245***	1.380***	1.404***	1.127***
	(0.704)	(0.231)	(0.133)	(0.196)	(0.128)
Cotton Management	0.000	0.065*	0.074**	0.033	0.048
	(0.289)	(0.037)	(0.034)	(0.030)	(0.035)
Wheat Management	-0.000	0.052	0.050	0.143*	0.027
	(0.433)	(0.077)	(0.041)	(0.084)	(0.043)
Cumin Management	0.000	-0.015	0.012	0.081	0.110*
	(0.347)	(0.051)	(0.045)	(0.051)	(0.062)
Pesticide Management	0.000	0.022	0.056	0.040	0.059
	(0.303)	(0.034)	(0.037)	(0.039)	(0.040)
Fertilizer Management	0.000	0.044	0.033	0.091*	0.100
	(0.306)	(0.054)	(0.037)	(0.049)	(0.067)
C. Sowing Decisions					
Planted Cotton	0.985***	-0.002	-0.002	0.012	-0.002
	(0.122)	(0.012)	(0.017)	(0.013)	(0.019)
Planted Wheat	0.776*	0.044	0.080*	0.047	0.056
	(0.417)	(0.044)	(0.045)	(0.045)	(0.045)
Planted Cumin	0.425	-0.031	-0.000	-0.003	-0.003
	(0.495)	(0.045)	(0.046)	(0.045)	(0.046)
D. Agricultural Outcomes (Input Expenditure, Profit, Yield)					
Total Input Expenditure (log rupees)	9.682	0.164	0.173	-0.185	-0.005
	(0.766)	(0.182)	(0.240)	(0.201)	(0.244)
Total profit (log rupees)	11.466	0.071	0.074	0.101	0.105
	(1.015)	(0.095)	(0.152)	(0.090)	(0.134)
Cotton Yield (kg/acre)	694.819	14.968	4.358	49.256	35.137
	(468.752)	(39.144)	(39.568)	(37.905)	(36.641)
Wheat Yield (kg/acre)	981.132	-90.107	1.048	25.363	-81.650
	(702.002)	(88.367)	(85.476)	(95.205)	(79.429)
Cumin Yield (kg/acre)	172.570	-20.500	29.530	5.614	70.126***
	(191.017)	(29.383)	(27.173)	(24.958)	(26.755)
E. Agricultural Knowledge					
Total Correct Answers (44 questions)	14.156	0.315	0.523	0.661	1.000*
	(5.279)	(0.494)	(0.568)	(0.517)	(0.583)
N	398	1557	1525	1539	1512

Notes

1. This table reports the main treatment effects for the two treatment groups – AOE and AO.
2. The AOE group refers to the 403 treatment respondents that had access to AO along with physical extension. AO group refers to the 399 treatment respondents that only had access to AO.
3. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
4. The indices aggregate information over multiple outcomes for which we expect unidirectional treatment effects. Each index consists of the average of the z-scores for each component of the index, with the control group mean and standard deviation as reference.
5. Mobile phone-based information usage index: Aggregates mobile phone use across crop decision, soil preparation, pest identification, weather, cotton pesticides, cotton fertilizers, wheat fertilizers, cumin pesticides and cumin fertilizers.
6. Management practices indices: seed usage + pesticide purchase + pesticide usage + pesticide quantities + pesticide expenditure + fertilizer purchase + fertilizer usage + fertilizer quantities + fertilizer expenditure for the three different crops – cotton, wheat and cumin.
7. Pesticide management index: dummy to indicate purchase/use of a pesticide + pesticide expenditure + pesticide quantities across purchase and usage decisions
8. Fertilizer management index: dummy to indicate purchase/use of a fertilizer + fertilizer expenditure + fertilizer quantities across purchase and usage decisions
9. Seed management index: dummy to indicate purchase/use of recommended seeds + seed expenditure + seed quantities across purchase and usage decisions
10. Total input expenditure refers to money spent on seeds, fertilizers, pesticides and irrigation in the past year.
11. Profit from agriculture is the difference between revenue from all crops and total input expenditure in the past year.
12. Respondents were asked agricultural questions across crop and topic, and a knowledge score was computed based on the proportion of correct answers.
13. Column 1 provides the mean and standard deviation for the control group at baseline.
14. Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.
15. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

APPENDIX TABLE D7: BREAKDOWN OF INPUT ADOPTION DECISIONS

Dependent Variable	Control Mean	Difference-in-Difference Estimates (Treat*Post Coefficient)			
		Treat vs. Control	Treat vs. Control	Treat+Reminder vs. Control	Treat+Reminder vs. Control
	(Baseline)	(Midline)	(Endline)	(Midline)	(Endline)
	(1)	(2)	(3)	(4)	(5)
A. Pesticides					
Index for pesticide management	0.000	0.031	0.057*	0.028	0.041
practices	(0.303)	(0.030)	(0.032)	(0.033)	(0.035)
Purchased monocrotophos	0.854	0.001	-0.025	-0.010	-0.018
	(0.353)	(0.024)	(0.028)	(0.026)	(0.030)
Quantity of monocrotophos purchased	2.878	0.087	0.086	0.071	0.084
(liter)	(3.756)	(0.276)	(0.230)	(0.307)	(0.260)
Total spent on monocrotophos	2.856	-0.348	-0.258	-0.456	-0.262
(log rupees)	(3.323)	(0.250)	(0.264)	(0.278)	(0.289)
Used monocrotophos	0.721	-0.011	-0.030	-0.010	-0.018
	(0.449)	(0.029)	(0.033)	(0.032)	(0.036)
Quantity of monocrotophos used	2.259	0.061	0.058	0.218	0.202
(liter)	(3.383)	(0.281)	(0.239)	(0.313)	(0.268)
Purchased imidachloropid	0.480	0.029	0.031	0.037	0.038
	(0.500)	(0.042)	(0.042)	(0.046)	(0.046)
Quantity of imidachloropid purchased	0.498	0.043	0.034	0.098	0.058
(liter)	(0.972)	(0.095)	(0.075)	(0.110)	(0.084)
Total spent on imidachloropid	2.588	-0.032	0.078	0.094	0.228
(log rupees)	(3.370)	(0.297)	(0.275)	(0.327)	(0.302)
Used imidachloropid	0.440	0.021	0.022	0.027	0.025
	(0.497)	(0.042)	(0.042)	(0.046)	(0.046)
Quantity of imidachloropid used	0.208	0.054	0.056	0.131	0.110
(liter)	(0.865)	(0.092)	(0.073)	(0.100)	(0.075)
N	398	2323	2280	1743	1716

Notes

1. This table reports a detailed break-down of the impact of AO on purchase and usage decisions for agricultural inputs, aggregated across cotton, cumin and wheat.
2. The indices aggregate information over multiple outcomes for which we expect unidirectional treatment effects. Each index consists of the average of the z-scores for each component of the index, with the control group mean and standard deviation as reference.
3. Pesticide management index: dummy to indicate purchase/use of a pesticide + pesticide expenditure + pesticide quantities across purchase and usage decisions.
4. Fertilizer management index: dummy to indicate purchase/use of a fertilizer + fertilizer expenditure + fertilizer quantities across purchase and usage decisions.
5. Seed management index: dummy to indicate purchase/use of recommended seeds + seed expenditure + seed quantities across purchase and usage decisions.
6. `Treat' group refers to the 802 farmers that received access to AO.
7. `Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
8. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
9. Column 1 provides the mean and standard deviation for the control group at baseline.

10. Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.

11. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

APPENDIX TABLE D8: BREAKDOWN OF INPUT COSTS

Dependent Variable	Control Mean (Baseline) (1)	Difference-in-Difference Estimates (Treat*Post Coefficient) Treat vs. Control (Midline) (2)	Treat vs. Control (Endline) (3)	Treat+Reminder vs. Control (Midline) (4)	Treat+Reminder vs. Control (Endline) (5)
A. Acreage					
Total Acres of Cotton	4.448	-0.031	0.115	-0.071	0.250
	(3.622)	(0.181)	(0.182)	(0.198)	(0.209)
Total Acres of Wheat	1.171	0.050	0.086	0.022	0.079
	(1.346)	(0.092)	(0.089)	(0.109)	(0.106)
Total Acres of Cumin	0.762	0.055	0.005	0.060	0.012
	(1.406)	(0.109)	(0.113)	(0.122)	(0.126)
B. Input Costs					
Total input expenditure	9.682	-0.012	0.082	-0.028	0.266
(log rupees)	(0.766)	(0.164)	(0.210)	(0.181)	(0.223)
Expenditure on seeds	6.105	0.071	0.309	0.023	0.416
(log rupees)	(3.359)	(0.281)	(0.271)	(0.311)	(0.298)
Expenditure on fertilizers	8.914	0.085	-0.017	0.091	0.014
(log rupees)	(0.898)	(0.102)	(0.145)	(0.110)	(0.156)
Expenditure on pesticides	6.587	-0.212	-0.089	-0.255	-0.022
(log rupees)	(2.725)	(0.213)	(0.240)	(0.234)	(0.258)
Expenditure on irrigation	4.821	1.009***	0.605	1.118***	0.817**
(log rupees)	(4.469)	(0.376)	(0.369)	(0.414)	(0.404)
N	398	2323	2280	1743	1716

Notes

1. This table reports a detailed break-down of the input costs, aggregated across cotton, cumin and wheat.

2. All expenditure figures are computed on a yearly basis. Total input expenditure refers to money spent on pesticides, fertilizers, seeds and irrigation.

3.`Treat' group refers to the 802 farmers that received access to AO.
4.`Reminder' group refers to the 502 treatment farmers that also received bi-weekly calls reminding them to call in to the AO line.
5. The midline survey took place between 4th June and 8th July 2012. The Endline survey took place between 23rd July and 30th August 2013.
6. Column 1 provides the mean and standard deviation for the control group at baseline.

7. Columns 2-5 report the coefficient on the interaction term between a dummy for treatment and a dummy for the 'post' variable from a difference-in-difference specification. All specifications include village fixed effects, a control for the amount of baseline cotton grown and its interaction with the post variable.

8. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01

APPDENDIX D9: KNOWLEDGE INDEX QUESTIONS

The following are the agricultural questions used to gauge agricultural knowledge. The analysis of this index is presented in Table 6.

A. General

Q1. Which essential plant nutrients does urea contain?
Q2. Which is the best fertilizer for adding phosphorus in the soil?
Q3. If you had the option of using 50 kg (1 bag) of Di-Ammonium Phosphate (DAP) or 50 kg (1 bag) of 20-20-20 grade NPK fertilizer, which would you use to add phosphorus to the soil?
Q4. Which is the best fertilizer for adding potash in the soil?
Q5. If you had the option of using 50kg (1 bag) of Murate of Potash or 50kg (1 bag) of 12-32-36 grade NPK fertilizer, which would you use to add potash in the soil?
Q6. Which is the best fertilizer for adding sulphur in the soil?
Q7. If you had the option of using 50 kg of Ammonium Sulphate or 50 kg of Sulphur fertilizer, which would you use to add sulphur to the soil?
Q8. When mixing pesticides in the pump, do you add powder concentrate or liquid concentrate first?

B. Cotton-Related Questions

Q1. What types of pests does BT cotton provide resistance against?
Q2. Do you know what a pheromone trap is?
Q3. What is the use of a pheromone trap in agriculture?
Q4. After the flowering stage, which type of fertilizers should you spray for good development of bolls and to stop falling of flower buds?
Q5. During the flowering stage, which fertillizer should you spray to stop yellowing of plants and to increase production?
Q6. Monocrotophos is used to control which pests?
Q7. Have you heard of Imidachloropid (or Confidor/Tatamida/Imidagold)
Q8. Imidachloropid (or Confidor/Tatamida/Imidagold) is used to contol which pests?
Q9. Have you heard of Acetamapride?
Q10. Acetamapride is used to control which pests?
Q11. Which pests is acephate pesticide used to control ?
Q12. If you had the option of using 1 litre of Prophanophos or 1 litre of Monocrotophos to treat Mealybug in cotton, which would you use?
Q13. If you had the option of using 1 litre of Acetamapride or 1 litre of Monocrotophos to treat Whitefly in cotton, which would you use?
Q14. If you had the option of using 1 litre of Imidachloropid or 1 litre of Monocrotophos to treat Leaf Curl or Aphid in cotton, which would you use?
Q15. If you had the option of using 1 litre of Dithan or 1 litre of Monocrotophos to treat Wilt disease in cotton, which would you use?
Q16. Which fungus or bio-product can be used with compost as a seed treatment or soil application to control Wilt disease?

C. Wheat Related Questions

Q1. What is the ideal time period for sowing of wheat?
Q2. For those practicing late sowing, wheat crop should be planted by when at the latest?
Q3. Which disease affects the grain quality, and ultimately the price of wheat grains
Q4. Which variety of wheat is recommended in Gujarat for those practicing late sowing?
Q5. What is the recommended dose of nitrogen in irrigated wheat?
Q6. What is the recommended dose of phosphorus in irrigated wheat?
Q7. After the first irrigation at the time of sowing, when should the next irrigation for wheat take place?

D. Cumin -Related Questions

Q1. Which recommended varieties of cumin are resistant to wilt?
Q2. What is the best time for planting cumin?
Q3. What should be done to cumin seeds before sowing to prevent fungal diseases?
Q4. What is the recommended dose of nitrogen for cumin?
Q5. Which fungicide is used to control the harmful effects of Wilt disease in cumin?
Q6. If you had the option of 1 kg of Mancozeb or 1 liter of Monocrotophos, which would you use to treat Wilt disease in cumin?
Q7. If you had the option of 1 kg of Sulphur or 1 liter of Monocrotophos, which would you use to treat powdery mildew in cumin?
Q8. Which herbicide is used to control weed growth in cumin?
Q9. Which fungus or bio-product can be used as a seed treatment or soil application to control Wilt disease in cumin?

Appendix E

Appendix to Chapter 3

E.1 Supplementary Tables and Figures

APPENDIX TABLE E1 : PROJECT TIMELINE

Date	Event
May/2011	Cotton planting decisions begin
May/2011	Listing for baseline survey
Jul/2011	Baseline (paper) survey
Aug/2011	AO training for treatment respondents
Aug/2011	AO service activated for all treatment respondents
Sep/2011	Reminder calls started
Nov/2011	Physical extension Round 1
Nov/2011	Phone Survey Round 1
Dec/2011	Phone Survey Round 2
Mar/2012	Peer Survey
Jun/2012	Midline (Paper) Survey

Notes:

1. Phone surveys were conducting with roughly half the treatment sample i.e. 400 respondents who had access to AO and the 398 control respondents who did not have access to AO

2. Peer surveys reached out to roughly 1000 farmers listed as farmer friends by respondents.

APPENDIX TABLE E2— CHARACTERISTICS OF PHONE SURVEY ATTRITORS BY INITIAL TREATMENT ASSIGNMENT

	Control Mean (1)	Treat-Control (2)
Age	35.37	2.22
	(9.05)	(3.16)
Years of Education	2.59	0.34
	(2.99)	(0.92)
Landholdings - Acres	4.58	-0.05
	(3.84)	(0.85)
Agricultural Income ('000s)	147.26	-30.93
	(151.72)	(33.47)
Planted Cotton in K'10	1.00	-0.03
	(0.00)	(0.04)
Area of Cotton Planted	4.33	-0.54
	(4.04)	(0.85)
Planted Wheat in K'10	0.667	0.092
	(0.480)	(0.122)
Area of Wheat Planted	1.111	0.137
	(1.649)	(0.412)
Planted Cumin in 2010	0.407	0.179
	(0.501)	(0.134)
Area Cumin Planted in 2010	0.563	0.194
	(0.808)	(0.235)
N	27	56

Notes: This table compares baseline characterstics of the 27 of 398 control, and 29 of 400 treatment individuals who could not be reached in the phone survey that concluded in December, 2012. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

APPENDIX TABLE E3: EFFECTS OF AO ON SOURCES OF INFORMATION AND PEER INTERACTIONS

Cell contents:	Control Group Mean (1)	AO Only Mean (2)	AO-Control ITT (3)
A.Type of Information Shared with Peers			
Shared info: Crop decision	0.122	0.082	-0.037*
Shared info: Field preparation	0.033	0.030	-0.004
Shared info: Seeds	0.283	0.244	-0.038
Shared info: Fertilizers	0.312	0.258	-0.053
Shared info: Pesticides	0.454	0.449	-0.005
Shared info: Irrigation	0.084	0.085	0.000
Shared info: Weather	0.005	0.011	0.006
Shared info: Harvesting	0.014	0.000	-0.014**
Shared info: Prices	0.008	0.003	-0.006
B. Type of Information Received from Peers			
Received info: Crop decision	0.114	0.055	-0.054***
Received info: Field preparation	0.038	0.016	-0.020*
Received info: Seeds	0.179	0.134	-0.045*
Received info: Fertilizers	0.204	0.151	-0.052*
Received info: Pesticides & Fungicides	0.296	0.279	-0.016
Received info: Irrigation	0.071	0.055	-0.016
Received info: Weather	0.014	0.005	-0.008
Received info: Harvesting	0.003	0.005	0.003
Received info: Prices	0.008	0.003	-0.005
C. Type of Information Observed from Peer's Field			
Observed info: Crop decision	0.049	0.008	-0.040***
Observed info: Field preparation	0.041	0.016	-0.024**
Observed info: Seeds	0.035	0.014	-0.021*
Observed info: Fertilizers	0.041	0.022	-0.020
Observed info: Pest and disease id	0.014	0.003	-0.011
Observed info: Irrigation	0.041	0.016	-0.025**
Observed info: Weather	0.005	0.000	-0.006
Observed info: Harvesting	0.000	0.000	0.000
Observed info: Prices	0.000	0.000	0.000
N	398	399	797

Notes: This table reports the effect of AO on social interactions and the exchange of information using data from the 'phone survey' conducted in Nov, 2011. 'AO' refers to 399 respondents who were assigned access to the treatment and surveyed by phone. Column 1 provides the mean and standard deviation of the control group. Column 2 provides the mean and standard deviation for the AO group. Column 3 provide an Intention to Treat (ITT) estimate of the difference in means (and the robust standard error) between the AO group and the control group. Robust standard errors used but not reported. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

APPENDIX TABLE E4 - EFFECTS OF AO ON STRUCTURE OF SOCIAL INTERACTIONS

Cell contents:	Control Group Mean (1)	AO (2)	Post (3)	AO*Post (4)
Panel A: Difference-in-Differences				
Spoke to peers more than once a month about ag	0.598	-0.008	0.226***	-0.006
	(0.491)	(0.034)	(0.031)	0.044
Spoke to a peer who is not a relative about agriculture	1.000	0.000	-0.415***	-0.010
	(0.000)	(0.005)	(0.025)	0.035
Spoke to a peer who is not a neighbor	0.879	-0.022	-0.098***	0.015
	(0.326)	(0.024)	(0.026)	0.038
Spoke to a peer on the phone about agriculture	0.033	-0.003	0.003	0.017
	(0.178)	(0.012)	(0.013)	0.019
Subjective rating of peer's agricultural knowledge	3.872	0.025	-0.071	-0.071
	(0.704)	(0.047)	0.075	0.075
N	398	-	-	1596
Panel B: Simple Differences				
Went to peer's house to discuss ag topics	0.510	-0.045	-	-
	(0.501)	(0.037)		
Peer came to own house to discuss ag topics	0.504	0.011	-	-
	(0.501)	(0.037)		
N	363	720	-	-

Notes: This table reports difference-in-difference estimates (Panel A) and simple difference estimates (Panel B) of the effect of AO on the structure of social interactions. The difference-in-differences estimates compare baseline data to data from the midline. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

APPENDIX TABLE E5 - HETEROGENEOUS EFFECTS OF AO ON STRUCTURE OF SOCIAL INTERACTIONS

Cell contents:	Control Group Mean (1)	AO*Educ (2)	AO*Inc (3)	AO*Age (4)
Shared info?	0.693	-0.001	0.000	-0.002
A. Type of Information Shared				
Crop decision	0.122	0.009	0.000	-0.003
Field preparation	0.033	-0.000	-0.000	0.000
Seeds	0.283	0.001	0.000	0.000
Fertilizers	0.312	-0.008	0.000	-0.002
Pesticides	0.043	0.000	-0.000	0.003*
Irrigation	0.084	0.001	0.000	0.001
Weather	0.005	0.002	0.000	-0.000
Harvesting	0.014	-0.001	0.000	-0.001
Prices	0.008	-0.000	-0.000	-0.000
Animal husbandry	0.003	0.001	0.000	-0.000
B. Type of Information Received				
Crop decision	0.114	-0.002	-0.000	-0.001
Field preparation	0.038	0.002	-0.000	0.001
Seeds	0.179	0.012*	-0.000	0.002
Fertilizers	0.204	0.008	-0.000	0.001
Pesticides	0.296	0.001	0.000	0.003
Irrigation	0.071	0.003	0.000	-0.000
Weather	0.014	-0.003*	0.000	0.001
Harvesting	0.003	0.000	0.000	0.001
Prices	0.008	-0.002	0.000	0.001
N	398	797	797	797

Notes: This table reports the heterogeneous effects of AO on social interactions by education (column 2), income (column 3) and age (column 4).Column 1 reports the mean and standard deviation of the control group. In the case of education a dummy is coded as 1 if a respondent has above median education. In each case, the table reports the coefficient on the interaction term in a simple difference specification. Robust standard errors are reported in parentheses. Asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

APPENDIX TABLE E6: BALANCE FOR PEER REGRESSIONS

	Treatment Group		Non-Study Peers	
Dependent Variable	Control Peer Group	Fraction of Peers Treated	Control Peer Group	Fraction of Peers Treated
	(1)	(2)	(3)	(4)
Panel A: Sample Size				
Entire Sample	654	802	393	1114
Panel B: Individual Characteristics				
Age	36.245	-0.111	33.232	0.226
	(10.608)	(1.596)	(9.706)	(0.715)
Years of Education	4.064	-0.732	5.321	-0.017
	(3.948)	(0.515)	(4.217)	(0.317)
Landholdings- Acres	6.331	-0.855	6.681	-0.004
	(6.153)	(0.680)	(10.534)	(0.007)
Panel C: Historical Agricultural Activity, 2010				
Planted Cotton	0.983	-0.007	0.781	0.026
	(0.129)	(0.015)	(0.414)	(0.026)
Area Cotton Planted (Ac	4.960	-0.528	4.111	0.846**
	(4.380)	(0.564)	(6.002)	(0.401)

Notes:
This table assesses whether the fraction of one's peers assigned to the treatment group is independent of observable characteristics preceeding the treatment. Column 1 reports the mean and standard deviation for all treated respondents who did not reference peers -- a maximum of 3 were elicited -- who were themselves assigned to the treatment. Column 3 reports the mean and standard deviation for peers who were not respondents in the main study and who were not referenced by a treatment respondent. Column 2 & 4 report the coefficient on the number of peers who were assigned to the treatment group, from a regression of the characteristic in question on this variable. The regresson specifications in column 2 & 4 include dummies for the number of peers referenced, the amount of cotton grown at baseline, and village fixed effects. Asterisks denote statistical significance: * p<0.10, ** p<0.05, *** p<0.01.