Essays on Development

by

Puja Vijesh Parmar

A dissertation accepted and approved in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in **Economics**

Dissertation Committee:
Alfredo Burlando, Chair
Shankha Chakraborty, Core Member
Michael Kuhn, Core Member
Stephen Wooten, Institutional Representative

University of Oregon

Spring 2025

License

Dissertation Abstract

Puja Vijesh Parmar

Doctor of Philosophy in Economics

Title: **Essays on Development**

This dissertation is a collection of 3 chapters on development economics pertaining to financial inclusion, gender and marriage markets and impact on farmers' acquisition of information as a result of agrarian policies. The dissertation aligns with my research interest in applied microeconometrics with a focus on development. I focus on two developing countries in my papers – Uganda and India each having salient issues which provide a lot of scope for research in development. I am particularly interested in the topics of financial inclusion and gender issues which I study in my dissertation.

The first chapter is titled 'Unlocking Potential: The Impact of Savings Groups on Household Welfare Outcomes'. In this paper, I leverage the variation in timing of entry of a savings group in a community parish to assess its impact on household welfare outcomes. I use data from Ugandan National Panel surveys in the period of 2010-2020 which was a period of rapid expansion of savings groups using TW Fixed Effects approach for my analysis. I study outcomes like household assets, earnings, bulk purchases as well as food security measures for the purpose of my paper. I find that exposure to savings groups indeed helps increase household assets, improve food diversity and provide an insurance against price volatility via the channel of bulk purchasing.

The second chapter is titled 'Impact of Groundwater Conservation Laws on Indian Farmers' Profitability and Land Productivity' which is a joint work with Pramod Dudhe[1]. In this chapter I

[1]Assistant Professor, Department of Economics and Development Studies, IIM Udaipur, Rajasthan, 313001, India, pramod.dudhe@iimu.ac.in

study the Preservation of Subsoil Water Act, which was implemented by the northwestern states of Punjab and Haryana in India with the intention of conserving groundwater resources. The aim of the act was to push monsoon-cropping season to align it with the monsoon rains which would reduce the dependency on groundwater resources by farmers. We investigate the impact of this Act on farmers' choice of input, yield and profitability outcomes. In our analysis, we combine a country-wide survey of farmers with district-level rainfall, and show that farmers' irrigation investments become negatively correlated with the early monsoon rains in states with the legislation. Despite this, yields and revenues decline, suggesting that early Monsoon rainfall is not a perfect substitute for irrigation.

The third chapter is titled 'The rise of singlehood and delayed marriages in modern India'. India has traditionally been a country where marriages are universal and mostly performed in an individual's early-to-mid 20s especially in the case of women. However, recent trends show an increase in never-married women in the age group 25-34 years and the numbers are higher for women who are highly educated professionals. Since delayed marriage has implications on a variety of factors including economic independence of women, autonomy and bargaining power at home, child-bearing decisions, etc. I study the trends in singlehood and factors affecting age at first marriage among women in India. I first perform a descriptive analysis to confirm the trends in age at first marriage. Next, I adopt a Survival Analysis approach to identify the factors that might influence women's decisions to marry late. Finally, I perform a Decomposition Analysis to understand how the opposing factors of empowerment vs social norms play out in overall marriage age decisions.

This dissertation includes previously unpublished coauthored material.

Acknowledgements

I express my deepest gratitude to my committee chair – Alfredo Burlando. His support and guidance has been the backbone of my journey through graduate school. I am also indebted to Shankha Chakraborty, Michael Kuhn, Stephen Wooten and Ed Rubins. Their insightful feedback has been invaluable, shaping different parts of this dissertation. I wish to express my appreciation for my family and friends, both within and beyond the University of Oregon, for their endless support. I would like to specifically acknowledge my mother, Indu Parmar for being my steadfast pillar and helping me achieve my dreams every step of the way.

”To my mother, Sanjay, and KK sir who kept me sane, hopeful, motivated and gave me the strength and courage to fight all odds to achieve this milestone with grace and resilience.”

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation focuses on three important topics pertaining to development- finance, gender and agriculture. Access to finance is one of the most crucial requirements for the progress of any developing country. Access to financial services plays an important role in economic development, particularly for the under-served populations. In rural areas, means to formal savings are limited and access to savings is often achieved by participating in savings groups which may provide similar benefits. In Chapter 2, I study the impact of savings groups in the context of Uganda using National Panel Survey data from 2010-2020, during a period of rapid expansion of savings groups. I use a TWFE model that exploits the variation in the timing of arrival of a savings groups in a community. I find that savings groups significantly increase the average value of the assets of households by roughly 19%, however, they do not have a significant impact on monthly household earnings. I also observe that savings groups improve the variety of foods consumed by households as well as the expenditure on food consumption. I also validate that bulk quantities do trade at a discount to small quantities, which means that households with savings would be less sensitive to price volatility if they made bulk purchases from savings. However, I do not find a significant evidence of increase in bulk purchases of storable foodstuffs due to exposure to savings groups.

Chapter 3, jointly developed with Pramod Dudhe focuses on The Preservation of Subsoil Water Act, implemented by Indian states of Punjab and Haryana, which shifted the planting of rice crop closer to the arrival of monsoon rains, to reduce the extraction of groundwater resulting from early planting. In this paper, we empirically investigate the impact of this law on input choices, production, crop yield, and revenue for farmers in the treatment states of Punjab and Haryana. In principle, delayed planting should allow farmers to better target inputs to growing conditions, po-

tentially leading to increased profitability. We model difference-in-differences and triple-difference specifications by exploiting variation in the localities where this law was implemented, the timing of the law and the monsoon season, and identify these effects. Our findings indicate that, as expected, irrigation input per hectare is negatively correlated and revenue per hectare is negatively correlated with policy implementation.

Late marriage (above age 25) seemed to be a common feature of modern western societies at the end of the 20th century. In recent years the proportion of unmarried females especially in the age group of 25-34 is on the rise in India, despite the stigma associated with delayed marriage and single-hood. In Chapter 4, I analyze the trends in single-hood and factors influencing the age at first marriage in India. I attempt to explore the impact of select predictors on the marital outcomes for females in the age groups 15-49 years. Higher levels of education among women, rising urbanization and modernization are observed to be key factors influencing women to push forward their marriage, although no drastic rise in single-hood is observed.

Chapter 2

Unlocking Potential: The Impact of Savings Groups on Household Welfare Outcomes

2.1 Introduction

Uganda, a landlocked country in East Africa, while rich in natural resources and cultural diversity, grapples with significant challenges related to poverty, hunger, and financial accessibility, which affects its development. According to World Bank, approximately 21.4% of the population lives below the national poverty line as of 2020, with rural areas experiencing higher rates of deprivation. Food insecurity remains a pressing issue, with 34% of Ugandans classified as food insecure, as reported by the Global Network Against Food Crises (2021). Access to nutritious food remains a critical issue as many families struggle with food insecurity, exacerbated by climate change and inconsistent agricultural yields. Financial accessibility is also a major hurdle, as only about 49% of adults have access to formal financial institutions, as noted by the Uganda Bureau of Statistics (2019). The limited banking infrastructure and high interest rates hinder many from obtaining loans or opening and maintaining savings accounts. This combination of factors hinders economic growth and development, perpetuating a cycle of deprivation, making it difficult for communities to break free from poverty and improve their living standards.

In recent years, Uganda has witnessed a significant rise in the prominence and impact of savings groups and Village Savings and Loan Associations (VSLAs). These grass-root financial mechanisms have gained remarkable traction in the country's economic landscape due to their potential to enhance financial inclusion, stimulate local economies, and improve livelihoods. Savings groups and VSLAs represent a critical innovation in the realm of financial services, particularly in rural and under-served areas, especially when access to traditional financial services has been limited, with many communities lacking proximity to formal banking institutions. Thus, savings groups and VSLAs have emerged as vital alternatives, offering community-driven solutions to address the gaps in financial accessibility and in encouraging entrepreneurial activities, savings behavior as well as bringing people out of perpetuated cycles of poverty. Understanding their role and importance is crucial for optimizing the benefits of savings groups and VSLAs and for informed policy interventions aimed at enhancing their efficacy.

I divide the paper into two sections. In the first half of the paper, my primary objective is to critically analyze the role and impact of savings groups in Uganda, examining their contributions to household welfare as measured by asset ownership value, earnings, household amenities and food security. In the second half of the paper, I am particularly interested in the mechanism through which savings groups can impact food security. Even without an increase in income, savings groups could impact food security by relaxing liquidity constraints on a household. Households could use loans or savings to take advantage of the arbitrage opportunities in the purchasing of

durable storable foodstuffs. This is because prices of basic foodstuffs have large and predictable seasonal variation and bulk quantities trade at a discount to small quantities.

My paper contributes to the growing literature that examines the impact of savings groups on economic outcomes in Uganda and the broader Sub-Saharan African context, focusing on key themes including financial inclusion, economic empowerment, and social outcomes. While previous studies have looked at the effect of VSLAs on poverty, wealth, resilience, improving food security, overall consumption smoothing, livestock holding, household business outcomes and women's empowerment, etc. (see Beaman et al., 2014, Gash and Odell, 2013) through randomized control trials and evaluation studies; using a full decade of panel data on a broader sample of households and exploring the channel of bulk purchases for food security are two novel features of my paper. The existing studies have been based on randomized control trials of limited duration or they have focused on a very specialized subset of population (see Amponsah et al., 2023) or a specific purpose of analysis (see Dupas and Robinson, 2013). Despite its increasing popularity in the rural developing world and among donors, there are only a few rigorous experimental studies that are able to address endogeneity and selection issues in program roll-out and treatment take-up to accurately measure the role of VSLAs on household well-being (Gash and Odell, 2013). To my knowledge, my paper is among the first ones to shift to a longer time horizon rich dataset spanning 2010-2020; that replicates some of the previous results from RCTs and uncovers new channels of impact. I also perform robustness checks to ensure that there is no selection in terms of where savings groups get established. I also tackle endogeneity and selection concerns by showing that NGOs randomly establish savings groups in different communities in Uganda irrespective of the community characteristics.

I find that access to savings groups significantly increases the value of assets held by households as well as improves their food security, without any real increase in monthly earnings. My findings on the role of savings groups in improving household asset ownership is consistent with the results from Annan et al., 2013. Marguerie and Premand, 2023 also observe that VSLAs have significant impact on assets but insignificant impact on earnings in their RCT intervention in Côte d'Ivoire. Burlando et al., 2020 perform a randomized control trial that enhanced financial intermediation by introducing two formal banking products–a savings account and a loan account–to existing savings groups in five districts in Uganda. They find that this linkage intermediation results in a statistically insignificant 13% increase in self-reported income which is consistent with my findings of regressing conditional monthly household earnings on treatment variable. Gash, 2017 in their report compile results from 53 studies that assess the role of savings groups on different economic and social outcomes. The report also validates that a fair amount of evidence

exists supporting that savings groups participation leads to an increase in asset ownership; while they also found a mixed evidence on consumption and food expenditure. The report also confirms that there is only a small evidence that savings groups lead to an increase in income, as also a small evidence that savings groups increase business ownership. Finally, the report also supports my findings on improvements in food variety consumed and increase in food expenditures as a result of exposure to savings groups. Frisancho and Martin, 2020 on the other hand observe an insignificant effect on household assets but significant improvements in roof quality in rural Peru using a cluster randomized control trial. My results are also consistent with the fact that the literature on saving groups rarely reports improvements in business outcomes like business expansion or increase in entrepreneurial activities (Beaman et al., 2014, Ksoll et al., 2016), with the notable exception of Karlan et al., 2017.

My paper also contributes in the literature of assessing impact of savings groups on outcomes other than household income and assets; for instance household quality and entrepreneurial outcomes. Using a cluster randomized trial, Ksoll et al., 2016 investigate the impact of VSLAs in Northern Malawi over a two year period and find evidence of positive and significant intention-to-treat effects on several outcomes, including the number of meals consumed per day, household expenditure, and the number of rooms in the dwelling, although they find no effect on the quality of floor, walls and roofs. In this paper, I extend the analysis to look at additional outcomes like quality of roof, access to electricity, own toilets and own water source, improved walls and floor quality of houses and find that savings groups do not have any significantly positive impact on those outcomes. Karlan et al., 2017 in their paper using a clustered randomized evaluation spanning three African countries (Ghana, Malawi, and Uganda), find that the promotion of community-based microfinance groups leads to an improvement in household business outcomes and women's empowerment. However, they do not find evidence of impacts on average consumption or other livelihoods. My paper looks at similar measures albeit only for Uganda over a larger time horizon. I also build on the literature studying role of savings in improving food security by virtue of greater diversity of foods consumed, improved food security and increased food expenditure/consumption (Bass et al., 2016, Kiiza and George, 2021, Beaman et al., 2014) find that savings groups have a positive and significant impact on household dietary diversity score, food consumption score, and food expenditure.

Since, I have staggered adoption of treatment, one of the potential concerns is that the fixed effects estimates may be biased, since we do not know precisely how it compares mean outcomes across groups. When there are heterogeneous treatment effects in a staggered treatment design, the standard "two-way fixed effects" (TWFE) estimator can potentially misrepresent the

true average treatment effect by providing a weighted average that may not accurately reflect the diverse impacts across different treatment groups and time periods. Thus I run Goodman Bacon Decomposition Analysis (Goodman-Bacon, 2021) combined with Heteregeneous treatment on treated estimators by Wooldridge, 2021 to deal with this issue. I observe that the magnitude and direction of estimates remain consistent with my original TWFE results, albeit they become smaller in size with less significance. I show these results in robustness section and in Appendix.

The rest of the paper is organized as follows. In the first part of the paper, section 2 provides a background on savings groups while section 3 provides information on the data used for analysis in this paper. I discuss thoroughly the methodology in section 4 followed by results from regression analysis in section 5. I extend the results from heterogeneity analysis in next section and perform robustness checks in the following one. In the second half of the paper, I study the mechanism of bulk purchases explicitly in the next section followed by results corresponding to the same. I finally conclude my paper in the final section.

2.2 Background on Savings Groups

Savings groups are community-based financial institutions which typically consist of small, self-managed assemblies of individuals who pool their resources to provide mutual financial support. They are generally comprised of 20 to 30 members, who meet weekly over the operating cycle (typically lasting one year). At the end of the cycle all funds are shared among the group's members in proportion to the amount saved during the period of operation. Members of a savings groups pool savings within the group, borrow from the group at an interest, and receive a return on their savings. Each week, members contribute funds to the group, repay previous loans, and request new loans. Hence, a member who wishes to borrow must first save. Funds that are not lent out are stored in a safe and can be lent out in the future. Finally, at the end of the operating cycle of the group, all loans are repaid and each member receives back the amount saved with the group, plus a return on her savings that depends on the total interest payment collected by the group. After the share out, the composition of the group may change, and a new cycle may start.

Thus, savings groups facilitate financial flows within local communities, and have received a lot of attention from policy makers in the past decade. Because they can be set up and maintained with minimal outside intervention, savings groups have spread very fast in sub-Saharan Africa and other developing countries. Savings groups are an innovative instrument for bringing financial inclusion to ultra-poor, vulnerable households who are usually not reached by traditional banking or microfinance interventions. For instance, the Gates Foundation has provided signifi-

cant resources to Catholic Relief Services, CARE International, and Oxfam to develop such groups in sub-Saharan Africa[1] Consequently, membership in savings groups reached 10.5 millions people worldwide in 2014, a tenfold increase relative to 2008, and continues to climb (Burlando and Canidio, 2016). In Fig. 4.1, we can see a rapid expansion in the parishes exposed to savings groups from 51% in 2010 to 86% by 2019-20 in Uganda. By facilitating access to credit and encouraging saving habits, savings groups and VSLAs contribute to increased household resilience, improved business opportunities, and enhanced community cohesion. In Uganda, the formalization of savings groups began in the 1990s, with support from NGOs and development agencies. In this paper, I empirically study how the timing of entry of savings groups in a community affects the consumption behavior and welfare of community members.

[1] See https://docs.gatesfoundation.org/Documents/one-early-success-story.pdf.

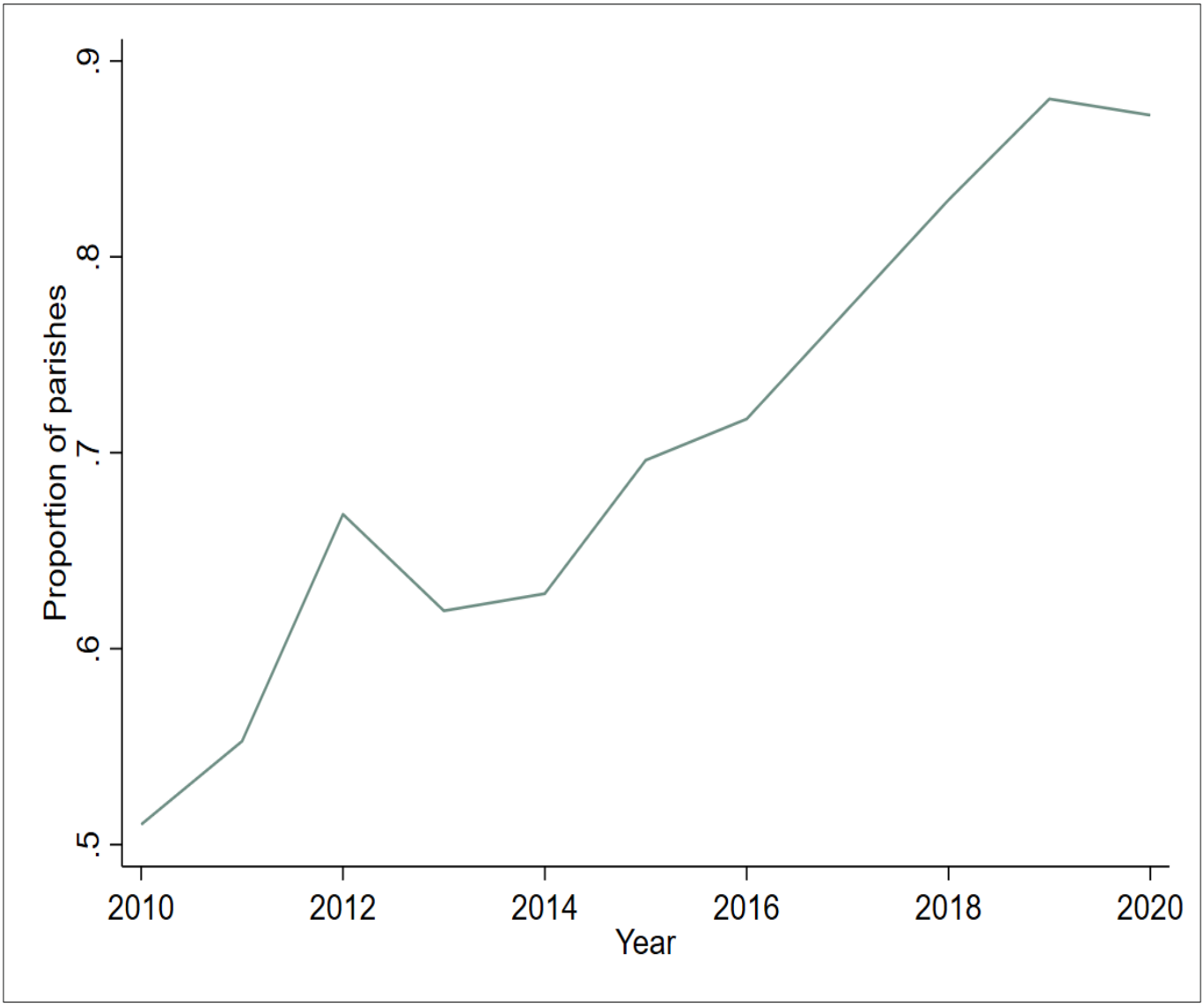

Figure 2.1: Expansion of savings group in local parish over year of survey in Uganda

(Notes: The graph shows proportion of parishes that are treated i.e. exposed to at least 1 savings group by year of interview. There was quite a bit of parish restructuring that happened across our panel survey years. This means that we have certain parishes which didn't follow through in community module post 2011-12. Also, a lot of new parishes started emerging in community module since 2013-14 for which earlier community data is not available. Thus, I do not have a consistent baseline of parishes to track the growth of savings groups exposure. However, I do have the quantitative information to make comparisons. To elaborate this further, I have 224 unique parishes in wave 2 out of which 110 were exposed to savings groups. On the other hand I have 465 parishes in wave 7 out of which 402 were exposed to savings groups. This demonstrates that expansion of savings groups has in fact be significant, since we have more total number of parishes in the most recent wave and also more significant chunk of those parishes as being exposed to savings groups.

2.3 Data

For the purpose of this study, I use data from Uganda National Panel Survey from 2010-2020. The UNPS is carried out annually, over a twelve-month period on a nationally representative sample of households, for the purpose of accommodating the seasonality associated with the composition of and expenditures on consumption. The survey is conducted in two visits in order to better capture agricultural outcomes associated with the two cropping seasons of the country. The UNPS has been set out to track and re-interview 3,123 households that were distributed over 322 enumeration areas (EAs), selected out of the 783 EAs that had been visited by the Uganda National Household Survey (UNHS) in 2005-06. The UNPS aims at producing annual estimates in key policy areas and at providing a platform for experimenting with and assessing of national policies and programs.

I gather information on household demographics (see Appendix), monthly earnings, asset ownership value as well as household amenities from the household module. I use the GDP deflator linked series and Official Exchange Rate data from World Bank to convert household (HH) assets and earnings in terms of 2017 inflation-adjusted US dollars. Each survey year collects data on the primary occupation and the secondary occupation for each member of the household. It is important to note here that different members can be earning income at different frequencies (weekly/daily/monthly/hourly) and so I first standardize the frequency of payments to 'Monthly'. I then add up the monthly earnings from both the primary and the secondary occupations for each household member. Finally, I also add up the monthly earnings for each member to give me monthly total household earnings. I deflate the monthly earnings and value of assets held by households and convert it to USD. To deal with skewness in the distribution of monthly earnings and value of household assets, I winsorize 5% from each side of the tail of the distribution for both the outcome variables.

To measure improvements in housing conditions, I construct a household quality index using data regarding the materials used in roof, floors and walls of the dwelling, the number of rooms per person in dwelling, access to electricity, own private toilets and own improved drinking water source by household. Another outcome of interest in this paper is Food Index which is constructed as follows. I first construct a food variety index which can take any value ranging from 1 to 11 depending on how many categories of food the household consumes. It is reasonable to believe that the more diverse consumption patterns a household has, the better the household ranks in terms of nutrition and food security. I also compute the total value of all food items consumed by each household in a given time period and aggregate it to get the expenditure on food which I

deflate to 2017 prices and convert to US dollars. I then combine both food variety index and food expenditure into a single standard normal variable which I call the Food Index. Simply put, higher the value on Food Index, better the household is doing in terms of food security.

It is to be noted that the UNPS survey revised the manner in which the unique HH - ID key was designed multiple times across years - this is important for us since we want to track the households right from 2010-11 until 2019-20 which means we need to have consistent HH - ID keys to panelize them. I have listed the detailed process I use to get consistent household IDs across years in Appendix.

The community section module collects information on presence and characteristics of savings groups in the communities in which the households belong. These characteristics include information on group beginning year, number of female members in the group and frequency of group meetings. Even though I cannot link directly household's participation into savings groups, I proxy it by parish exposure. I obtain parish level information on the savings groups and construct measures for timing of entry of savings groups into a parish. I then link all the households that correspond to these parishes with the savings groups information. I describe the detailed process of linking community survey with household survey in Appendix. In other words if a household belongs to a parish that got introduced to savings groups, I mark the household as 'treated' thereon. Thus, savings groups affects households by virtue of their timing of entry into a given parish.

Note that the community section collects data on a limited sample of parishes, hence for households belonging to any parish outside of this scope, I have to drop the households as they do not have any corresponding savings group information. This leaves me with roughly 400 parishes that can be linked between household and community modules and that can offer information on savings groups. In this way I retain survey data on monthly earnings, value of assets owned, primary occupation as well as other household and demographic characteristics like sex of household head, marital status, educational level, age composition of members, etc. and their corresponding community module information for 3670 unique households.

The treatment variable of interest is *'SavingsGroups'* dummy which takes the value of 1 from the time that it is reported when there is at least 1 savings group in parish and 0 otherwise. All households belonging to a given parish will experience the same treatment. Since different parishes can be exposed to savings groups in different time periods we have a staggered adoption of treatment.

2.3.1 Summary statistics

Table 3.1 reports the summary statistics of the key variables of interest. Over the period of 2010-2020, the sampled households have on average a monthly earnings of ($61). In order to prevent extreme value outliers in data I have winsorized 5% of the data from top and bottom which gives me an average monthly earnings of $27. Similarly, I observe an yearly ownership of assets worth $5031 which come down to $2659 after winsorization. Savings group have expanded to most of the parishes with an average of 74.7% expansion rate out of all the parishes in our sample.

Comparing the household characteristics shows us that roughly 33% of the households are female headed. Most of the household heads report to be married and have at least some level of positive schooling. Only about 23% of the households live in urban areas, and most of the household members belong in the age group 16-59 years. Almost 19% of the households report to be poor on average. Finally, approximately 38% of the households have some kind of business, 72% have at at least 1 member owning a mobile phone and almost 83% of the households have at least 1 member who has at least some form of savings.

The household composition is such that on average we have 1.6 members in the age group of 6-15 years, 2.25 members in the age group of 16-59 years and 0.33 members in the age group of 60 and above. The households live in a dwelling comprising of 0.57 rooms per capita on average, with 13.4% of the households having access to electricity while 19% of the households have ownership of some land. Almost 34% of the households have their own private toilet and around 32% have own source of water; while 71% of the households report having improved roof quality (non-mud, non-thatch roofs), 33% report improved floor materials and 44% report improved materials used for external wall construction. Next I also break down the dietary patterns of the households in the last 7 days into 11 different categories of consumption - on average I observe that households consume at least 8 different categories of food. They also report an average food expenditure of 21$. Finally, I observe that almost 41% of the households have faced some kind of shock.

Table 2.1: Summary Statistics

Variable	Mean	Std.Dev.	Min	Max	N
Monthly HH earnings deflated (in $)	61.352	1845.989	0	1,77,256	9,629
Monthly HH earnings deflated (in $), 5%	27.128	53.940	0	188.220	9,629
Total value of assets deflated (in $)	5,031	28,192	0	1,671,263	9,614
Total value of assets deflated (in $), 5%w	2,659	4,679	45.173	18,313	9,614
SavingsGroups in parish	0.747	0.434	0	1	9,629
HH head is female	0.335	0.472	0	1	9,629
HH head is married	0.721	0.448	0	1	9,629
HH is urban	0.229	0.420	0	1	9,629
HH head has some schooling	0.846	0.361	0	1	9,629
No. of members (6-15 yrs)	1.605	1.530	0	11	9,629
No. of members (16-59 yrs)	2.252	1.420	0	12	9,629
No. of members (60 yrs or above)	0.329	0.589	0	4	9,629
HH is poor	0.194	0.395	0	1	9,629
HH with current earnings	0.295	0.456	0	1	9,629
HH-Business dummy	0.378	0.485	0	1	9,629
Mobile ownership dummy	0.717	0.450	0	1	9,613
Savings dummy	0.831	0.374	0	1	4,415
Number of rooms per capita	0.571	0.382	0.111	3	9,583
Access to Electricity	0.134	0.341	0	1	9,616
Ownership of Land	0.188	0.391	0	1	9,613
Access to own toilet	0.340	0.474	0	1	9,614
Access to own water source	0.322	0.467	0	1	9,614
Improved roof	0.707	0.455	0	1	9,616
Improved floor	0.331	0.471	0	1	9,614
Improved wall	0.440	0.496	0	1	9,614
Amenities Index	0	1	-1.442	3.501	9,616
Food variety index	8.208	2.111	1	11	9,446
Food expenditure (in deflated $)	20.835	19.146	0.943	1,039.549	9,446
Any shock dummy	0.408	0.491	0	1	8,167
Food index	0	1	-2.693	33.220	9,446

2.4 Methodology

In this paper, I hypothesize that savings groups improve household welfare by increasing household's assets and earnings, improving household quality and enhancing food security. In order to test these hypotheses individually, I structure the model as follows:

$$\boldsymbol{y_{hpt} = \beta SavingsGroups_{pt} + f_h + f_p + f_t + \delta X_{hpt} + \epsilon_{hpt}} \tag{2.1}$$

where:

1. Household h is observed over t $>$ 1 years (it's a panel).

2. Household h is interviewed at different times of the year across different years.

3. y stands for outcome variables like total value of household assets, ownership of land dummy, total monthly household earnings, food index, household quality index, etc.

4. f_t is the year-month fixed effect. It captures the fact that aggregate household consumption may be seasonal. The time fixed effects control for factors changing each year that are common to all households and parishes for a given year.

5. f_p is the parish fixed effect. It captures the group level at which our treatment is happening. It estimates the common difference (to all years) in the outcome variable in parish p relative parish 1, controlling for household characteristics and time-specific characteristics common to all parishes. It is called parish fixed effect precisely because the difference is common to all year-months in parish p; in other words, the 'effect' of parish p is 'fixed' across all year-months, the parish fixed effects control for baseline differences between parishes.

6. f_h is the household fixed effect. It captures the household level preferences which may change across years.

7. X_{hpt} includes demographic variables like sex of household head, marital status, educational status of household head, number of household members belonging to different age categories in panel. To the extent that composition changes (say, kids age out and eat more) it can influence the decision to buy certain foods. Note that X contains both time-variant as well as time-invariant characteristics. The time-invariant characteristics will automatically drop when we incorporate household fixed effects.

8. β measures the change in outcome variable due to the treatment *'SavingsGroups'* in a parish.

I use the variation in timing of entry of savings groups in a community parish to identify

the effect on household assets, earnings, amenities and food security based on when the community parish reports entry of at least 1 savings group in it, controlling for household characteristics. The first regression in column 1 controls for time fixed effects, parish level fixed effects and household controls. Column 2 includes household fixed effects to account for within household changes across time. The standard errors are clustered at the parish level.

An additional point to consider is that sometimes savings groups could enter a parish but for a variety of reasons, they may stop operating in between in any given year - this means in a given parish once a savings group has entered it could become inoperative either temporarily or permanently. However, in my paper I do not account for this switch on and switch off of savings groups, rather I only look at the time of entry of the savings groups in a parish and then hold the treatment dummy to be 1 for all years subsequently. Thus our estimates can be interpreted as "intent to treat". In order to find the actual magnitude of treatment effect, I divide the co-efficient estimates with the imputed average membership to savings group. The imputed average membership (IAM) is calculated as number of savings groups in parish times 25 divided by number of households in parish. I then compute the average membership across all parishes across all years which gives me an average value of 0.316. This means that on average for any parish in any year we have 31.6% of the population participating in savings groups. I report the magnitude of treatment effect along with the intent-to-treat estimates in my regression results in the following subsection.

Following Anderson, 2008, sharpened False Discovery Rate q-values are computed for each family of outcomes to deal with the potential issue of simultaneous inference. The FDR is the expected proportion of rejections that are type I errors (false rejections). Suppose we have 20 hypothesis tests and accordingly 20 p-values and that none of the treatments have any effect on any outcome (all null hypotheses are true), and that the outcomes are independent. If we just test the hypotheses one by one, then the probability of one or more false rejections when using a critical value of 0.05 is 1-(0.95^{20}) = 64% (and using a critical value of 0.10 is 88%). As a result, in order to reduce the likelihood of these false rejections, we want some way of adjusting for the fact that we are testing multiple hypotheses. That is what the false discovery rate sharpened q-values method does.

2.5 Results from regression analysis

2.5.1 Household Assets and Amenities

For the first set of analysis, I regress the total value of household assets on the treatment variable. I find that wealth in the form of total value of assets owned by households is strongly associated with our treatment of savings group exposure to parish increasing total value of assets by $519.3 dollars which is 19.5% of average value of assets owned by households. This gives me an actual magnitude of treatment effect on household assets worth $1,643.33, which may seem large but it is plausible considering that treatment to savings groups often has a spillover effect in the community thereby benefiting even non-participants to savings groups. The results on household assets suggest that savings groups may facilitate the accumulation of lumpsum money required to make investments in productive assets.

I also do not find any significant effect on household quality index - this could be due to a number of reasons. For instance, external shocks or even competing priorities like health and education could disrupt savings making it difficult to invest in amenities. Similarly, lack of basic infrastructure like electricity cannot entirely be controlled or influenced by community savings. Finally, to ensure that my identification strategy is correct, I run a placebo regression of land ownership on treatment. The idea is that purchasing land is harder on just savings groups revenues, and therefore we should not observe significant increase in land ownership by households. As expected, exposure to savings groups causes a 7% increase in ownership of land which is statistically insignificant, thereby strengthening my identification.

Table 2.2: Fixed-Effects regression of total value of assets, amenities index and ownership of land on treatment

	Total HH value of assets		**Ownership of land**		**Amenities Index**	
Variables	(1)	(2)	(1)	(2)	(1)	(2)
SavingsGroups	399.556*	519.293***	0.017	0.013	-0.047	-0.024
	(206.704)	(194.194)	(0.024)	(0.024)	(0.031)	(0.028)
q-values	[0.278]	[0.042]			[0.278]	[0.236]
Observations	9,612	9,602	9,611	9,600	9,614	9,604
R-squared	0.382	0.768	0.347	0.563	0.623	0.885
Mean of Outcome	2659.310	2659.310	0.188	0.188	-0.000	-0.000
Time F.E.	YES	YES	YES	YES	YES	YES
HH Controls	YES	YES	YES	YES	YES	YES
Parish F.E.	YES	YES	YES	YES	YES	YES
HH F.E.	NO	YES	NO	YES	NO	YES

(Notes for table: Outcome variable total household (HH) value of assets is derived by first aggregating the total value of all the assets owned by a household in a given period. Then they are deflated to 2017 prices and finally converted to US dollars using Official Exchange rates. In order to avoid extreme outliers I winsorize 5% of the distribution on either sides. For Amenities Index, I create a standard normal variable by combining factors like number of rooms per capita, access to electricity, access to own water source and own toilets and roof quality, floor quality and wall quality. Outcome variable ownership of land dummy is a dummy variable that takes the value of 1 if in a given time period, the household reports land as one of the assets owned by household and 0 otherwise. Column 1 represents regression with time and parish fixed effects, column 2 includes household fixed effects as well. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

2.5.2 Household Earnings and business

Next I regress household monthly earnings on the treatment variable. I do three separate set of regressions here - in the full sample I regress on the full sample of households; in the all earners only regression I only take households that report some positive monthly earnings. Finally, in the wage earners only regression, I further reduce my sample by taking only those households that report positive wage earnings. Exposure to savings groups does not seem to have any significant effect on monthly earnings of households in either case. It's important to note that almost half of the sample reports subsistence farming as their most important source of earning - this means we do not have a corresponding measure of HH earnings for most of these households. Thus it's hard to accurately track the impact of savings groups on household's income due to our data limitations. I also regress household business dummy on treatment and find negative significant results at 5%

level. This seems to be consistent with the idea that access to savings groups reduces households' needs to mitigate risk through engaging in activities outside of agriculture (like owning a business). The safety net that the group provides allows households to better manage risk.

Table 2.3: Fixed-Effects regression of total household monthly earnings and business dummy on treatment variable

	All HH		**All earners only**		**Wage earners only**		**HH with business**	
Variables	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
SavingsGroups	0.426	0.704	0.555	8.993	2.134	1.871	-0.045*	-0.057**
	(2.411)	(2.320)	(6.301)	(6.717)	(8.988)	(11.272)	(0.026)	(0.026)
q-values	[0.387]	[0.439]					[0.278]	[0.053]
Observations	9,628	9,618	2,815	2,121	1,390	1,033	9,628	9,618
R-squared	0.276	0.662	0.408	0.752	0.544	0.821	0.184	0.603
Mean of Outcome	27.128	27.128	91.946	91.946	108.243	108.243	0.378	0.378
Time F.E.	YES	YES	YES	YES	YES	YES	YES	YES
HH Controls	YES	YES	YES	YES	YES	YES	YES	YES
Parish F.E.	YES	YES	YES	YES	YES	YES	YES	YES
HH F.E.	NO	YES	NO	YES	NO	YES	NO	YES

(Notes for table: Outcome variable total household (HH) monthly earnings is derived by first aggregating the total monthly earnings of each household member across jobs, and then summing up the monthly earnings for all the members in the household that report to be working. Then they are deflated to 2017 prices and finally converted to US dollars using Official Exchange rates. In order to avoid extreme outliers I winsorize 5% of the distribution on either sides. I extract two sub-samples of households - one that reports at least some positive earnings and other reports at least some positive wage earnings only. Outcome variable household (HH) having business is derived by looking at if a household member ran a business in the last 12 months and then aggregating it for all household members; if at least one household member reports running a business in the last 12 months then the household business dummy will take the value of 1. Column 1 represents regression with time and parish fixed effects, column 2 includes household fixed effects as well. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

2.5.3 Food diversity and expenditures

I next delve into the food security results. It is important to note that households are prone to a variety of shocks in any given year. The shocks can take any form like irregular rains, theft, death of income earner or serious illness or even unusually high level of crop diseases and pests. On average, 40% of the households report facing at least one shock in the last year. Thus, it is important to look at how households' exposure to shock affects their food consumption and if being exposed to savings groups helps in their food security or not. Thus, I run 3 separate sets of regressions for food index. Specifically, I first regress food index on full set of households. Then, I also regress food index on only those sample of households that faced any shock in the previous year and on only those sample of households that did not face any shock in the previous year.

In terms of food index I observe that there is a significant increase in food index coefficient at 5% level for the full sample of households with a coefficient of 0.089 which implies a magnitude of treatment effect of 0.281. This indicates that households with exposure to savings groups consume more categories of food on average and/or spend more money on food consumption compared to households without exposure to savings groups. I observe an even stronger effect of exposure to savings groups on households that did not experience any shock with a coefficient of 0.193 significant at 1% level. This translates to a magnitude of treatment effect of 0.610. The findings above are in tandem with those from Kiiza and George, 2021 who find that access to SACCOs has a positive effect on the household dietary diversity score during the period of 2009-10 and 2010-11. However, for households that did experience a shock I observe negative but insignificant coefficient - this means that for households who experienced some shock there was a decline in variety of foods consumed and/or food expenditure but the drop was not significant for households that were exposed to savings groups. This implies that savings groups have an important contribution in enhancing food security by increasing the variety of goods consumed by households and enabling them to spend more money on food in absence of shocks and in minimizing the effect of shock on food security.

Table 2.4: Fixed-Effects regression of food index on treatment

	Food Index					
Variables	**(Full)**	**(Full)**	**(No shock)**	**(No shock)**	**(Shock)**	**(Shock)**
SavingsGroups	0.056	0.088**	0.135*	0.193***	-0.039	-0.014
	(0.046)	(0.044)	(0.069)	(0.074)	(0.073)	(0.071)
q-values	[0.278]	[0.067]				
Observations	9,445	9,404	4,804	4,060	3,283	2,451
R-squared	0.360	0.624	0.429	0.733	0.338	0.653
Mean of Outcome	0.000	0.000	0.056	0.056	-0.076	-0.076
Time F.E.	YES	YES	YES	YES	YES	YES
HH Controls	YES	YES	YES	YES	YES	YES
Parish F.E.	YES	YES	YES	YES	YES	YES
HH F.E.	NO	YES	NO	YES	NO	YES

(Notes for table: Outcome variable represents a standard normal variable which is a combination of food variety index and food expenditure by household. The food variety index measures the variety of food items consumed by households which I divide into 11 different categories - these categories include Beverages, Cereals, Fruits, Meat/Fish, Milk/Egg/Dairy products, Nuts and Seeds, Oil and Spices, Pulses, Starches, Sugar & sweets and Vegetables. The greater the score out of 11 the better the household ranks in terms of diverse foods consumed. Food expenditure is obtained by looking at the total value of food consumed by a household in a given time period - I then deflate the value to 2017 prices and convert it to USD using Official Exchange rate. I run 3 separate set of regressions - one for the full sample of households, one for those subset of households who did not experience any shocks in the last 12 months, and one for those subset of households which experienced at least 1 shock in the last 12 months. Column 1 represents regression with time and parish fixed effects and column 2 includes household fixed effects as well. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

2.6 Heterogeneity Analysis and Results

To better understand the mechanism through which the timing of entry of savings groups influences household's welfare outcomes, I next show how treatment effects vary by characteristics of household. Here, I only look at the differential treatment effects on household's total value of assets and monthly earnings. Savings groups helps in significantly boosting the total value of assets both in rural and urban areas, but the effect in urban areas is almost 4 times more compared to that in rural areas. Savings groups do not seem to be significantly improving assets in female headed households, but I do find a significant impact on households with male household heads. Similarly, savings groups have a significant impact on asset value of the richer households, but an insignificant impact on poor households. Finally, savings groups are seen to be much more significantly increasing household assets for households that report the primary occupation and main source of earnings to be "Farming" than for the non-farmer households. On the other hand, savings groups do not seem to be significantly improving monthly earnings for any of the heterogeneous type, however we do see a larger magnitude effect on urban households, households with female heads and non-poor, non-farmer households. One of the potential explanations for a significantly negative effect on poor households' earnings could be that savings groups help households to climb up the income ladder and migrate from poor to non-poor category. If the number of households reporting themselves to be poor declines, then average income level of poor households might decline as well.

Table 2.5: Heterogeneous Fixed-Effects regression of total value of assets on treatment variable

Variables	Rural	Urban	Female HH head	Male HH head	Poor	Non-poor	Farmer	Non-farmer
SavingsGroups	330.938**	1,907.872***	283.197	690.614***	-306.270	730.717***	489.336**	545.482*
	(164.112)	(696.833)	(279.057)	(250.773)	(213.887)	(237.768)	(226.273)	(328.297)
Observations	7,220	1,974	3,084	6,259	1,309	7,455	4,516	4,156
R-squared	0.705	0.801	0.761	0.786	0.707	0.773	0.697	0.814
Mean of Outcome	1854.169	5369.440	2276.686	2852.269	917.893	3082.054	1866.568	3488.837
Time F.E.	YES	YES	YES	YES	YES	YES	YES	YES
HH Controls	YES	YES	YES	YES	YES	YES	YES	YES
Parish F.E.	YES	YES	YES	YES	YES	YES	YES	YES
HH F.E.	YES	YES	YES	YES	YES	YES	YES	YES

(Notes for table: In this table I investigate the heterogeneous effects of treatment on household assets by breaking down the sample of households into 8 different categories and running the regression for each one of them. The 8 categories are Rural, Urban, Poor, Non-poor, Farmer, Non-Farmer, Female-headed households and Male-headed households. I run the regressions using time, parish and household fixed effects. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Table 2.6: Heterogeneous Fixed-Effects regression of total monthly earnings on treatment variable

Variables	Rural	Urban	Female HH head	Male HH head	Poor	Non-poor	Farmer	Non-farmer
SavingsGroups	-0.552	4.644	3.830	-2.040	-11.775***	2.892	-2.119	6.024
	(2.440)	(8.334)	(3.873)	(2.728)	(4.056)	(2.916)	(2.206)	(5.156)
Observations	7,232	1,978	3,093	6,266	1,313	7,469	4,519	4,171
R-squared	0.612	0.710	0.653	0.687	0.630	0.690	0.532	0.714
Mean of Outcome	19.292	53.498	22.671	29.379	15.621	30.013	10.805	44.176
Time F.E.	YES	YES	YES	YES	YES	YES	YES	YES
HH Controls	YES	YES	YES	YES	YES	YES	YES	YES
Parish F.E.	YES	YES	YES	YES	YES	YES	YES	YES
HH F.E.	YES	YES	YES	YES	YES	YES	YES	YES

(Notes for table: In this table I investigate the heterogeneous effects of treatment on household total monthly earnings by breaking down the sample of households into 8 different categories and running the regression for each one of them. The 8 categories are Rural, Urban, Poor, Non-poor, Farmer, Non-Farmer, Female-headed households and Male-headed households. I run the regressions using time, parish and household fixed effects. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

2.7 Robustness

It is important to note that we do not know why a community gets a savings group when it gets one which could make identification of a clean model trickier. However, I make the following arguments to support exogeneity of treatment. Firstly, savings groups are founded by NGOs and are supply driven- this means that presence of savings groups is not directly related to existing consumption and welfare outcomes of household nor is it related to characteristics of a community. I demonstrate this in table 2.7, where I observe that whether a savings group enters a parish or not is unrelated to how many services a parish has to offer. Similarly, it is also unrelated to whether the parish is rural or urban or small or large in terms of its population. I, however, do find that most savings groups will first enter parish in urban areas and then expand on to the rural areas. In Fig 4.3, I also test for parallel trends using event study graphs which show how outcome variables respond to the timing of the arrival of the savings groups. I do not see any pre-trends for either of the key variables which strengthens my identification argument. Additionally, as reported earlier, the size of estimated impacts of savings groups on different household outcome measures compare with the size of estimates from RCT literature.

One of the key issues with staggered adoption of treatment, is that the fixed effects estimates may be biased, since we do not know precisely how it compares mean outcomes across groups. When there are heterogeneous treatment effects in a staggered treatment design, the standard "two-way fixed effects" (TWFE) estimator can produce biased results, potentially misrepresenting the true average treatment effect by providing a weighted average that may not accurately reflect the diverse impacts across different treatment groups and time periods; this can lead to incorrect inferences about the direction and magnitude of the treatment effect, especially if the treatment effects vary significantly depending on when and how a unit was treated. In extreme cases, the estimated treatment effect could even have the opposite sign compared to the true average treatment effect due to the weighting of different treatment groups. As discussed in Goodman-Bacon, 2021 paper, two-way fixed effects estimator is a weighted average of all possible 2x2 DD estimators that compare timing groups to each other which means some of them may have negative weights. Negative weights occur when already-treated units act as controls and changes in their treatment effects over time get subtracted from the DD estimate. This negative weighting arises when treatment effects vary over time, in which case it typically biases regression DD estimates away from the sign of the true treatment effect. To ensure that my estimates are fully robust, I also perform Goodman Bacon Decomposition of weights. As can be seen from table 3 and figure 1 in Appendix Appendix,it seems that most of the weights come from treated vs already treated. This could be a

potential issue if we have heterogeneous effects across time.

To correct for this, I employ "Heterogeneous difference in differences" model using "hdidregress twfe" estimators in Stata as outlined in Wooldridge, 2021. It estimates average treatment effects on the treated (ATETs) that may vary over time and over treatment cohorts. Treatment cohorts are groups subject to treatment at different points in time. Because there are multiple DID parameters, Wooldridge, 2021 refers to them as heterogeneous treatment effects or as heterogeneous DID. The advantage of using this estimator is that instead of potentially having a bias coming from already treated units, the estimator limits the control group to never treated units. I then use the post-estimation aggregation command to obtain a single ATET for each cohort aggregated over all cohorts and time - I also tabulate the same for key outcome variables. As can be seen in Table 4 in Appendix Appendix, I observe that most of the significance in my results comes from early treatment effects before they fade out in the later years, making the overall treatment effect insignificant for assets and food index. Nonetheless, the magnitude and sign of coefficients are preserved which is desirable.

Event Study Plot

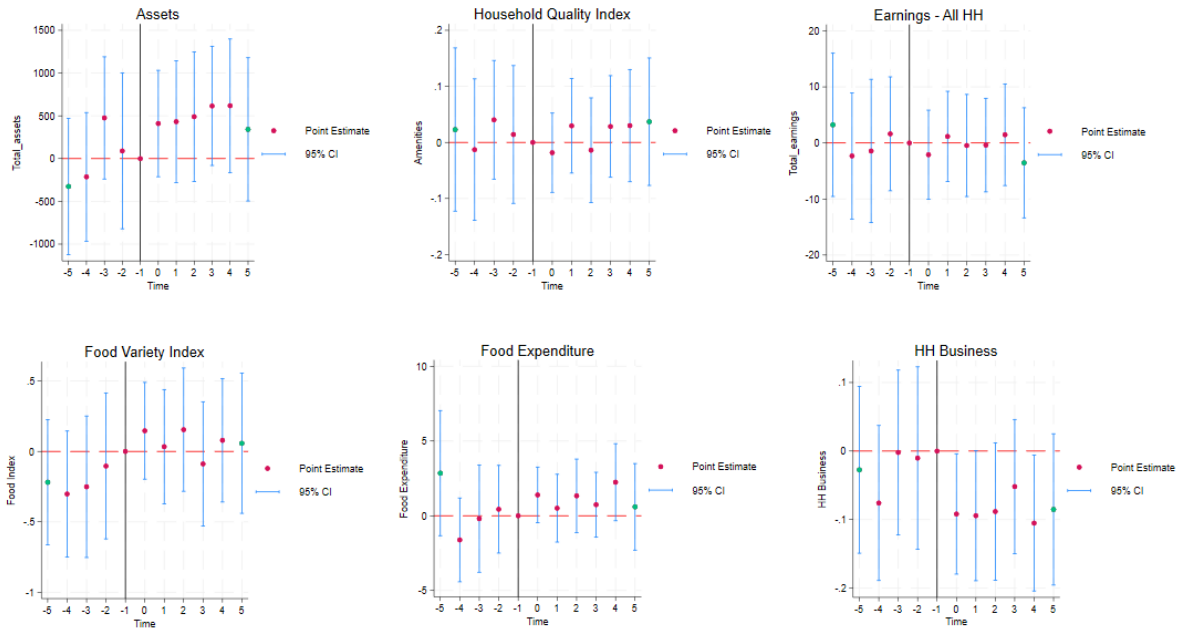

Figure 2.2: Event DD graphs

In this graph, I perform an event study analysis to assess how different outcome variables respond before and after treatment. The time variable stands for the difference between year of interview and year in which the parish was exposed to savings groups. All event study regressions are based on parish fixed effects with standard errors clustered at parish level except for regressions related to number of parish services, urban status of parish and number of households in parish all of which are based on district fixed effects with standard errors clustered at district level.

Table 2.7: Fixed-Effects regression of treatment and timing of treatment on parish characteristics, parish population and rural-urban status of parish

	Savings Group in Parish			**Late Treatment**		
Number of Services	0.011***	-0.000	0.008	0.014***	0.003	0.004
	(0.002)	(0.007)	(0.005)	(0.002)	(0.005)	(0.004)
Urban Parishes	-0.038	-0.034	0.010	-0.072*	-0.065*	-0.097***
	(0.056)	(0.057)	(0.040)	(0.041)	(0.039)	(0.028)
Number of HH	0.006	0.006	0.017***	-0.003	-0.002	-0.007**
in Parish (in 000's)	(0.011)	(0.01)	(0.004)	(0.003)	(0.004)	(0.003)
Observations	2,045	2,045	2,044	1,448	1,448	1,444
R-squared	0.027	0.041	0.342	0.076	0.112	0.433
Mean of Outcome	0.708	0.708	0.708	0.153	0.153	0.153
Time F.E.	NO	YES	YES	NO	YES	YES
District F.E.	NO	NO	YES	NO	NO	YES

(Notes for table: In this table I investigate if there is a selection of savings groups into parish depending on parish characteristics, urban status or parish population. The parish population is approximated by the number of households in a parish. I construct 18 dummy variables to look at 18 different services - namely - community road, government hospital, government primary school, government secondary school, private hospital, private primary school, private secondary school, private NGO clinic, pharmacy, police station, post office, veterinary service, Agricultural extension services, markets, trunk road, feeder road, technical school and government health center. I then add up these dummies to construct a measure for how many services are available to community members within the community. I run the regressions using no fixed effects, time fixed effects and district fixed effects. Standard errors are clustered at the district level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Bulk purchasing as a mechanism for food security

In this section, I study the bulk purchasing behavior of households as a motivation to test one of the channels through which households could achieve food security, even when there is no real improvement in income. Purchasing in bulk offers the buyer an income effect from the price discount, moreover bulk purchasing offers an insurance against price volatility. However bulk purchasing requires access to large lump-sum amount of money that can potentially come from savings. Households with savings can purchase in bulk when the prices are low and consume it later, so households can adjust the timing of their expenditure to the season when prices are lower, meaning they can even out the consumption timing from savings. Thus, households without savings would be more exposed to variation in prices whereas households with savings would be less sensitive to price volatility.

Bulk purchases directly offer a price discount to the buyers. This is because the bulkier the purchase the less price you have to pay per unit for the commodity. I demonstrate this using data by first normalizing the commodities on a common scale of kilograms/liter and then estimating price per kg/liter on units purchased using commodity, district and time fixed effects. The units purchased in kg/liter tells us how much bulky the purchase is; for instance a factor of 10 means that the commodity is being purchased in a unit of 10kg / 10litres. The expectation is to get an inverse relationship between bulkiness of commodity and cost of purchase – for instance, purchasing a 120 kg sack should cost less per kg than purchasing a 10 kg sack, everything else equal. As can be seen from Fig. 4.2, I indeed observe a negative relationship between price per kg/liter of commodity and actual unit of purchase of commodity (more the kilograms purchased, lesser the price per kg).

I also capture seasonal variation in the price of these food items in my data. This is important because households will benefit from purchasing in bulk only if they can arbitrage from monthly price variation. As seen in Fig. 2 using rice as an example, I do observe monthly fluctuations in the price of rice indicating that households can benefit by making bulk purchases of rice bags in the months of April, July and October when the prices are low and use it for consumption in the subsequent months when prices are higher.

While most of the theoretical models on commodity storage are developed in developing country's context, there are only a few empirical studies that investigate household grain storage decisions in Sub-Saharan Africa. My paper relates to the literature on strategic holding of bulk food items. Tesfaye and Gebremariam, 2020 shed light on maize storage behavior and investigate the effect of climatic factors, improved storage practices, and household characteristics on grain storage behavior. They attribute grain storage practice in Ethiopian households to factors like crop

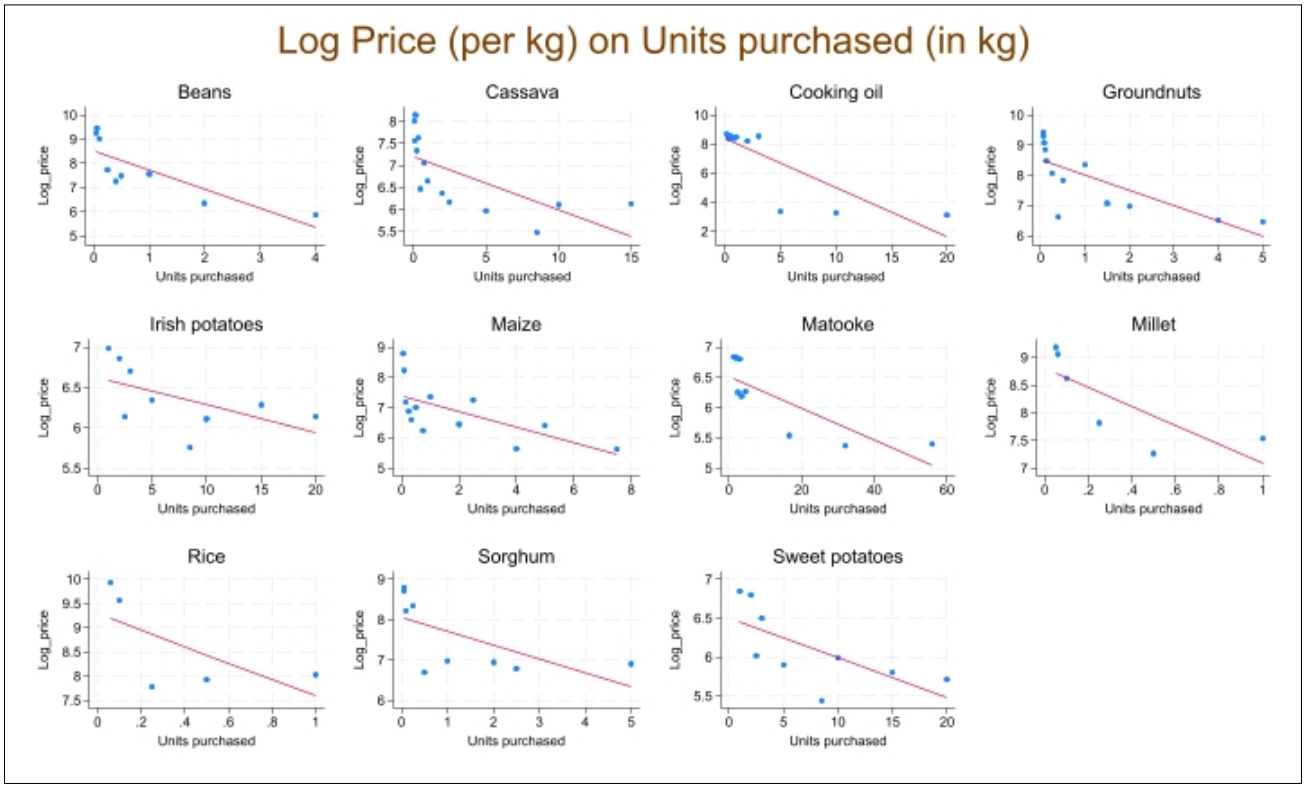

Figure 2.3: Price of commodity per kg/liter on bulkiness of purchase

protection technology, rainfall shocks, temperature and health shocks among other things. My paper on the other hand studies the channel of savings groups as a contributor to bulk purchasing behavior in Ugandan households. Aggarwal et al., 2018 in their paper used a novel approach where they encouraged randomly selected ROSCAs to set aside maize together in communal bags, stored at a single member's house (usually the ROSCA treasurer) in Kenya. They provided ROSCAs with storage supplies, namely triple-layered plastic bags capable of being hermetically sealed and designed specifically for the purpose of grain storage, and a heavily subsidized wooden stand to keep the maize elevated from the ground (and less susceptible to pests and water damage). They found that this technique increased storage among smallholder farmers providing savings clubs with a simple way to set aside maize for storage.

The consumption module provides detailed information on consumption behavior of households for 100+ commodity items that households can choose to consume from. I fetch information on all products consumed by the household in the last 7 days including their market prices and units of purchase from the consumption module. Firstly, I combine many of these commodities into larger commodity buckets, for instance maize flour, maize cobs and maize grains will all

Figure 2.4: Average Market price of rice across the months

be classified together under Maize. The dataset also has many non-standardized unit codes which I convert in terms of kilograms / liters as standards of unit. Since Matooke is one of the most important dietary items of consumption, I describe the methodology of converting Matooke units of consumption to standardized units in Appendix.

Since all households do not consume all the available products, I have an unbalanced panel when it comes to the products consumed. Hence, I balance out my panel by adding extra rows for each household - one for each of the product not consumed by the household in a given year. Then I populate all the remaining household and community particulars except for the consumption and pricing related information which are left missing for the products not consumed. After all of the data cleaning operations performed so far, I recheck if there are any households that appear just a single year data and therefore cannot be panelized and I drop all such households. This leaves me with 2756 unique households that repeat across years.

Then for all of these larger commodity buckets, I compute the share of expenditure on the particular commodity as a proportion of total value consumed of all commodities by a household in a given year. I drop all the commodities which have less than 1% share in the total value of food consumed by household. In the final step of data cleaning I create a new dummy variable which takes the value of 1 if the product is perishable and 0 otherwise. I further, eliminate all the commodities which are perishable and therefore cannot be purchased in bulk or stored for many days. This leaves me with 11 commodity items for my analysis purpose. I am left with household-demographic-product-community information for 2753 unique households, which I use for my regression analyses. Out of the total purchases made by households for consumption purposes, 15.5% account for purchases in bulk.

Next, the idea is that savings groups encourages bulk purchasing by making you less price sensitive. I test this using the following model equation:

$$\boldsymbol{bulkpurchases_{chpt} = \beta SavingsGroups_{pt} + f_c + f_h + f_p + f_t + \delta X_{hpt} + \epsilon_{chpt}} \quad (2.2)$$

where f_c is the fixed effect for commodity. This parameter identifies household-level preferences for good c. (for example, some household like to eat a certain crop, others don't).

Results from regression analysis

I regress the bulk purchase dummy on treatment. However, I do not observe any strong association between exposure to savings groups and bulk purchases. Note that here I am not looking at if the household consumes the entire product in bulk or not; but rather I am interested in observing whether the household reports a product that was purchased in bulk units. One of the reasons why savings groups still do not result in bulk purchases could be that goods kept in bulk at home could be subject to spoilage or even theft (see Aggarwal et al., 2018). A further analysis requires digging into the seasonality of bulk purchases based on the commodities. In Appendix Table 2, I also regress price of commodities (in standardized units of measurement) on the bulk purchase dummy. As expected, I find a significantly negative association between the two - the bulkier the purchase, the more price discount obtained by the consumer.

Table 2.8: Fixed-Effects regression of bulk purchases on treatment variable

	Bulk Purchase dummy		
Variables	(1)	(2)	(3)
SavingsGroups	0.001	0.003	0.003
	(0.006)	(0.006)	(0.006)
Observations	105,913	105,913	105,913
R-squared	0.024	0.042	0.298
Mean of Outcome	0.156	0.156	0.156
Time F.E.	YES	YES	YES
HH Controls	YES	YES	YES
Parish F.E.	YES	YES	YES
HH F.E.	NO	YES	YES
Product F.E.	NO	NO	YES

(Notes for table: Outcome variable represents the bulk purchase dummy which takes the value of 1 if the household reports purchasing the product in bulk units of 5 kgs and above. Note that if the household does not consume a given product in the given time period, I impute the bulk purchase dummy as 0. Column 1 represents regression with time and parish fixed effects, column 2 includes household fixed effects as well while column 3 also adds product fixed effects. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Conclusion

Access to financial services is crucial for development, as it enhances resource mobilization needed for productive investment and facilitates consumption smoothing. Reaching the rural poor with appropriate savings and credit services has been a persistent challenge for formal financial institutions globally. Access to savings and credit in low-income settings can improve food security, increase investment in productive assets and facilitate empowerment especially in the rural community. In this paper, I studied the channel of a growing financial institution in the form of Savings groups in the context of Uganda using National Panel Survey data from 2010-2020. The period corresponded to a rapid expansion of savings groups and I employed a TWFE model that exploited the variation in the timing of arrival of a savings groups in a community. I observed that savings groups significantly increase household's average value of assets, however they do not have a significant impact on monthly household earnings or on household quality. I also found a significant positive effect of savings groups on food security characterized by variety of food consumed and expenditure on food especially for households not facing any shocks during any given year. I also explored bulk purchasing as a potential mechanism to explain improvements in food security without any real increase in income. I observed that bulk quantities do trade at a discount to small quantities which means that households with savings would be less sensitive to price volatility. However, I did not find a significant evidence of increase in bulk purchases of storable foodstuffs due to exposure to savings groups. In conclusion, my study provides a comprehensive assessment of the impact of savings groups on households in Uganda, offering valuable insights into the potential of informal financial institutions to promote economic empowerment, asset accumulation and to enhance food security among the vulnerable households.

My results on bulk purchase behavior though insignificant is not fully without credit. It's reasonable to believe that households who save with formal or informal savings mechanisms are more likely to rely on savings in the event of an external shock, which means instead of purchasing durable foodstuffs in bulk today, households could very well be saving up money to spend on consumption and other expenses at the time of an external shock. I intend to study this aspect in my subsequent paper.

A number of impact evaluation studies found that the introduction of savings groups improves food security, overall consumption smoothing, livestock holding, household business outcomes and women's empowerment (Beaman et al., 2014, Gash and Odell, 2013, Karlan et al., 2017); however, these welfare impacts are quite muted, raising the question of why the increase in financial intermediation created by savings groups does not improve outcomes. Flynn

and Sumberg, 2018 in their paper observe that while savings groups can help to facilitate operational expenses and cash flow thereby supporting members' micro-enterprises – in opportunity starved contexts their transformational potential is at best limited, especially in the short run. Savings groups are known to have a profound impact on financial inclusion, economic empowerment, and social outcomes. They offer a viable alternative to traditional financial services, particularly in under-served areas. However, challenges related to management, sustainability, and scalability need to be addressed to maximize their impact. The success to sustain the savings groups may depend on several factors such as the cohesion between group members; the group's ability to deal with rotation rates and work with ultra poor members (Burlando and Canidio, 2016) and the continued support of the sponsoring institution that promotes the group and endows its members with basic skills to self-manage them (Greaney et al., 2016). For instance, Beaman et al., 2014 shows that the positive effects of a savings group program on savings, housing quality, and food security in Mali only materialized when recruitment activities were organized and directly structured by an NGO, while the implementation of the model with soft support from the NGO did not yield much benefit. Ensuring the long-term sustainability of savings groups can be challenging, considering that while many groups are successful in the short term, maintaining their operations over time requires ongoing support and capacity-building. Besides, poor management and lack of financial literacy could also result in mismanagement or fraud. Future research should focus on strategies for overcoming these challenges and enhancing the effectiveness of savings groups. Access to the VSLA methodology may also be an important stepping stone in terms of deeper financial inclusion levels. Self-management of the savings groups allows individuals to learn important financial concepts, which can help them overcome financial literacy and trust issues that restrict their demand for credit from external financial institutions. Savings groups may thus foster borrowers' progression up the lending ladder, promoting graduation into access to credit from formal financial institutions.

Chapter 3

Impact of Groundwater Conservation Laws on Indian Farmers' Profitability and Land Productivity

This chapter is based on a coauthored project with Pramod Dudhe, Assistant Professor at the Indian Institute of Management Udaipur. Pramod Dudhe and I jointly conceptualized the research design and conducted the empirical analysis. While Pramod did most of the data collection, I was primarily responsible for data analysis, coding the empirical specifications into software, writing parts of the manuscript, preparing the figures and tables, and refining the estimation strategy. Both authors contributed equally to interpreting the results and drawing policy implications.

3.1 Introduction

Punjab and Haryana, the northwestern states of India, are also known as the 'breadbasket' of India, producing two-thirds of food grains in the country with a significantly increased paddy (rice) area. As rice is a water-intensive crop, the rapid expansion of rice cultivation has meant considerable dependence on groundwater resources in addition to the monsoon (June-September) rains. The 2009 Preservation of Subsoil Water Act (SSWA) was introduced to conserve groundwater in the Indian states of Punjab and Haryana Government of Haryana, 2009; Government of Punjab, 2009 by shifting the paddy plantation closer to the arrival of monsoon rains. The law prohibited the sowing of paddy seeds in nurseries before a fixed date in May and the transplantation of the seedlings before a fixed date in June every year, pushing back its crop cycle. As a result, the paddy is sown mainly during the monsoon months of June and harvested in October-November. State governments of Punjab and Haryana announce these dates well in advance based on monsoon forecasts, and they may vary from year to year.

The act's main objective is to conserve groundwater by syncing the planting cycle of paddy with monsoon rains and dissuading farmers from using groundwater during the summer months. There is considerable evidence supporting this objective Singh, 2009; Tripathi et al., 2016. The act also resulted in unintended consequences, one being the reduced temporal window between harvesting monsoon crops and sowing winter crops. This led to farmers clearing the paddy residue by burning it to prepare the farms for winter crops. The burning has substantially contributed to increasing air pollution in the winter season in the northern states of India Agarwala et al., 2022; Kant et al., 2022; McDonald et al., 2019. In this paper, we are interested in using the exogenous policy variation of the SSWA to identify the effect of groundwater conservation laws on the usage of inputs, productivity, and profitability outcomes of farmers.

The paper uses difference-in-differences estimation to compare the policy-based outcome differences between the districts of treated states (Punjab and Haryana) and control states for a selected set of crops from 2008 to 2020, this being the first difference. The second difference comes from the variation in the June monsoon rains in these districts. The empirical investigations are divided by monsoon and winter season crops – this helps us isolate the main effects of the policy on the monsoon crops and the spillover effects, if any, on the winter crops. The primary data source is the "Cost of Cultivation Surveys" (CCS) Ministry of Agriculture & Farmers Welfare, 2020, which provides farm-level information on various inputs and outputs for various crops across districts in India.[1] We also perform separate analyses on paddy and all other crops to assess

[1]Since district identifiers in the data are available only since 2008, around the same time the SSWA came into

if there is any impact on crops other than paddy, i.e., spurious regressions do not drive our statistical results. Since state-level data are available in the CCS from 2000 to 2020, a triple-differences model is also estimated, with the third difference being the pre- and post-policy periods. Its results are presented in Appendix.

So far, we have estimated the results for a farm input directly affected by the SSWA and monsoon rains, irrigation, along with yield (quintals per hectare) and revenue (Indian rupees per hectare). A positive shock to monsoon rains seems to decrease the irrigation input (measured by irrigation hours per hectare) for paddy in the districts affected by the SSWA and does not affect other monsoon crops. These results are on an expected line, with the increased groundwater levels due to the SSWA and favorable monsoon rain shock leading to reduced irrigation hours in the treated districts. Moreover, there doesn't seem to be any spillover to winter crops for irrigation input. Next, we discuss the effect of the SSWA on yield and revenue.

Based on the extant literature, there are arguments on how the SSWA will affect the yield and, by extension, revenue. On the positive side, pushing the crop cycle of paddy away from the hot summer months and closer to the favorable monsoon climate should allow farmers to target inputs to growing conditions better and increase rice yield (Chahal et al., 2007). On the negative side, potential delaying of the paddy cultivation may result in yield penalties, as the literature argues (Shah et al., 2020). Although preliminary, the results suggest an adverse effect on yield and revenue. In the future, we plan to dig deeper into the impact of SSWA on yield and revenue by controlling for other potentially important factors, such as the variable agro-climatic conditions of districts. The future plan also includes estimating the effect of the policy on other inputs such as fertilizer and labor costs and farmers' profitability.

The paper plans to contribute to three streams of literature, the first of which is based on the effects of the SSWA – increase in the groundwater level and unintended increase in the paddy residue burning contributing to air pollution – already discussed in the previous paragraphs Agarwala et al., 2022; Chahal et al., 2007; Kant et al., 2022; McDonald et al., 2019; Shah et al., 2020; Singh, 2009; Tripathi et al., 2016. The second literature strand relates to using the CCS for economic analysis. It remains underutilized as a database despite providing farm-level data across various Indian states for principal crops Nawn, 2013; Srivastava et al., 2017. The third literature strand is linked to monsoon forecasts and how they can benefit farmers Burlig et al., 2024; Rosenzweig and Udry, 2019. Delaying paddy sowing due to the SSWA forces farmers to observe the start of the monsoon season, and it can potentially benefit them by tailoring the inputs to the monsoon forecasts.

effect, we cannot capture the pre- and post-policy differences for district-level specifications.

The rest of the paper is organized as follows. The next section provides a brief background of the SSWA. Section 3 then elaborates on the data sources, followed by the methodology for empirical estimation. The subsequent sections focus on results and spillover effects from the statistical analysis. The final section concludes.

3.2 Background of the SSWA act

Since the mid-70s, an assured income support from the government through 'Minimum Support Price' for rice and provision of subsidized power to access the irrigation system had driven the farmers to adopt the resource-intensive rice-wheat cropping system. The groundwater used by these irrigation systems enabled paddy rice to be sown several months before the onset of the Indian southwest monsoon, allowing the crop to mature earlier when pests are typically less prevalent. However, intensive rice cropping led to extensive groundwater extraction and depleted groundwater resources with evidence that the water table could be lowered in a way that could threaten the nation's future food requirements. In order to prevent a looming groundwater crisis and reduce the further use of groundwater for irrigation, the states of Punjab and Haryana implemented the 'Preservation of Subsoil Water Act' in 2009. As per the provisions of the Act, the sowing of the irrigated rice crop was delayed, syncing the rice crop growth period with monsoon season, so that the reliance on groundwater irrigation could be reduced.

3.3 Data

Our main data source is the "Cost of Cultivation Surveys" from the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Govt. of India. These surveys collect intensive information on the costs of various major crops across India, which are used in policy formulation, mainly to set the minimum support price for these crops. The surveys follow three-stage stratified random sampling, with tehsils as the first stage units, followed by villages and occupational holdings (farmlands). Each state is divided into homogenous agro-climatic zones, and the first stage sampling units are allocated to different zones in proportion to the crop area. Surveys began in 2000, but district identifiers were introduced in 2008. Moreover, cultivators included in the sample are surveyed over a period of three years. We therefore limit our study to the 2008-2020 period, and treat the data as cross-sectional. Choice of the crops is driven by major crops cultivated in treated states Punjab and Haryana. For monsoon season, we focus on bajra,

cotton, maize, potato, sunflower and paddy while for winter season we focus on bajra, gram, masur, lentil, mustard, wheat, cotton, maize, potato, sunflower and paddy. All these crops follow similar plantation cycles across the states for the respective seasons. We link the surveys to rainfall data from the "Indian Institute of Tropical Meteorology (IITM) Indian regional/subdivisional Monthly Rainfall data set". The monthly rainfall data for various districts of India are collated from daily rainfall data in Pai et al., 2014. We merge the rainfall data to our cost of cultivation data using district-time identifiers. Finally, we use use the consumer price index obtained from MOSPI to deflate all revenues and expenses based on the price index with 2015 as the base year.

3.3.1 Summary statistics

Table 3.1 reports the summary statistics of the key variables of interest. Over the period of 2008-2020, the sampled farmers irrigated for an average of 115.96 hours per hectare in Punjab and Haryana, compared to 74.31 hours per hectare in the rest of India. Farmers spent an average of Rs. 2149.16 per hectare in treated districts compared to Rs. 1813.20 per hectare elsewhere. This does imply that farmers in the state of Punjab and Haryana do rely on extensive irrigation for cultivation which potentially explains the need to implement SSWA act. Farmers in treated districts have a slightly higher yield of 73.02 quintals per hectare compared to farmers in control states with an average yield of 71.76 quintals per hectare. Farmers in treated districts also earn a higher revenue of Rs. 56000 on average compared to farmers in control districts who enjoy an average revenue of Rs. 42000. To ensure that we are accounting for real revenue, we deflate the actual revenue with CPI (price index with baseline year of 2015). We also winsorize 5% of the tail on each side to eliminate outlier values. This gives us deflated winsorized revenue of Rs. 54000 in control districts compared to Rs. 75000 in treated districts.

We observe that out of all the crops from our dataset cultivated by farmers, paddy accounts for 39% in control districts and 29% in treated districts. Wheat on the other hand accounts for 18% in the control districts and 39% in treated districts. Other crops account for 42% and 32% respectively in control and treated states. On average, the crop area under cultivation is 1.27 hectares in control districts and 2.40 hectares in treated districts. The average annual rainfall is 1191.05 mm in control districts and 1217.32 mm in treated districts. The average rainfall in June is 175.18 mm in control districts and 180.54 mm in treated districts.

Table 3.1: Summary Statistics

	Control			Groundwater Act		
Variable	**Mean**	**Std.Dev.**	**N**	**Mean**	**Std.Dev.**	**N**
Total Irrigation Hours (per ha)	74.31	161.68	260,197	115.96	129.10	28,775
Total Irrigation Rs (per ha)	1813.20	3395.54	260,197	2149.16	2971.11	28,775
Productivity (Qtls per ha)	71.76	184.69	260,197	47.23	73.02	28,775
Revenue (in 000s Rs)	42.64	41.88	260,197	56.12	35.99	28,775
Revenue (deflated in 000s Rs)	57.53	48.35	260,197	76.63	34.79	28,775
Revenue (deflated in 000s Rs - 5%)	54.07	33.99	260,197	75.64	30.13	28,775
Paddy	0.39	0.49	26,0197	0.29	0.45	28,775
Wheat	0.18	0.39	260,197	0.39	0.49	28,775
Other crops	0.42	0.49	260,197	0.32	0.46	28,775
Total Annual Rainfall (in mm)	1,191.0	617.1	260,197	1,217.3	641.6	28,775
June Rainfall (in mm)	175.18	168.72	260,197	180.54	180.75	28,775
Crop Area (in ha)	1.27	1.32	260,197	2.40	2.26	28,775

Summary statistics from 2008-2020 Cost of Cultivation Surveys and Pai et al., 2014 rainfall dataset. The unit of observation is the farmer. Groundwater Act states are Punjab and Haryana. Paddy, Wheat and Other crops are indicator variables that take the value of one if the farmer planted the specific crop.

3.4 Methodology

We are interested in identifying the effects of the groundwater act on farmers' irrigation choices in the presence of early Monsoon rainfall shocks. By shifting the earliest date on which farmers can plant paddy, it forces farmers to observe the start of the Monsoon season. We therefore expect the amount of irrigation to respond more closely to the June start of the season. We however, do not have a clear expectation about how the yields and revenue would respond to June rains. Consequently, we also expect that yields and revenue might be more responsive to the amount of June rains. More rainfall could be more beneficial for the crop in terms of yield and revenue but on the other hand there could be a negative penalty to the farmers by pushing them away from their optimal choice of planting the crops earlier.

For each farmer i in district d and cultivating crop c at time t, we therefore estimate

$$y_{icdt} = \alpha SSWA_d \times JuneRains_{dt} + X_{icdt}\beta + \delta_t + \delta_{cd} + \gamma y_{icd(t-1)} + \epsilon_{icdt}. \tag{3.1}$$

where y represents farm-crop-season level outcomes. The explanatory variable of interest is $SSWA \times JuneRains$, which is the interaction between the indicator for the district being located in one of the two SSWA states and a measure of the amount of rain the district received in June of calendar year t. We standardize the amount of rainfall in June by subtracting the average June rainfall in each district across years from actual June rainfall in the district in a given year and dividing by the standard deviation of the amount of June rainfall in each district across years. This measures 1 standard deviation shock in the amount of June rains in year t in district d.

Our regression includes also a number of additional explanatory variables. The control matrix X comprises of all control variables like amount of crop area under cultivation in hectares, July to September rainfall, Winter rainfall, Summer rainfall, as well as rainfall interacted with crop dummies. To control for variation in district-specific outcomes arising from differences in agro-ecology and state-level agricultural policies, we should include district fixed effects, δ_d. In addition, different crops require different input choices and generate different yields, requiring the inclusion of crop-specific fixed effect, δ_c[2]. Since crop effects may well interact with agro-ecological environment, we utilize district-specific effects δ_{cd}.

One limitation of our data is that we do not observe many farmer characteristics that may be relevant in the choice of inputs and determination of outcomes. For example, certain plots of land

[2]To be precise, the data specifies varieties within each crop, i.e., whether a paddy is from a local, improved or hybrid variety. We define the "crop" as the variety within the crop.

may be particularly suitable for irrigation, while others are not.. Fortunately, we are able to address these concerns by exploiting the panel aspect of the Cost of Cultivation Survey. Since each farmer is observed over a period of three years, we can use input choices and outcomes from the previous crop cycle to control for unobserved, time-varying characteristics of the farm. We thus include the one-year lagged outcome variable in the set of control. Our estimates are thus ANCOVA estimates.

Finally, we cluster standard errors at the district level to account for unobserved correlations in outcomes across units of observation, and estimate our regressions over the Monsoon and Winter seasons separately. For Monsoon instead of using total annual rainfall we take only June rainfall into our policy-district or policy-state-year-interaction. All the specifications would be repeated for winter season to assess the spillover effects of policy across seasons. We also perform regression analysis by splitting data into only paddy vs. all other crops and run two separate regression panels for the same. The specification remains the same, except that we no longer control for crop fixed effects when we are running regression on just Paddy.

3.5 Results and Analysis

In this section, we present the results for difference-in differences specification for irrigation inputs, revenue and productivity yield outcomes. Note that the act pushed the sowing and transplanting of crops to align with the monsoon season, hence we have run our regression only taking monsoon season. We also run regressions for winter season separately as part of spillover across seasons analysis.

The diff-in-diff specifications help us capture the treatment effect of being in the policy districts (districts in Punjab and Haryana) after the act was implemented relative to control. We indeed observe that the treatment effect was negative for total irrigation hours (per hectare) which confirms our hypothesis that the act reduced dependency on irrigation by aligning the cropping season with monsoon rains. We also find that the act significantly decreases crop yield as also decreases the aggregate revenue of farmers by a significant amount.

When we run our difference in differences specification for just Paddy, our results become stronger. We however do not observe any significant effects of the policy interacted with monsoon shock on irrigation inputs and revenue for other crops which aligns with our expectation since the policy was only meant to delay paddy sowing and transplantation. The negative effect on yields could potentially be explained by the shortened growing period for rice. Late transplantation meant rice had a shorter maturity window before winter crops like wheat needed to be sown. The reduced growing period could have led to lower yields, affecting farmers' total production and

revenue. In non-policy states, farmers could harvest rice earlier thereby capturing higher market prices. Farmers in Punjab and Haryana on the other hand would have to enter the market late due to late harvesting - if supply was already high, this would reduce selling prices. Thus, while the Act helped conserve groundwater, it cut farmers' earnings by lowering crop yields, delaying sales, and increasing costs—ultimately reducing revenue compared to farmers in states without the policy.

Table 3.2 examines the impact of the SSWA Act, as interacted with rainfall shocks, on irrigation inputs, agricultural yields, and farm revenues during the monsoon season. Using district-level rainfall data from 2008–2020 and a fixed-effects regression framework, we analyze how a one-standard-deviation increase in June rainfall influences irrigation hours, irrigation expenditure, yield (in quintals per hectare), and revenue (in rupees). Standard errors are clustered at the district level to account for within-district correlation. Additionally, we control for lagged dependent variables, monthly rainfall levels, their interactions with crop type, and total cultivated area to mitigate endogeneity concerns.

Our findings reveal that higher June rainfall significantly reduces irrigation hours, suggesting a substitution effect between rainfall and artificial irrigation. Specifically, irrigation hours decrease by approximately 8.76 hours without fixed effects and by 7.73 hours with crop, district, variety and year fixed effects included. While the reduction in irrigation expenditure is not statistically significant, yield declines significantly by 1.45 quintals per hectare, highlighting potential adverse effects of early monsoon rainfall variability. Correspondingly, farm revenue exhibits a significant decline of approximately Rs. 2,717 per hectare in the fully specified model.

Importantly, when we restrict our analysis to just Paddy, we find that the effects are even more pronounced. A one-standard-deviation increase in June rainfall results in a 13.09 hours reduction in irrigation hours and a statistically significant decline in yield (1.32 quintals per hectare). The decline in revenue is also substantial, with estimated losses around Rs. 1,793 per hectare. These results suggest that the Act made Paddy, a monsoon-dependent crop, particularly sensitive to rainfall variability at critical growth stages.

In contrast, when we examine all other monsoon-season crops (excluding Paddy), the effects differ markedly. A one-standard-deviation increase in June rainfall leads to a statistically insignificant change in irrigation hours and irrigation expenditure, indicating that non-Paddy farmers in the treated districts do not reduce artificial irrigation as much in response to increased rainfall. However, we observe a larger decline in yield (-1.38 quintals per hectare) and a more substantial drop in farm revenue (-3,320 per hectare), though the drop is not significant.

Overall, these results suggest that while increased early monsoon rainfall reduces irrigation needs, it also negatively impacts agricultural productivity and farm revenues, particularly

for Paddy farmers, possibly due to disruption of optimal growing conditions. These findings have important policy implications, particularly in designing climate adaptation strategies and irrigation infrastructure planning for monsoon-dependent agricultural economies.

Table 3.2: Impact of early rains in SSWA states on Irrigation and farm outcomes during the Monsoon season

Variables	**Irrigation Hrs**		**Irrigation Rs**		**Yield (Qtls perha)**		**Revenue (Rs)**	
SSWA x JuneRains	-8.762**	-7.729**	-105.259	-42.619	-1.573***	-1.454***	-2,135.733	-2,717.036**
	(3.482)	(3.855)	(69.681)	(78.791)	(0.464)	(0.447)	(1,358.239)	(1,380.227)
Mean of Dep Var	51.31	51.31	1,234.34	1,234.34	31.07	31.07	54,451.49	54,451.49
Observations	30,727	30,727	30,727	30,727	30,727	30,727	30,727	30,727
R-squared	0.656	0.762	0.539	0.648	0.756	0.837	0.606	0.767
Adjusted R-squared	0.656	0.749	0.539	0.629	0.756	0.829	0.606	0.755
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variety FE	No	Yes	No	Yes	No	Yes	No	Yes
Crop - District FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

The set of controls include rainfall levels in all other months, rain-crop interaction of total annual rainfall with paddy and wheat respectively, and crop area under cultivation. We restrict the sample to crops that are grown in monsoon season. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level.

Table 3.3: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Monsoon season using district level rainfall data for just Paddy

Variables	Irrigation Hrs		Irrigation Rs		Yield (Qtls perha)		Revenue (Rs)	
SSWA x JuneRains	-13.371***	-13.093**	-193.167**	-163.987*	-1.145**	-1.319***	-706.271	-1,792.797**
	(4.536)	(5.235)	(90.474)	(89.581)	(0.540)	(0.507)	(906.227)	(736.475)
Mean of Dep Var	60.25	60.25	1,281.91	1,281.91	37.62	37.62	55,262.45	55,262.45
Observations	20,456	20,456	20,456	20,456	20,456	20,456	20,456	20,456
R-squared	0.687	0.778	0.555	0.647	0.700	0.793	0.701	0.814
Adjusted R-squared	0.687	0.767	0.555	0.630	0.700	0.783	0.701	0.805
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variety FE	No	Yes	No	Yes	No	Yes	No	Yes
District FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

The set of controls include rainfall levels in all other months, and crop area under cultivation. Here, we filter the data to only include Paddy and drop all other crops. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level.

Table 3.4: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Monsoon season using district level rainfall data for all other crops

Variables	Irrigation Hrs		Irrigation Rs		Yield (Qtls perha)		Revenue (Rs)	
SSWA x JuneRains	2.765	2.233	90.349	184.521	-1.954**	-1.383**	-4,000.463	-3,320.109
	(2.308)	(2.513)	(116.538)	(136.526)	(0.762)	(0.668)	(3,759.703)	(3,164.932)
Mean of Dep Var	33.52	33.52	1139.6	1139.6	18.03	18.03	52836.36	52836.36
Observations	10,271	10,271	10,271	10,271	10,271	10,271	10,271	10,271
R-squared	0.490	0.667	0.515	0.660	0.468	0.683	0.537	0.746
Adjusted R-squared	0.490	0.638	0.514	0.629	0.467	0.655	0.536	0.724
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variety FE	No	Yes	No	Yes	No	Yes	No	Yes
Crop - District FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

The set of controls include rainfall levels in all other months, and crop area under cultivation. Here, we filter the data to exclude Paddy and include all other crops that are grown in monsoon season. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level.

3.6 Spillover across seasons Regression Analysis

In this section, we perform the same district level analysis but now for winter season alone. In contrast to the negative effects observed in the monsoon season, we find that increased June rainfall during the winter season is associated with significantly higher crop yields and higher farm revenues in the treated districts. Specifically, yield increases by 1.11 quintals per hectare, suggesting that additional early-season rainfall benefits winter crops rather than disrupting growth. Revenue increases by Rs. 450 per hectare in the model when crop, variety, time and district fixed effects are included. Unlike in the monsoon season, irrigation hours and irrigation expenditure do not exhibit significant reductions, indicating that winter farmers do not adjust irrigation intensity in response to increased rainfall.

One potential explanation could be that winter crops are mostly irrigated crops and additional rainfall in June may contribute to improved soil moisture retention, reducing water stress and supporting higher yields. The insignificant effect on irrigation suggests that irrigation practices in winter are more rigid or necessary due to the drier climate. These results suggest that rainfall shocks have seasonally distinct effects in treated districts, with adverse consequences on yield and revenue during the monsoon but marginally beneficial impacts in the winter.

Table 3.5: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Winter season using district level rainfall data

Variables	**Irrigation Hrs**		**Irrigation Rs**		**Yield (Qtls perha)**		**Revenue (Rs)**	
SSWA x JuneRains	0.264	-1.239	9.117	70.789	2.225***	1.110*	2,058.598***	449.796
	(1.576)	(1.702)	(68.155)	(77.513)	(0.666)	(0.581)	(627.846)	(531.368)
Mean of Dep Var	67.24	67.24	2359.54	2359.54	37.97	37.97	58532.96	58532.96
Observations	31,901	31,901	31,901	31,901	31,901	31,901	31,901	31,901
R-squared	0.491	0.610	0.610	0.690	0.788	0.914	0.594	0.792
Adjusted R-squared	0.491	0.586	0.610	0.671	0.788	0.908	0.593	0.779
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variety FE	No	Yes	No	Yes	No	Yes	No	Yes
Crop - District FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

The set of controls include rainfall levels in all other months, rain-crop interaction of total annual rainfall with paddy and wheat respectively, and crop area under cultivation. We restrict the sample to crops that are grown in winter season. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level.

3.7 Conclusion

In this paper, we study the exogenous policy variation impact of the SSWA act aimed at groundwater conservation on input, yield and profitability outcomes. Treatment effect was negative for total irrigation hours (per hectare) which confirms our hypothesis that the act reduced dependency on irrigation by aligning the cropping season with monsoon rains. We find a significantly negative effect on the crop yield and crop revenue of farmers. The effects are stronger for paddy compared to all other crops.

Chapter 4

The rise of singlehood and delayed marriages in modern India

4.1 Introduction

Late or delayed marriage refers to a situation where a person who has reached the state of maturity in every facet of life (mentally, spiritually, financially, and physically) is not married either probably as a result of unavailability of spouse (unmarried by circumstance) or some other factors which may possibly be difficult to identify (unmarried by choice). Late marriage and late parenting are two recent trends observed globally especially among employed women due to the significant political-social and economic shifts that have come up as the outcome of globalization and liberalization. Marriage is gradually transforming into an institution based more on individual free choice and romantic love rather than a means of survival, postponed to later in life as people adjust to changes in educational opportunities, increased non-agricultural employment, and the proliferation of an urban lifestyle (Guilmoto, 2012).

One of the most striking changes in Indian society over the last 20 years has been the gradually changing marriage markets characterized by a decline and delay in first marriage.
According to the Census of 2011, Indian women prefer to get married at an older age than the previous decade, making late marriage a reality for the women of today. Adults tend to postpone marriage in exchange for freedom, independence, more attention given to education and career-building and to fulfill their own needs as individuals before entering into marriage. Official reports from National Statistical Office show that the proportion of unmarried females increased from 13.5% in 2011 to 19.9% in 2019. There is a growing body of literature studying the phenomenon of single-hood and changing trajectories of marriage among women for developed countries. Berg-Cross et al., 2004 studied the phenomenon of globalization of elite single professional women in the US, Poland, and Germany and found that there is an increasing desire of young elite women to have developmental parity with men, including a long period in which to continue their education and focus on their own development before committing to marriage or a permanent relationship. Hendi, 2019 examined the role of increasing age at first marriage, declining mortality, narrowing sex differences in mortality, more intense educational assortative mating, and educational expansion on how first-marriage trajectories changed between 1960 and 2010 in the United States. This phenomenon of delayed marriage and single-hood is picking pace among more educated working women in the age group of 25+ years, however it hasn't been studied much in the context of developing countries like India.

It is interesting to note that delayed marriage is often stigmatized in many developing countries especially with rigid social norms surrounding family. For instance, unmarried women above 26 years of age are labeled as 'sheng-nu' or 'Leftover' women in China (To, 2013). In

traditional Japan, women who remained single and childless in the 30s were considered as useless and even called "loser dogs (makeinu)" as per Yamaguchi, 2006. Similarly, women in India too have to face stigma and societal pressure of getting married once they cross 25 years of age, though there is no explicit labeling of such unmarried females in India. Often single women are either perceived as unhappy, miserable, deficient, "desperate for men" or as unmarriageable, arrogant, and too independent. They have to juggle between their personal aspirations and traditional gender role expectations which emphasize on getting married and having children at an "appropriate age". The emergence of "leftover women" could be regarded as a manifestation of social progress, as more women pursue higher education, careers and "have more control over their own lives" instead of rushing into marriage (Gui, 2020). In recent decades, we can observe that women are fighting the stigma and postponing their marriage age because of various factors. Lamb, 2018 explored the stories of single women living in the urban metropolis of Kolkata, and in smaller towns and villages of West Bengal, as a means to examine the ways gendered identities intersect with the institution of heterosexual marriage, class mobilities, etc. and suggest that social recognition and belonging are even more important than independence and true single-hood in the lives of those who live outside marriage in India.

Women with greater economic and personal independence could use their increased incomes and modify their consumption patterns to fight the stigma associated with single- hood. This could plausibly open up avenues for new markets to emerge just to cater to single women (e.g., single women tourism). A woman marrying at an older, more mature age could potentially have higher bargaining power and autonomy in decision making which is significant not just for their own well-being but also for their children's well-being. An increasing age at marriage also hints towards fall in fertility rates (Kirdar et al., 2009) which is especially important for India considering it surpassed China in terms of population in 2023 to become the most populous country in the world. In India, premarital sex and birth outside marriage are frowned upon for cultural and religious reasons; hence marriage becomes the basis of family formation and marks the beginning of regular exposure to the risk of fertility. Thus, it serves to determine the future trend of demographic transition in India. Heath, 2014 have observed that women who work for pay may face greater rates of domestic violence only if they have less education or if they marry young, implying that domestic violence may be used instrumentally by husbands to counteract the increase in bargaining power that employment provides to women. However, women whose bargaining power is sufficiently high are more capable of breaking away from abusive marriages and thus do not face this increase in violence.

It is important to note that marital experiences of other women, especially instances of

conflictual and unhappy parents' marriage do have a profound influence on marital decisions of never married women, since they would rather wait longer to find a compatible partner or even not marry, than be stuck in a marriage that they cannot get out of (Ferguson, 2000). This is also owing to the fact that despite the recent increase in divorce rates in India, divorced women are highly stigmatized and discriminated against. While the traditional norms of hypergamy ("marrying-up") are still widely prevalent in India, increasingly females are considering "marrying-down" meaning marrying a man who is younger in age or has lesser educational attainment or even earns lesser than the female counterpart, as long as they could find a "soulmate" in them.

Because of the rigid gender division of responsibilities, often working women are expected to shoulder the double burden from both workplace and family. Thus, even if women decide to get married, they may want to enjoy their financial and personal independence before taking on the traditional roles in marriage by marrying at a later age. In the context where the gender norm has started to shift, the progressive achievement of women in education and economic participation are not followed by male's adaptation to domestic sphere. In emphasizing the gains to specialization derived from gender differences in comparative advantage in domestic and market labor, the Becker's (Becker, 1973) exchange model suggests that, by reducing women's comparative advantage in domestic work, women's increasing educational attainment and economic independence should reduce the gains to marriage. Highly educated women are likely to find marriage less attractive and shall continue to stay single longer as they expect greater incompatibility between work and familial responsibilities and it increases their opportunity cost of marriage.

It seems that policy programs aiming to increase educational and employment opportunities for women contribute to the marriage markets by influencing the marital decisions of women. Jensen, 2012 provided 3 years of recruiting services to randomly selected women from rural Indian villages to get jobs in the business process outsourcing Industry, thereby exogenously increasing women's labor force opportunities. He found that young women in treatment villages were significantly less likely to get married or have children during this period, choosing instead to enter the labor market or obtain more schooling or post-school training, consistent with increased aspirations for a career. Kritz and Gurak, 1989 argue that schooling exposes girls to new ideas, attitudes and aspirations that give them power to question traditional values. These ideational changes tend to delay marriage, give more autonomy to women in marriage decisions, encourage alternative institutions to marriage, such as cohabitation, and make it more acceptable to remain single. Loughran, Zissimopoulos, et al., 2004 estimate that delaying marriage increases hourly wages of women by nearly four percent for each year they delay resulting in substantial differences in hourly wages at later ages. The authors hypothesize that the effect of delaying marriage on wages is largely at-

tributable to the central role mobility plays in facilitating career development at younger ages, and that geographic mobility declines considerably for high ability individuals after marriage. Taken all together, if the careers of women have to take a backseat to the careers of their husbands, i.e., if marriage itself complicates career development we might expect individuals to take steps to minimize those negative effects.

India ranks among the top countries in the world having high proportions of child marriages and early marriages. While the literature abundantly studies causes and consequences of early marriage in India (ICRW Report, Jensen and Thornton, 2003, Corno et al., 2020), it is important to consider the rise of delayed marriages as it has crucial implications for the economy. My aim is to contribute to filling the gap in literature by analyzing the trend of delayed marriages and identifying the significant determinants of the same. The objective of my paper is threefold- 1) to demonstrate that there is a gradual but significant rise in proportion of never married women in India 2) to assess if there a significant delay in age at first marriage over the years 3) to determine plausible explanation for the delay and decline in marriage, more specifically, to attribute the changes to modernization characterized by rising female education, urbanization, participation in service sector jobs, etc.

For the purpose of this study, I use data from 3 rounds of DHS for India- this corresponds to survey data for 2005-06, 2015-16 and 2019-21. I specifically focus on the Women's module covering all eligible women in age group 15-49 years. This covers 124385 observations in 2005-06, 699686 observations in 2015-16 and 724115 observations in 2019-21 dataset. The outcome variables of interest are the marital status of women (to check the proportions of ever married women relative to single women) and age at first marriage. Time to age at first marriage of women is the duration of time until the age at which they started living with their first partner. The analysis is divided into 3 categories- descriptive analysis, survival analysis and decomposition analysis. The remainder of the paper will briefly describe the methodology utilised for each type of analysis as well as the primary set of results obtained from the same.

4.2 Descriptive and Regression Analysis and Results

In India, union formation is still a family-oriented matter with unions formed based on the similarities between partners regarding their social class, level of education, employment, religion, ethnic group, family background etc. These factors play an important role in collectively determining the timing of marriage. In this section I explore the trends in marital status and age at first marriage. The objective is to test if the proportion of never married women is increasing sig-

nificantly in India. To test this, I undertake a descriptive analysis whereby I assess the trends in marriage rate over the 3 survey rounds. For each age cohort, I calculate the proportion of women who are never married relative to all the women in that age cohort with different marital statuses and observe if these proportions change for each age cohort over time. I undertake similar descriptive approach to check trends in education levels across different age cohorts over time as also if the proportion never married changes with rural/urban status for all the age cohorts. I also examine the timing of marriage among the women across time and age cohorts to check if there is indeed a rising proportion of women who are choosing to delay their marriage.

In many countries the gender gap in education has reversed: women are getting more education than men. As parents increasingly invest more in their daughter's education and health, thereby resulting in women receiving more education and employment opportunities, it is important to see if this culminates into new generation of women developing different perceptions on gender, marriage, career, and family. In communities where some girls are in school or participate in non-family work, their marital behavior could bring about normative changes and encourage late marriage even among those not studying or working. Thus, the influence of schooling can go beyond the individual's own educational attainment and marriage age could be influenced by broader community factors.

Table 2 in Appendix looks at the proportion of women with different educational status. At aggregate, I observe rising education levels among women with declining proportions of women with no education between 2005-06 and 2019-21 and a corresponding increase in secondary and higher education levels among women over time. It is interesting to note that these proportions become sharper for never married women. It is of considerable interest to note that the proportion of women with higher education is almost twice as large among never married women compared to all women. This indicates that an important characteristic of never married women is that they are highly educated (and possibly professional working) women.

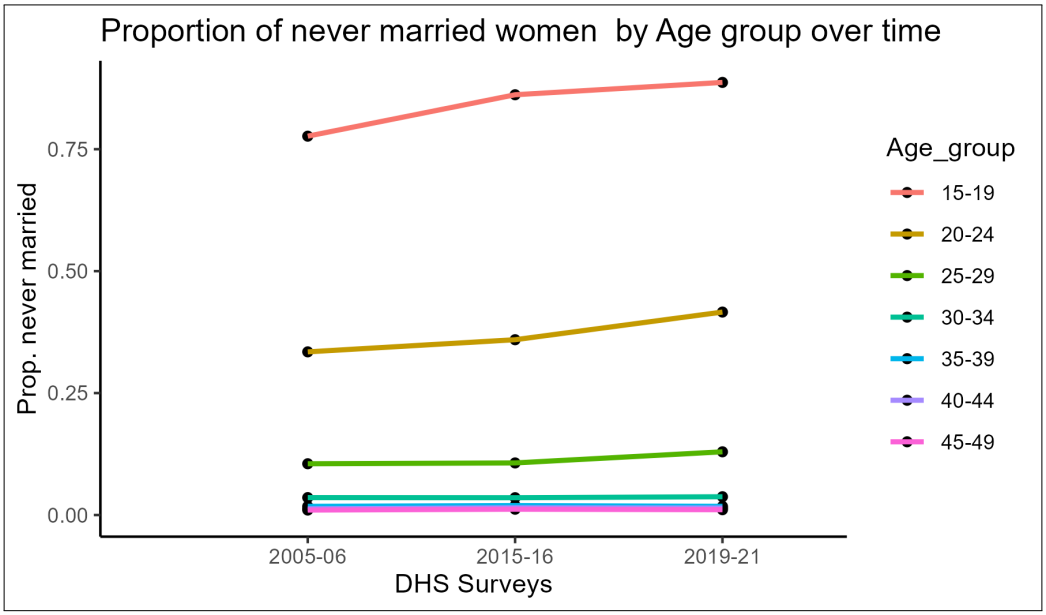

Figure 4.1: Proportion of never married women by age group over time

(Note: For each age group, I divide the number of women who report marital status as never married by the total number of women in each age group to get the proportion of never married women. I then chart the trend in proportion of never married women across the DHS surveys from 2005-06 to 2019-21.)

The above graph shows the proportion of never married women across age groups over time. For women in the age group of 15-19 years, we can see a substantial increase in never married females from 79% to almost 90%. I also see an increase in proportion of never married women for age group 20-24 from 35% to over 41% and for 25-29 from 11% to 13%. However, I see a constant proportion of never married women over time for women above 30 years of age. This indicates that even though single-hood may be part of women's reality today, most of the women still embrace marriage as a universal norm in India.

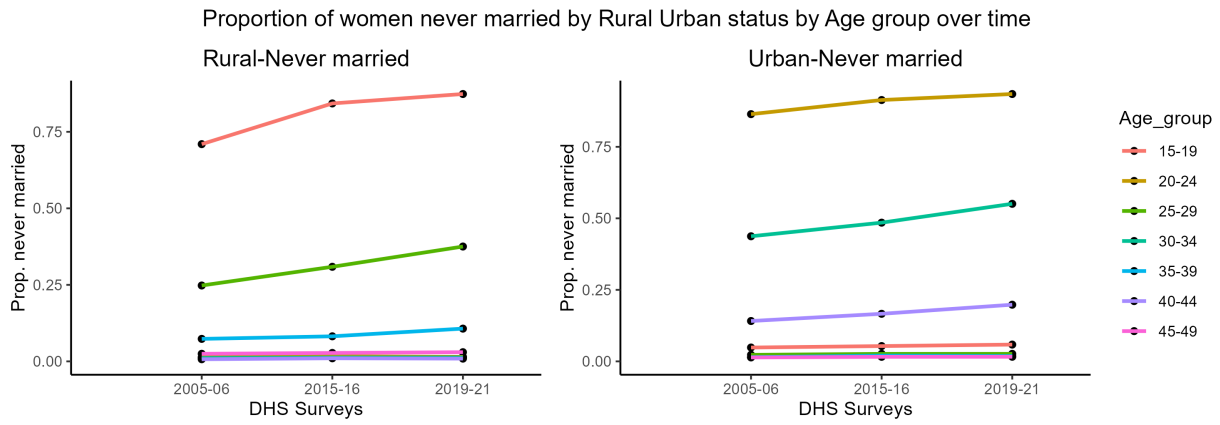

Figure 4.2: Proportion of never married women by rural urban status by age group over time
(Note: The data is first grouped by rural-urban status and age group. Then for each such group I divide the number of women who report marital status as never married by the total number of women in each age group to get the proportion of never married women. I then chart the trend in proportion of never married women across the DHS surveys from 2005-06 to 2019-21. An example to read the graph is out of all the women who report being in rural area and are in age group of 15-19, 71% report their marital status as never married in 2005-06 while 28% report other marital statuses in 2005-06.)

While there is an aggregate increase in proportion of never married women, it is important to note that this trend is not driven just by urban areas. I observe a rising proportion of never married women both in rural and urban areas. However, it is important to note that most of the increase for rural women is observed for females in age cohort 15-19 and 25-29, whereas for urban women most of the increase is observed for women in age cohort 20-24 and 30-34. This indicates that there may be different factors contributing to single-hood in rural and urban areas. Since early marriage is a widely prevalent phenomenon in rural areas particularly for women with no education, we may attribute the rising single-hood among the youngest cohort to policies encouraging education and preventing child marriage in rural areas. On the other hand, we may attribute rising single-hood among older cohorts in urban areas to employment in service sector jobs and dating/cohabitation culture which is increasingly becoming popular.

Finally, it is important to have a disaggregated view on proportion of never married women corresponding to different education levels. Here, I observe two interesting patterns. The proportion of never married women is low but increasing among females with no schooling or primary schooling. In contrast, the proportion of never married women is substantially high but constant for women with highest education level attended as secondary education or higher education. This indicates that even though women are increasingly pursuing higher education, the rigid social norms may inhibit them from shunning marriage altogether. Thus, persistent universal

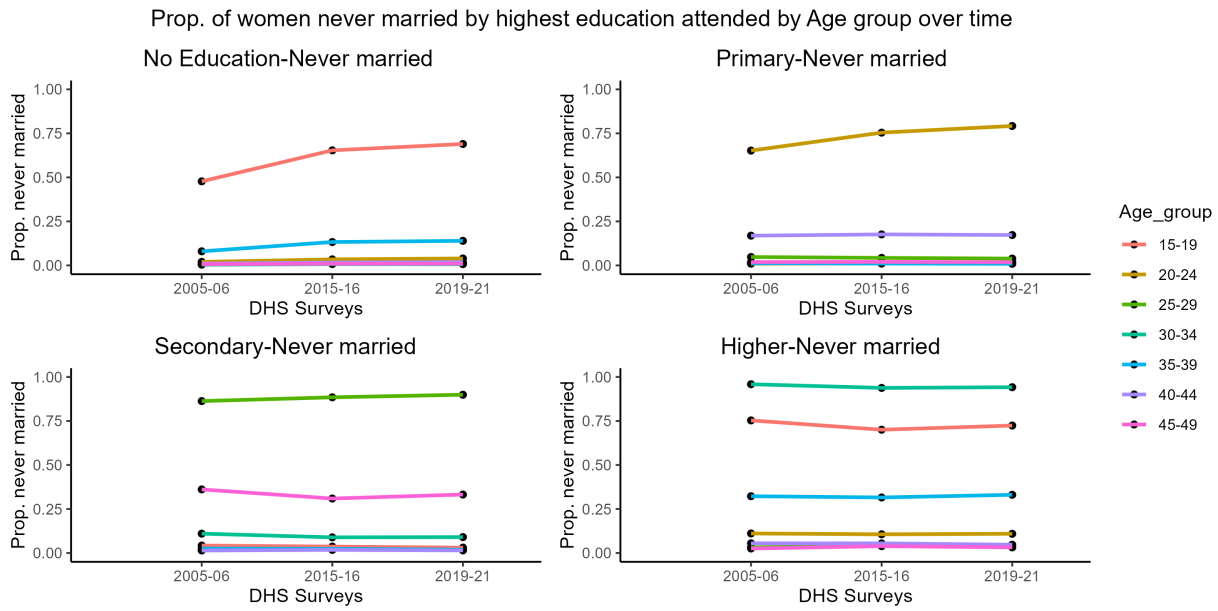

Figure 4.3: Proportion of never married women by highest education level attended by age group over time

(Note: The data is first grouped by educational status and age group. Then for each such group I divide the number of women who report marital status as never married by the total number of women in each age group to get the proportion of never married women. I then chart the trend in proportion of never married women across the DHS surveys from 2005-06 to 2019-21. An example to read the graph is out of all the women who report highest education level as secondary schooling and are in age group of 25-29, 86% report their marital status as never married in 2005-06 while 14% report other marital statuses in 2005-06.)

norms towards marriage and social stigma towards singleness continue to overshadow the effect of increase in educational achievement to single-hood in India.

Next, I look at the mean age at first marriage among women across all age cohorts and between 2005-06 and 2019-21 DHS surveys. Conditional on being married, I look at the changes in aggregate mean age at first marriage as well as break it down based on educational and rural urban status. Table 1 presents the summary of mean first age at marriage. From Table 1, we can observe that mean age at first marriage has consistently increased across all age cohorts between 2005-06 and 2019-21. However, the increase is small for women above 40 years of age, indicating that older cohort of women belonging to early generations do not experience a significant shift in age at marriage relative to younger cohorts. I also observe that the mean age at first marriage is lower for rural women compared to women living in urban areas across all age cohorts. Finally, I observe that mean age at first marriage increases with the highest level of schooling attended by women. Specifically, I observe a large gap of 6 years between women with no education and

Table 4.1: Mean age at first marriage

Age Group	Aggregate		Rural		Urban		No Education		Primary	
	2005-06	2019-21	2005-06	2019-21	2005-06	2019-21	2005-06	2019-21	2005-06	2019-21
15-19	15.73	16.65	15.57	16.62	16.18	16.85	15.16	16.24	15.56	16.29
20-24	17.34	18.30	16.96	18.21	17.93	18.71	16.05	17.35	16.70	17.46
25-29	18.20	19.20	17.53	18.94	19.02	20.05	16.30	17.49	17.06	17.71
30-34	18.38	19.26	17.57	18.87	19.33	20.40	16.23	17.39	17.14	17.79
35-39	18.21	19.03	17.54	18.67	18.98	20.07	16.36	17.45	17.43	17.98
40-44	18.11	18.73	17.42	18.36	18.87	19.74	16.48	17.44	17.39	17.92
45-49	18.22	18.91	17.64	18.66	18.86	19.63	16.92	18.00	17.60	18.39

Age Group	Secondary		Higher	
	2005-06	2019-21	2005-06	2019-21
15-19	16.30	16.72	17.51	17.77
20-24	18.05	18.27	20.44	20.03
25-29	19.00	19.19	22.46	22.08
30-34	19.50	19.57	23.49	23.24
35-39	19.63	19.74	23.49	23.47
40-44	19.51	19.69	23.28	23.33
45-49	19.58	20.00	22.95	23.48

(Note: Mean age at first marriage is calculated by restricting the sample of females who are ever married and calculating the mean age of marriage for the restricted sample for each age group)

women with higher education, indicating that delayed marriage is more prominent among women with higher education.

From the below regression summary (Hlavac, 2022) it is evident that mean age at marriage increased by a small yet significant amount across the DHS surveys. This seems to indicate that controlling for the observable characteristics like educational levels, rural/urban status, etc. there is also a gradual shift in societal norms and attitudes that might contribute to increasing the age at first marriage among women.

Model 1: Regression of Age at marriage on DHS survey dummy

Regression of Age at marriage on DHS survey dummy

	Dependent variable:			
	Age_at_marriage		c_Age_at_marriage_w	
	Unweighted-Full sample	Weighted-Full sample	Unweighted-Agregated sample	Weighted-Aggregated sample
	(1)	(2)	(3)	(4)
Survey 201516	0.654***	0.712***		
	(0.076)	(0.098)		
Survey 201921	0.844***	0.870***		
	(0.096)	(0.116)		
Survey 201516			0.617***	0.657***
			(0.061)	(0.078)
Survey 201921			0.828***	0.849***
			(0.075)	(0.099)
Observations	1,042,613	1,042,613	99,213	99,213
R^2	0.206	0.211	0.394	0.624
Adjusted R^2	0.206	0.211	0.393	0.624
Residual Std. Error	3.682 (df = 1042559)	3.586 (df = 1042559)	2.873 (df = 99159)	4.818 (df = 99159)
Note:				*p<0.1; **p<0.05; ***p<0.01

(Note: For the full sample I look at all married women and calculate the mean age at marriage with and without DHS survey weights. For the aggregated sample, I cluster the observations based on the state, rural/urban status, religion, caste, age, highest level of schooling attended and survey year participated in and then compute the weighted mean age at marriage using DHS weights. I further re-weight them based on observations corresponding to each cluster, the result of which is shown in column 4. Further, I apply state and age cohort fixed effects in all 4 regressions and cluster standard errors by State. The variable "`c_Age_at_marriage_w`" stands for cluster-weighted age at marriage.)

My next objective is to assess if there is a rise in delayed marriage among women in India across age groups and over time. For this purpose, I break the age at first marriage into 3 categories- a) women married before the legal age of 18 years b) women married between 18-25 years c) women with age at first marriage as above 25 years (late marriage). The three categories are mutually exclusive and a woman can fall into either of the three categories. Following graph shows the patterns in timing of marriage among women.

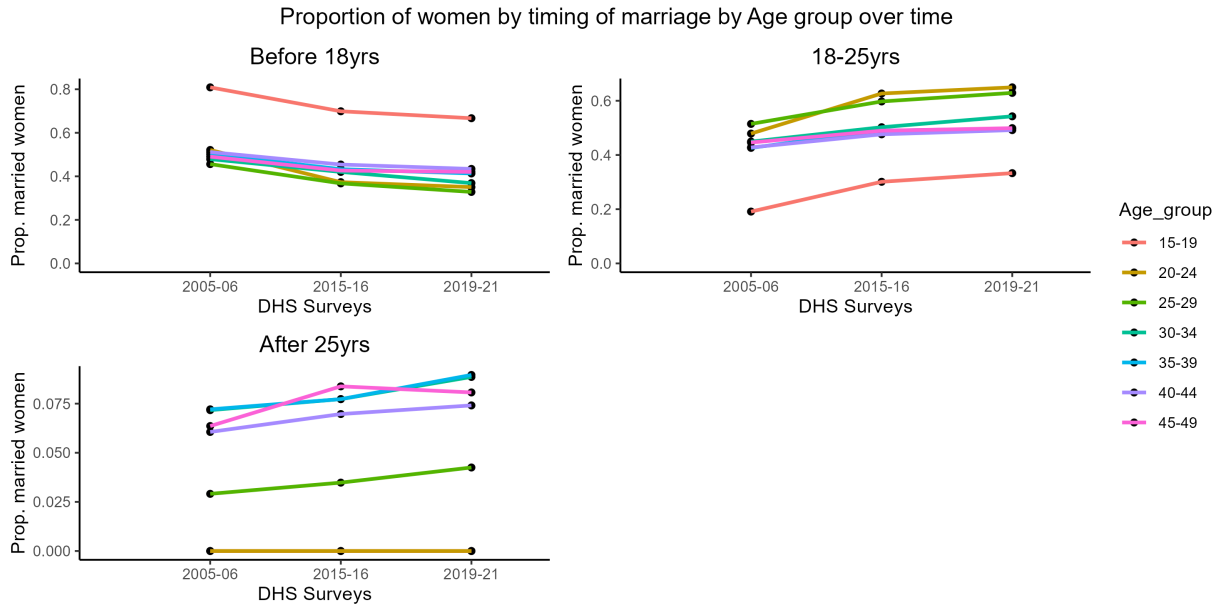

Figure 4.4: Proportion of women by timing of marriage by age group over time
(Note: The data is first grouped into the three categories of timing of marriage and the proportion of women belonging to each group is calculated relative to all married women)

As we can see from the above graph, across all age cohorts we see a declining proportion of women married before the legal age of 18 years. This is contrasted by a rising proportion of women married between 18-25 years of age. While the proportion of women married post 25 is still less than 10%, it is important to note that I do see an increasing trend for almost all age cohorts indicating that the trend is gradually getting accepted in the Indian society. Once again it may be of interest to look at patterns disaggregated by residential status and educational status (Table 3, Table 4, and Table 5 in appendix). I observe a steady increase in proportion of late marriage among women in rural areas, as well as a steady yet higher proportion of late marriage among urban women. While the proportion of women with delayed marriage is substantially low for females with highest level of education up to primary schooling, I do notice an increasing proportion of late marriage among more educated women. Particularly, I observe the phenomenon of late

marriage to be dominant among women with higher education. Once again this is consistent with my hypothesis that education is an important factor contributing to decline and delay in marriage among women in India. Since marital and student roles are perceived to be incompatible, with increasing levels of school enrolment and higher education take-up, we observe a corresponding delay in marriage age for females.

Finally, I assess how the sample of never married women has changed over time. In particular, I collect a bunch of covariates and measure the changes in proportions of never married women between 2005-06 and 2015-16 surveys. From Appendix Table 6 and Table 7, we can see that there was almost a 8% increase in proportion of never married women with higher education and a corresponding decline in proportion of women with no or primary education. When I adjust for state and age group fixed effects the proportion of never married women with higher education falls slightly, yet it is a significant 6% rise from 2005-06 proportion. The proportion of never married women living in urban areas falls by 4% whereas the proportion of women living in female headed households increases by less that 1% and is not significant when I adjust for state and age group fixed effects. There seems to be a significant drop in wealthier never married women and a corresponding increase in poorer never married women. This might indicate that richer women get married earlier on average than poorer women, however an in-depth analysis of the mechanism is beyond the scope of this paper. There doesn't seem to be much change in the contraceptive usage pattern among never married women, which may be consistent with the fact that pre-marital sex is still regarded as a taboo in India and hence majority of the never married females do not use contraceptives.The exposure to media largely remains unchanged among never married women. This might be because the drop in proportion of women with exposure to newspaper, magazines or radio is compensated by the rise in proportion of women with exposure to TV. On the whole, the employment and occupational variables show a significant decline in later survey consistent with the fact that labor force participation rates are declining among Indian women despite the rise in education levels. However, the silver lining is that we see an increase in proportion of never married women working in sales, professional, technical and managerial jobs. This indicates that the trend of decline in marriage is predominantly observed among professional working women. Finally there is less than 1% increase in proportion of never married women who think that domestic violence of any kind is unjustified, but the absolute proportions of never married women is high enough to begin with. Thus on the whole we can observe that never married women are more highly educated, poorer, professional working women and this trend is increasing over time.

4.3 Survival Analysis and Results

In this section, my primary goal is to identify the factors that aid women in delaying forward her marriage. To test this, I employ survival analysis approach to calculate the cumulative probability of age at first marriage for married women aged 15 to 49 years old according to characteristics like highest education level and their place of residence. Survival analysis consists of studies of the survival time of a subject which is the time that elapses between the baseline and the moment an event of interest occurs, or the subject drops out of the trial. Here, I plan to assess the probability of an individual woman delaying marriage by an additional year accounting for variation in socio-economic characteristics. My contribution therefore would be to analyze the trends in age at marriage and isolate the factors contributing to increased probability of delayed marriage. I employ the method of Cox Proportional Hazard model in this paper for the same.

4.3.1 Cox Proportional Hazard Model

This is a popular model in survival analysis used to assess the importance of various covariates in the survival times of individuals or objects through the hazard function. The model is a semi-parametric model because it makes no assumption about the shape of the baseline hazard function. The Cox Proportional Hazards model (hereon abbreviated as CPH) gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables (Region, Religion, Education, Caste, and Residence) denoted by X which are predictor variables to predict an individual's hazard (Age at Marriage). The survival time is assumed to begin at birth and ends when the individual gets married. The duration of life as single is computed as equivalent to "age at first marriage" for married women while that duration is equal to "current age" for women never married. It is censored for an individual still unmarried as at the time of the survey. The hazard function gives the instantaneous potential per unit time for marriage to occur, given that the individual has survived up to time t.

For an individual i with a set of attributes represented by a vector of independent variables, X, the Cox Proportional Hazard Model is as follows:

$$\boldsymbol{h_i(t, x_i) = h_0(t) Exp[\sum_{i=1}^{p} \beta_i x_i]} \tag{4.1}$$

h_i(t,x_i) is the resultant hazard given the values of the p covariates for X = (x_1,x_2,. . . .,x_p), B=(β_1,β_2,. . . .,β_p) and the respective survival time t = T. It is the hazard rate at the survival time t

for the individual i with fixed covariates.

h_0(t) is the duration-dependent risk component of the hazard function otherwise known as the baseline hazard. It is the hazard for the respective individual when the values of all the covariates are equal to zero.

The exponential term is relative risk component of the hazard function associated with having attribute X.

X is a vector of covariates- this includes highest education level of individual, rural/urban status, state, religion, caste, age group, etc.

B is a 1 x p vector of regression parameters to be estimated.

T denotes duration to first marriage measured in years (age)

The interpretation of the Cox model is done using hazard ratios, defined as the ratio of the predicted hazard function under two different values of a predictor variable. The hazard rate is measured by the ratio of number of cases experiencing the event at the end of a time interval to the total number of cases exposed to the risk of experience the event at the beginning of the time interval. A hazard ratio greater than 1 means the event is more likely to occur, and a ratio less than one means an event is less likely to occur. In this study, a lower hazard rate implies a longer duration of waiting time for the event to occur, i.e., older ages at first marriage. The coefficients of the covariates will be transformed by exponentiation and interpreted as risk ratios.

During the DHS survey all women were asked a series of questions regarding their marital status and whether they had ever lived with a man. All those who reported that they were ever married or ever lived with a man, were asked to indicate how old they were at the time when they started, for the first time ever, living with a man as a wife, irrespective of the legality or otherwise of their union. The response to this question constitutes the woman's age at first marriage. All the women who indicated that they had never been in a union or lived with a man were considered single and as a result they were not asked the question about the age at first marriage. Since marriage and schooling are generally considered to be incompatible roles, in order to avoid confounding effect of school enrolment on age at marriage, I exclude women below 22 years of age since that is the time by which most of the individuals complete their education.

The below table summarizes the result from Cox-proportional hazard model estimating hazard risk of marriage relative to different covariates. I run the model using State fixed effects and age cohort fixed effects. In the context of India which still has highly segregated gender norms, and the decision to participate in the labor market may be tied to marital outcomes thereby creating endogeneity bias, it is of interest to test if there exists a negative relationship between women's educational attainment and the risk of marriage. It can be argued that economic opportunities

proxied by higher educational attainment are associated with a lower risk of marriage only in societies in which sharply differentiated gender roles make it difficult for women to combine work and family.

A positive association is expected between educational attainment of females and their age at marriage. A woman with higher education is more likely to remain unmarried or to marry at a later age at a given point of time compared to a woman with no education at the same given point of time. The explanation can be twofold. First, the continuation of education delays the entry of a woman into the marriage market. Second, education is often related to greater autonomy and opening up of new avenues for women besides their familial and reproductive roles. In developing countries, as formal education becomes more available, “no education” goes from being a catchall category to a stronger indicator of relative disadvantage and lower social status.

Cox Mixed-Effects Model Results

Model 2: Cox Proportional Hazard model with Mixed effects for risk of marriage

Variable	Coef	exp(Coef)	se(Coef)	z
Highest_Education1	-0.0624	0.9395	0.0032	-19.52
Highest_Education2	-0.4152	0.6602	0.0026	-157.16
Highest_Education3	-1.0517	0.3493	0.0041	-259.55
Rural_Urban1	-0.0911	0.9123	0.0025	-36.71
Religion1	-0.0337	0.9669	0.0035	-9.75
Religion2	-0.1371	0.8719	0.0059	-22.98
Religion3	-0.1338	0.8747	0.0056	-23.86
CasteOBC	0.0304	1.0309	0.0027	10.90
CasteSC	0.0567	1.0584	0.0033	17.36
CasteST	-0.0647	0.9373	0.0038	-16.75
surveyname201516	-0.2370	0.7889	0.0039	-59.91
surveyname201921	-0.2611	0.7701	0.0039	-65.79

(Notes: The analysis uses 115,331 observations. The reference category for highest education is women with no education, the baseline for rural-urban is women belonging in rural areas, the baseline for religion is Hindus and the baseline for caste is women belonging to general/open category. The baseline for survey name is DHS Survey of 2005-06. Religion1 stands for Muslims, Religion2 stands for Christians and Religion3 stands for all other religions. Highest Education1 is for primary schooling, 2 is for secondary schooling and 3 is for tertiary education.)

As we can see from the model results, compared to women with no education, women with highest education level attended up to primary schooling have 6.1% lesser likelihood of getting married. Women with secondary schooling have almost 34% lesser hazard risk of getting married whereas women with higher education have nearly 65.1% lesser likelihood of getting married compared to women with no education. This implies that given that a female was single until time t, the likelihood of getting married in next period is almost 66% lesser for a highly educated woman compared to a woman with no education. Thus, higher education seems to be a significant and substantial factor in delaying marriage among females.

While it is difficult to capture gender norms in a variable, religion and caste often serve as good proxies for culture and thereby the underlying cultural norms. Religious norms and beliefs

affect one's orientation towards marriage and childbearing, among other things; thus, religion is bound to affect a woman's age at first marriage. Besides, in India, marriages are often caste endogamous, hence caste may also play an important role in determining the timing of marriages. Hence, I include both religion and caste as explanatory variables in my model. Historically, it has been found that Hindus and Muslims have had lower ages at marriage as compared to Christians and other religions. Also, different castes do not have the same economic status and therefore they do not have the same ability to invest in the human capital formation of their daughters. The higher age at marriage for general/forward castes may be associated with their higher economic power and greater exposure to the outside world. On the other hand, women from forward castes may experience early marriage due to the fact that marriage customs are more rigid in the high caste community. High socioeconomic status families may be motivated, for religious and prestige reasons, to get their daughters married at an early age, preferably before menarche. Women from backward castes may have lower education levels and may experience earlier marriages if marriage is viewed as means for survival.

As expected, the risk of getting married early was found to be higher for high socioeconomic status females compared with low socioeconomic status females. Muslim women seem to have 3.4% lesser likelihood of getting married, Christians have 12.9% lesser hazards and other religion females have almost 12.6% less chances of getting married compared to a Hindu female. While women belong to Other backward class and Scheduled castes have higher likelihood of getting married compared to general caste, women belonging to Scheduled tribes have 6.3% less hazards of getting married. This is consistent with my hypothesis that the risk of early marriage is higher for women belonging to high socio-economic status. Additionally, scheduled tribes have been known to have matriarchal family structure, as also they are often economically the most disadvantaged groups. Both the factors could mean that females get married later compared to women belonging to other castes.

Rural areas tend to have institutional and normative structures such as the kinship and extended family that promote early marriage and childbearing (Dixon, 1971). On the other hand, people in urban areas need to develop skills, gain resources, and achieve maturity to manage an independent household and thus they have to delay marriage. Furthermore, urban women tend to be more educated and engaged in salaried employed than their rural counterparts. Compared to a rural woman, urban females have 8.8% lesser hazards of getting married. This implies that living in urban areas does help a woman delay marriage, albeit not drastically.

It is also important to note that compared to 2005-06 survey, women surveyed in 2015-16 have 21.2% lesser likelihood of getting married whereas women belonging to 2019-21 survey

have almost 23% lower hazards of marriage. This highlights the increasing trend of delayed and declined marriage in India as structural transformation is accompanied with changing social norms and growing acceptance towards delayed marriage or even single-hood.

One important consideration in employing the proportional hazards model is that it assumes that the hazard function for an individual depends on the values of the covariates and the value of the baseline hazard. As a result, given two individuals, say two women, with particular values for the covariates, the ratio of the estimated hazards over time will be constant. However, the assumption that the effects of all covariates do not change over time, is often violated in practice. Consequently, the relative risk for a covariate exhibiting non-proportional hazards obtained by Cox regression may be under or overestimated. For future study, it may be more desirable to incorporate time dependent covariates in analysis to deal with this issue.

4.4 Decomposition Analysis and Results

Blinder-Oaxaca decomposition is a commonly used statistical method that decomposes the gap in mean outcomes across two groups into a portion that is due to differences in group characteristics and a portion that cannot be explained by such differences (Oaxaca, 1973, Blinder, 1973). Although this method has been most widely used to study gender- and race-based discrimination in the labor market, Blinder-Oaxaca decomposition can be applied to explain differences in any continuous outcome across any two groups. Let's say we have two groups- Group A and Group B. The mean outcome difference to be explained is simply the difference of the mean outcomes for observations in Group A and Group B, denoted as $\Delta\overline{Y} = \overline{Y}_A - \overline{Y}_B$. We can perform a threefold Blinder-Oaxaca decomposition of the mean outcome difference. The endowments term represents the contribution of differences in explanatory variables across groups, and the coefficients term is the part that is due to group differences in the coefficients. Finally, the interaction term accounts for the fact that cross- group differences in explanatory variables and coefficients can occur at the same time.

$$\Delta\overline{Y} = \underbrace{(\overline{X}_A - \overline{X}_B)'\hat{\beta}_B}_{\text{endowments}} + \underbrace{\overline{X}'_B(\hat{\beta}_A - \hat{\beta}_B)}_{\text{coefficients}} + \underbrace{(\overline{X}_A - \overline{X}_B)'(\overline{X}'_B(\hat{\beta}_A - \hat{\beta}_B)}_{\text{interaction}}$$

In this paper, I adopt this approach to study the differences in mean age at first marriage between women part of 2005-06 surveys and 2019-21 surveys. The objective is to understand how much of the shift in age at first marriage can be attributed to characteristics like education, urban-

ization, etc. and how much of it comes from changing social norms over more than a decade. I also compare the differences in age at first marriage between women marrying before 25 years of age and women marrying after 25 years of age (late marriage) to explore what factors contribute in the decision to marry late for women.

Oaxaca-Blinder Decomposition Results

Mean age at marriage in 2005-06 vs 2019-21

```
$y.A$
18.15996

$y.B$
19.04258

$y.diff$
-0.8826198
```

Comparison of mean age at marriage for women marrying before 25 years of age vs after 25 years of age

```
$y.A$
18.2313

$y.B$
28.3925

$y.diff$
-10.1612
```

Model 3: Threefold Oaxaca Blinder Decomposition

coef(endowments)	se(endowments)	coef(coefficients)	se(coefficients)	coef(interaction)	se(interaction)
-0.040742931	0.009060237	-0.883445596	0.016599997	0.041568748	0.012983528
-0.36084476	0.0169842	-8.69832188	0.01819085	-1.10203657	0.01994305

(The first line of decomposition is for how much is the difference in the mean age at marriage between 2005-06 and 2019-21 is attributable to endowments, coefficients and interaction respectively. The second line of decomposition is for how much is the difference in the mean age at marriage between women married before 25 years of age and women married after 25 years of age is attributable to endowments, coefficients and interaction respectively.)

From the model results, I observe that there is a mean difference of -0.88 years between the two survey periods indicating that on average women are marrying one year later in 2019-21 compared to 2005-06. On the other hand, women marrying before 25 years of age have mean age at first marriage as 18 years whereas women marrying post 25 have mean age at first marriage as 28 indicating a 10 year difference between the two groups. Out of the 0.88 mean difference in age at first marriage between the 2 survey groups, -0.04 is attributed to group differences in endowments (education, rural-urban status, religion, etc.) whereas -0.88 is attributed to group differences in coefficients. The remaining 0.04 is attributed to the interaction between endowments and coefficients. Similarly, out of the 10 year mean difference between early and late married women, -0.36 is accounted for by group differences in endowments and -8.7 is accounted for by group differences in coefficients while the remaining -1.1 is attributed to their interaction. Doing a twofold decomposition analysis instead shows that out of the 0.88 mean difference in age at first marriage, -0.12 can be attributed to group differences in explanatory variables that can be explained whereas -0.76 is unexplained. Similarly, out of the 10 year mean difference in age at first marriage between early and late married women, -2.12 is explained by differences in explanatory variables whereas -8.04 remains unexplained.

Twofold Decomposition for Mean age at marriage in 2005-06 vs 2019-21

group.weight	coef(explained)	se(explained)	coef(unexplained)	se(unexplained)
0.0000	-0.04074	0.0090	-0.8418	0.0139
1.0000	0.0008	0.01299	-0.8834	0.0166
0.5000	-0.0199	0.0091	-0.8626	0.0139
0.1447	-0.0347	0.0116	-0.8478	0.0155
-1.0000	-0.1209	0.0085	-0.7616	0.0123
-2.0000	-0.0282	0.0086	-0.8543	0.0136

coef(unexplained A)	se(unexplained A)	coef(unexplained B)	se(unexplained B)
-0.8418	0.0139	0.0000	0.0000
0.0000	0.0000	-0.8834	0.0165
-0.4209	0.0069	-0.4417	0.0083
-0.7199	0.0020	-0.1279	0.0141
-0.6514	0.0105	-0.1103	0.0018
-0.0000	0.0000	-0.8544	0.0136

Twofold Decomposition for comparison of mean age at marriage for women marrying before 25 years of age vs after 25 years of age

group.weight	coef(explained)	se(explained)	coef(unexplained)	se(unexplained)
0.0000	-0.3608	0.0169	-9.8003	0.0223
1.0000	-1.4628	0.0100	-8.6983	0.0182
0.5000	-0.9118	0.0097	-9.2493	0.0177
0.9273	-1.3828	0.0157	-8.7784	0.0214
-1.0000	-2.1175	0.0156	-8.0436	0.0204
-2.0000	-1.3036	0.0088	-8.8575	0.0178

coef(unexplained A)	se(unexplained A)	coef(unexplained B)	se(unexplained B)
-9.8003	0.0223	0.0000	0.0000
0.0000	0.0000	-8.6983	0.0180
-4.9001	0.0116	-4.3491	0.0090
-0.7121	0.0207	-8.0662	0.0013
-0.5844	0.0031	-7.4591	0.0195
-0.0000	0.0000	-8.8575	0.0178

From the graphs (in Appendix), it can be inferred that among the explanatory variables, having secondary or higher education and living in urban areas are the most important determinants of the differences in age at first marriage between the two survey groups. The effect of higher education is even more prominent in explaining the differences in age at first marriage between early and late married groups. Caste and Religion also play a significant role in explaining some of the group differences although the magnitude of effect is smaller.

While the explanatory variables included in model explain only a small portion of the differences observed in age at first marriage, the remaining difference in coefficients can be attributed to two main things. Firstly, it is important to recognize that some of the potential explanatory variables could not be included in decomposition analysis due to data limitations. The trends we observe might also be driven by infrastructural, economic, or cultural transformations that occurred concurrently with increased educational access, which is not being fully captured in this model. For instance, urbanization and industrialization can result in women's marriage postponement by raising their access to education and non-familial paid work and also by reducing the cultural restrictions for women's engagement in socio-economic activities. A future investigation using different measures of development (e.g., percentage urbanized, per capita GDP), etc. could isolate the effect of these economic changes from the effect of education expansion. Delays in entry time to risk of marriage due to longer schooling years, increased awareness of the negative consequences of early marriage and pregnancy, better knowledge of contraceptive methods, higher opportunity cost of raising children, and higher bargaining power in fertility decisions for more educated women are all possible channels through which education would influence the age at marriage and first-birth.

In developed countries like USA, there has been a progression towards cohabitation meaning couples move from dating to living together, which then may or may not lead to marriage. While cohabitation is still not a generally accepted phenomenon in India, cohabitation or 'live-in' relationships are not entirely absent especially in urban India. This may partially explain delay in marriage for urban women if they choose to cohabit with their partners in quest of finding the 'right one' before deciding to tie the knot thereby postponing when they get married. The measure of age at marriage in the DHS does not distinguish between cohabitation and marriage; rather, this variable represents the age at which a woman began co-habitating or got married (whichever occurred first). With no alternative, I adopt this definition of marriage, however, we can expect that women in cohabitation may on average marry later than women directly entering in nuptial relationship.

The use of contraceptives gives young adults the possibility to postpone childbearing and avoid unplanned pregnancy that may compel them to enter into marriage earlier than desired. Even though it is still considered a taboo, premarital sex is increasingly becoming common especially in urban India. These two shifts combined could mean that young adults may delay partnership formation and live in single-hood for longer. However, the use of contraceptives and age at first sex could be endogenous to marital decisions of women and hence I do not include them as covariates in survival analysis model. One might expect that girls belonging to female headed households are comparatively immune from patriarchal norms and may choose to delay their marriage, however I do not have information on head of household for women from their own home before marrying into another household, hence I do not include sex of household head in analysis to avoid endogeneity.

In developing countries like India where dowry practice is common, unmarried women may take up employment to provide for their own dowries. In such a scenario, work may be viewed as a way to facilitate marriage rather than as an alternative to it. Data on dowries is not readily available in India, however one may expect that girls may have to work to arrange for not just the wedding expenses but also dowry and this might possibly push their marriage age forward if they need more time to make the dowry arrangements.

Exposure to media (newspapers/radio/TV) can be another determinant of the hazard risk of marriage, especially if it is used as means of creating awareness about the consequence of early marriage. However, whether women had frequent access to media before marriage is not known and hence, I do not include this variable in my model to avoid endogeneity bias. The experience of the rapid development of the internet in adolescence could potentially alter the gender attitudes and social norms making it more gender egalitarian. Future research may explore a more comprehensive dataset and directly examine the impact of digital technology on gender attitudes and other social and cultural norms.

Similarly, wealth index can be potentially used as an explanatory variable to assess if women belonging to richer households marry later than women belonging to poor households. Unfortunately, DHS does not provide data on personal wealth of women, nor does it provide information on wealth levels of women in their own maternal households before moving to their partner's household. Thus, I do not include wealth index or household assets measure in my model.

It is to be noted that changes in marriage timing patterns may also depend on exogenous factors like changes occurring in labor market, individual preferences for career development, duration of economic crises, etc. which potentially affect marriage conditions. However, keeping in mind the limitations of data, these factors could not be included in the model.

Taking into account these factors that were omitted from model, the remaining differences in coefficients should also be attributed to changing social norms and increasing societal acceptance for late or non-marriage. Unfortunately, I cannot fully isolate the effect of changing social norms in my model due to omission of above-mentioned variables. However, incorporating some of these variables could help extract the remaining effect of social norms on age at first marriage, but this is left for future study.

4.5 Conclusion

In this paper I adopted a descriptive regression analysis approach to analyze the trends of single-hood and age at first marriage in India. I also employed survival analysis and decomposition analysis to test the determinants of increased age at marriage. Cox proportional hazard model was used to study the effect of demographic, socioeconomic and socio-cultural variables, and to identify the magnitude and significance of their effects on age at first marriage. The hypothesis is that higher education levels, increased employment opportunities, urbanization should be some of the key factors encouraging women to marry later. I observed a small yet significant increase in proportion of never married women, as also women postponing their marriage age. Out of all the observable factors facilitated by data, I found higher education among women to be the dominant contributor in delay and decline of marriage among women.

Since the start of 1990s, the Indian economic system has undergone structural changes; it has become more urbanized; literacy rates among women have increased and level of educational attainment has gone up. However, there hasn't been a corresponding steep reduction in proportion of women marrying early or never marrying. Modern societies seem to be more and more demanding towards women by expecting them to be working and at the same time caring for others. Women's traditional roles as mother and wife are still valued over their professional attainments. Although women are 'expected' to marry at some point in their lives the societal pressure brought to bear on such views may not be as condemning as it was in the past. Marriage is likely to be delayed when it is incompatible with the attainment of one's personal goals and aspirations.

Even though the education reforms have resulted in universal primary education in India, a lot of girls drop out of school before completing their secondary or higher education. Provision of schools and basic amenities in schools, well-paying employment opportunities and strengthening of social infrastructure could encourage girls to pursue further studies especially in rural areas thereby preventing early marriage among girls. Especially in economically backward regions with high early marriage rates, efforts to increase education and employment opportunities should be

given more focus to reduce incidences of early marriage of girls due to poverty. In regions with more conservative and traditional gendered norms, there may be additional restrictions on women's mobility and higher cases of violence against women which may cause challenges to participate in the job market and eventually resulting in women succumbing to marriage pressure. However, we need better measures of gender norms and violence to establish a causal link with age at marriage.

Early marriage may interrupt accumulation of higher education, labor market skills, resulting in under-investment in human capital formation for females leading to their lower earnings in the labor market. Thus, marital delay is expected to increase the wages of women. This can be tested by examining the effect of age at first marriage on wages for women, however this is beyond the scope of the paper. Similarly, it has been observed that in countries with regulations around paid maternity leave, employees are often unwilling to recruit young women who have not had children yet to avoid the potential costs in future. It may therefore be of interest to study how employment patterns and wages are influenced by marital and motherhood status of women. The "motherhood earnings penalty" is a well-established finding in many Western countries. Whether family choices have consequences for women's employment and earnings in later life is not well known. However, if a motherhood penalty does exist then it is interesting to study if women delay their marriage to potentially delay their fertility and thereby delay when they have to face this penalty. This is a question I leave for further research.

It is also interesting to assess how singles fare in terms of personal wealth accumulation compared with their married counterparts. Wealth is an important measure of well-being since it provides resources to maintain living standards during economic hardship and is an important source of funds for living expenses during old age and for inter-generational transfers. However, due to limitation of data on personal wealth of individual women, this analysis was not included in this paper, though it constitutes a crucial area for future research. Additionally, if later and less marriage are indeed related to changing attitudes toward marriage and gender roles, irrespective of the educational attainments of women, then an important task for subsequent research would be to explore and clarify the relationship between attitudes and marriage timing by using direct measures of attitudes toward marriage and gender roles.

Age at first marriage is an important indirect determinant of age at first birth and in the long run it influences the number of children a woman bears through her reproductive cycle, especially when usage of contraceptives or any active fertility control is limited or absent. In addition to formal education, females should be encouraged to participate in economically productive activities, which will help them to generate income, as well as increase autonomy and empowerment. Motivating females to get involved in skill-based education and then, in the labor force is one of the

key challenges for policymakers. Late marriage permits women to complete their education, build labor force skills, and develop career interests that compete with childbearing within marriage.

Infrastructure development especially in rural areas can have important indirect implications for marital outcomes. Construction of all-weather roads for example can enable higher income generation for farmers, manual laborers, etc. Besides, access to all-weather roads can make commute to schools and nearby urban areas easier, thereby increasing educational outcomes and job opportunities for the youth. It may be interesting to study the impact of infrastructure development programs on marital outcomes of women , however, this is left for future research.

Chapter 5

Dissertation Conclusion

This dissertation combines topics of finance, gender and agricultural policies in developing countries (India and Uganda). In chapter 2, I studied the channel of a growing financial institution in the form of Savings groups in the context of Uganda using National Panel Survey data from 2010-2020. The period corresponded to a rapid expansion of savings groups and I employed a TWFE model that exploited the variation in the timing of arrival of a savings groups in a community. I observed that savings groups significantly increase household's average value of assets, however they do not have a significant impact on monthly household earnings or on household quality. I also found a significant positive effect of savings groups on food security characterized by variety of food consumed and expenditure on food especially for households not facing any shocks during any given year. I also explored bulk purchasing as a potential mechanism to explain improvements in food security without any real increase in income. I observed that bulk quantities do trade at a discount to small quantities which means that households with savings would be less sensitive to price volatility. However, I did not find a significant evidence of increase in bulk purchases of storable foodstuffs due to exposure to savings groups. My study provides a comprehensive assessment of the impact of savings groups on households in Uganda, offering valuable insights into the potential of informal financial institutions to promote economic empowerment, asset accumulation and to enhance food security among the vulnerable households.

In Chapter 3, joint work with Pramod Dudhe, we study the exogenous policy variation impact of the SSWA act aimed at groundwater conservation on input, yield and profitability outcomes. Treatment effect was negative for total irrigation hours (per hectare) which confirms our hypothesis that the act reduced dependency on irrigation by aligning the cropping season with monsoon rains. We

find a significantly negative effect on the crop yield and crop revenue of farmers. The effects are stronger for paddy compared to all other crops.

In Chapter 4, I adopted a descriptive regression analysis approach to analyze the trends of singlehood and age at first marriage in India. I also employed survival analysis and decomposition analysis to test the determinants of increased age at marriage. Cox proportional hazard model was used to study the effect of demographic, socioeconomic and socio-cultural variables, and to identify the magnitude and significance of their effects on age at first marriage. The hypothesis is that higher education levels, increased employment opportunities, urbanization should be some of the key factors encouraging women to marry later. I observed a small yet significant increase in proportion of never married women, as also women postponing their marriage age. Out of all the observable factors facilitated by data, I found higher education among women to be the dominant contributor in delay and decline of marriage among women.

Taken together, this dissertation fills the gap in applied micro-econometric literature, which is largely focused on advanced economies. It is achieved by leveraging panel extensive datasets combined with newest developments in econometric estimation to understand the role of informal savings groups on household welfare, role of information acquisition in agrarian policies and developments in women's marital decisions as they juggle between empowerment and social norms.

Appendix A

1. Linking Household IDs across survey years

The UNPS survey revised the manner in which the unique HH - ID key was designed multiple times across years. Hence before performing any data cleaning and analysis operations, it is important to have consistent HH - ID keys across survey years to panelize them. For the years that UNPS updated their HH- ID keys, they also provided the old linking keys to link households across years. I first do the linking process in ascending order of time. To elaborate, I take the oldest appearing household ID key and year by year merge the updated HH - ID keys with the help of the linking keys. This gives me the updated HH-ID keys for the oldest appearing households in survey. Then I perform a reverse merge-key operation in descending order of time. This is important since there could be households that did not appear in the oldest survey period but appeared in later survey years. Thus, I want to trace these households as further back in time as possible in my data. Here, I start with the HH - ID key of households in the most recent survey year and then year by year I merge the older HH - ID keys associated with them. Finally, I get a comprehensive set of HH- ID keys which are linked across time and I give a new set of unique HH - ID keys to each household which is now consistent across survey years. In this way I use the linking HH - ID keys and create a new column of HH - IDs that are consistent across the years for each household.

2. Conversion of Matooke quantities to standardized units of measurement

Table 1: Conversion of Matooke units in standardized units of measurement

Matooke	Number of plaintains	Total weight
Bunch Big	165-234	41.25 - 70.2 kg
Bunch Medium	88-140	22- 42 kg
Bunch Small	48-70	12- 21 kg
Cluster Large	15-18	3.75 - 5.4 kg
Cluster Medium	11-14	2.75 - 4.2 kg
Cluster Small	8-10	2 - 3 kg
Heap Large	10-12	2.5 - 3.6 kg
Heap Medium	6-8	1.5 - 2.4 kg
Heap Small	4-5	1 - 1.5 kg

3. Explanation of demographic variables from household module

I create 4 age bins and generate a count variable to track the number of household members who belong to each of the age bins. I obtain information on key demographic variables from the 'Demographic' module of the survey data. I first use the column of 'Relationship to HH head' to track the personal identifier observation corresponding to HH head. I then use that identifier to fetch the demographic information for HH head. I recode the Sex of HH head variable to be 1 if the household is headed by a Female and 0 if the household is headed by a Male. I also recode the educational status of Hh head to be 1 if the HH head has at least some level of schooling and 0 otherwise. For HH head that are married I recode the dummy to be 1 and if the HH head is single or divorced or widowed I recode the dummy to 0.

4. Linking Community survey with household survey

Since community module does not have information about households, I need to use a common linking factor that links community module with household module. For the initial 2 ways, this linking factor was in the form of community codes. Thus, I could merge the two modules using community codes and link household information with savings group information. For the subsequent survey waves, I have parish codes instead as the linking factor between the two modules. Thus for subsequent waves, I use parish codes to merge the community section with the household section.

It is to be noted that sometimes, for the same parish name, a different code could be assigned in the survey in the later survey years. Thus for the same household belonging in the same parish, if the parish code has changed, then I create a new column of parish codes, wherein I update the parish code with the original value rather than the new value. For households that move parishes, I create another new column of parish codes wherein I retain their original parish records to link the household module with the community module, and extract savings groups information corresponding to the original parish.

Finally, I append all the files to get the linked community-household information across all survey years. To have as minimum missing values as possible, for the households which belong to same parish, but have missing information over the years, I make sure that parish code is still the same and then populate the preceding year information in subsequent years for these observations. In the final linked file, all households have exactly 1 unique parish associated with them which enables me to track changes in household consumption and welfare before and after the entry of savings groups into the parish. All households which belong to a given parish, will have the same savings groups information corresponding to them. Once I have the completely linked file, I create a variable which records the first year of occurrence when a savings group entered the given parish, for each parish code. This is the year when the given parish gets exposed to at least 1 savings group for the first time. If in the panel, a given parish is never exposed to any savings group, then the variable will take the value of 9999. Lastly, I create my treatment variable *'SavingsGroups'* which takes the value of 1 if the year of interview is greater than or equal to the year of exposure to savings group and 0 otherwise.

5. Bulk purchase - Price Relationship

Table 2: Fixed-Effects regression of Price per kg in log on bulk purchases

Variables	(1)	(2)	(3)
Bulk purchase	-2.497***	-2.521***	-1.931***
	(0.038)	(0.038)	(0.040)
SavingsGroups	0.015	0.017	0.034
	(0.037)	(0.037)	(0.037)
Bulk savings interaction	-0.045	-0.030	-0.084**
	(0.049)	(0.049)	(0.040)
Observations	49,542	49,537	49,537
R-squared	0.585	0.606	0.688
Mean of Outcome	6.774	6.774	6.774
Time F.E.	YES	YES	YES
HH Controls	YES	YES	YES
Parish F.E.	YES	YES	YES
HH F.E.	NO	YES	YES
Product F.E.	NO	NO	YES

(Notes for table: Outcome variable represents the log of price of product per kg. Column 1 represents regression with time and parish fixed effects, column 2 includes household fixed effects as well while column 3 also adds product fixed effects. The HH controls includes variables like sex of HH head, marital status and educational status of HH head, age composition brackets for household members, etc. Standard errors are clustered at the parish level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

6. Goodman Bacon Decomposition Analysis

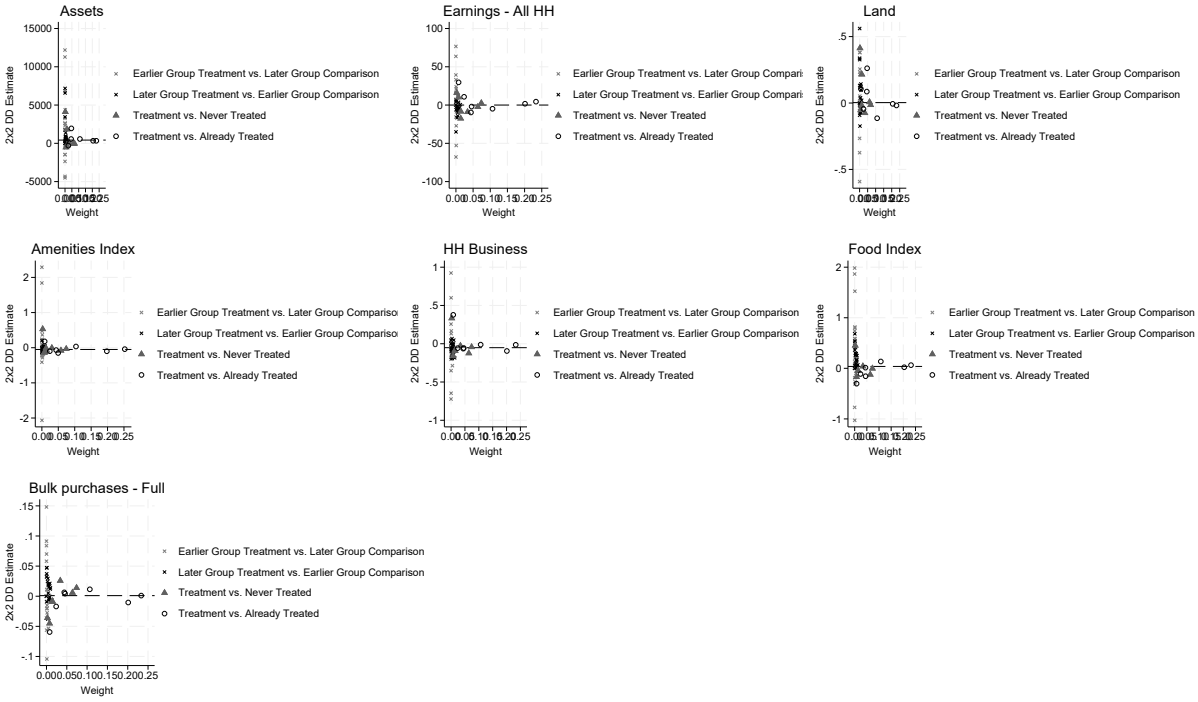

Figure 1: Goodman Bacon Decomposition of weights

Table 3: Goodman Bacon Decomposition of Weights

Variable	Diff-in-Diff Estimate	DD Comparison	Weight	Average DD Estimate
Assets	420.329	Earlier T vs. Later C	0.031	775.415
		Later T vs. Earlier C	0.097	547.175
		T vs. Never treated	0.203	118.587
		T vs. Already treated	0.669	477.244
Land	0.003	Earlier T vs. Later C	0.031	0.068
		Later T vs. Earlier C	0.098	0.027
		T vs. Never treated	0.201	0.002
		T vs. Already treated	0.67	-0.003
Amenities	-0.05	Earlier T vs. Later C	0.03	-0.05
		Later T vs. Earlier C	0.091	-0.024
		T vs. Never treated	0.2	-0.042
		T vs. Already treated	0.68	-0.055
Earnings	-0.035	Earlier T vs. Later C	0.03	-1.652
		Later T vs. Earlier C	0.097	-3.25
		T vs. Never treated	0.21	-2.459
		T vs. Already treated	0.662	1.278
HH-Business	-0.052	Earlier T vs. Later C	0.03	-0.078
		Later T vs. Earlier C	0.097	-0.073
		T vs. Never treated	0.21	-0.071
		T vs. Already treated	0.662	-0.041
Food Index	0.037	Earlier T vs. Later C	0.03	0.179
		Later T vs. Earlier C	0.096	0.189
		T vs. Never treated	0.21	-0.039
		T vs. Already treated	0.664	0.033
Bulk Purchase	0.001	Earlier T vs. Later C	0.03	-0.013
		Later T vs. Earlier C	0.097	0.008
		T vs. Never treated	0.21	0.008
		T vs. Already treated	0.662	-0.002

T: Treatment, C: Comparison

Note: *bacondecomp* package in Stata works only for balanced panel, hence I use the *ddtiming* package instead which works for unbalanced panel. The second column is obtained by multiplying average DD Estimate with corresponding weight for each comparison group and then summing it over the comparison groups. As observed, most of the weights come from T vs Already treated which could be an issue if we have heteregeneous treatment effects over different cohorts of treatment implementers.

7. Heterogeneous treatment-effects regression

Table 4: Heterogeneous treatment-effects regression for average treatment effects on the treated over time

Variables	ATET	Robust SE	p-value	Observations
Assets	394.459	(266.498)	0.140	8,741
Land dummy	0.019	(0.036)	0.593	8,740
Amenities Index	-0.007	(0.043)	0.861	8,739
Earnings	0.327	(2.919)	0.911	8,751
HH Business dummy	-0.077	(0.034)	0.023	8,751
Food Index	0.046	(0.064)	0.467	8,575
Bulk Purchase	0.007	(0.007)	0.377	96,256

(Notes for table: Above table is obtained by running the extended TWFE estimator by Wooldridge (2021) to compute the average treatment effects that control for heterogeneity across time and cohort for unbalanced panel. Standard errors are clustered at the parish level. Treatment level is considered at the HH level. Control group is taken to be never treated group. Heterogeneity is accounted for both at the treatment cohort and treatment time)

Appendix B

So far we presented district-level results, which offer cleaner identification due to localized rainfall variations. In this section, we transition to state-level analysis, which enables us to compare pre- and post-policy differences but at the cost of noisier rainfall measures. We use the following baseline estimating equation for our regressions:

$$y_{icvst} = \delta_t + \delta_s + \delta_c + \beta_2(PostAct \times PolicyState \times RainfallShock) + \beta_3 X_{st} + \beta_4 y_{icvs(t-1)}. \tag{1}$$

Explanation of subscripts:
s: state
$PostAct = 0$ for year until 2007, 1 for 2008 onward for Punjab and 2009 onward for Haryana

The interaction of Post-Act and PolicyState with Rainfall shocks significantly reduced irrigation hours to -13.889 but had no clear effect on irrigation expenditure. Yield effects were statistically insignificant, while revenue increased significantly to Rs. 4,228.719.

The policy's interaction with rainfall shocks had stronger effects on paddy cultivation. Irrigation hours declined even further to -16.076, and there was no significant impact on yield. Revenue impacts were negative but statistically insignificant. Unlike paddy, other crops saw a milder insignificant decline in irrigation inputs to -1.365 hours but a significant increase in yield to 1.240 quintals per hectare and insignificant increase in revenue to Rs. 6054.586.

One important thing to understand is that the negative revenue effects at the district level are more accurately capturing the localized impact of rainfall variation post-policy. In contrast, the observed positive revenue effects at the state level might be capturing broader economic or market-level compensatory factors rather than direct farm-level productivity changes. The positive revenue effects seen at state level could reflect structural adjustments post-policy, such as shifts in cropping patterns, adaptation strategies, or market price responses that are not as evident in the

district-level data.

Table 5: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Monsoon season using state level rainfall data

Variables	Irrigation Hrs		Irrigation Rs		Yield (Qtls perha)		Revenue (deflated Rs)	
SSWA x JuneRains	-15.069***	-13.889***	-291.932	-171.831	0.842	0.231	4,964.695***	4,228.719***
	(3.593)	(2.392)	(183.098)	(222.754)	(0.790)	(0.660)	(1,425.027)	(1,172.898)
Mean of Dep Var	48.34	48.34	874.23	874.23	28.35	28.35	47,821	47,821
Observations	47,167	47,167	47,167	47,167	47,167	47,167	47,167	47,167
R-squared	0.706	0.711	0.563	0.585	0.758	0.779	0.653	0.707
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crop FE	No	Yes	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes
WCB pval	0.292	0.114	0.623	0.676	0.281	0.108	0.211	0.243

The set of controls include rainfall levels in all other months, rain-crop interaction of total annual rainfall with paddy and wheat respectively, and crop area under cultivation. We restrict the sample to crops that are grown in monsoon season. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level. Since we have 20 state clusters, we also run the Wild-Cluster Bootstrap on top of the fixed effects model to get adjusted p-values.

Table 6: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Monsoon season using state level rainfall data for just Paddy

Variables	Irrigation Hrs		Irrigation Rs		Yield (Qtls perha)		Revenue (Rs)	
SSWA x JuneRains	-16.723***	-16.076***	-400.128	-223.343	-0.223	-0.382	-2,147.836	-750.705
	(4.457)	(3.976)	(412.483)	(396.874)	(1.059)	(1.032)	(1,578.335)	(1,073.737)
Mean of Dep Var	59.57	59.57	931.96	931.96	34.76	34.76	49,723.48	49,723.48
Observations	30,846	30,846	30,846	30,846	30,846	30,846	30,846	30,846
R-squared	0.727	0.731	0.578	0.598	0.699	0.732	0.720	0.766
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

The set of controls include rainfall levels in all other months, and crop area under cultivation. Here, we filter the data to only include Paddy and drop all other crops. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level. Since we have 20 state clusters, we also run the Wild-Cluster Bootstrap on top of the fixed effects model to get adjusted p-values.

Table 7: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Monsoon season using state level rainfall data for all other crops

Variables	Irrigation Hrs		Irrigation Rs		Yield (Qtls perha)		Revenue (Rs)	
SSWA x JuneRains	-5.102	-1.365	-82.601	-2.522	1.316**	1.240**	8,654.603***	6,054.586
	(3.726)	(2.576)	(186.824)	(138.716)	(0.573)	(0.435)	(1,897.166)	(3,847.559)
Mean of Dep Var	27.11	27.11	765.13	765.13	16.23	16.23	44225.4	44225.4
Observations	16,321	16,321	16,321	16,321	16,321	16,321	16,321	16,321
R-squared	0.504	0.547	0.542	0.577	0.479	0.522	0.606	0.677
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crop FE	No	Yes	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

The set of controls include rainfall levels in all other months, rain-crop interaction of total annual rainfall with paddy and wheat respectively, and crop area under cultivation. Here, we filter the data to exclude Paddy and include all other crops that are grown in monsoon season. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level. Since we have 20 state clusters, we also run the Wild-Cluster Bootstrap on top of the fixed effects model to get adjusted p-values.

Table 8: Regression analysis of effect of SSWA act interacted with rainfall shock on Irrigation inputs, Yield and Revenue - Winter season using state level rainfall data

Variables	Irrigation Hrs		Irrigation Rs		Yield (Qtls perha)		Revenue (deflated Rs)	
SSWA × JuneRains	8.608	10.468	406.360	354.568**	7.520***	4.852***	9,263.087***	4,879.907**
	(8.006)	(7.762)	(249.075)	(157.894)	(1.786)	(1.101)	(1,735.475)	(1,739.889)
Mean of Dep Var	65.24	65.24	1824.5	1824.5	35.64	35.64	54,153.74	54,153.74
Observations	46,603	46,603	46,603	46,603	46,603	46,603	46,603	46,603
R-squared	0.559	0.570	0.605	0.618	0.804	0.861	0.658	0.726
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged outcomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crop FE	No	Yes	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes
WCB pval	0.447	0.334	0.37	0.202	0.448	0.335	0.211	0.243

The set of controls include rainfall levels in all other months, rain-crop interaction of total annual rainfall with paddy and wheat respectively, and crop area under cultivation. We restrict the sample to crops that are grown in winter season. Revenue is deflated using 2015 as baseline. We winsorize the revenue variable by 5% on each side of the tail. June rain is a standardized variable. Standard errors are clustered at the district level. Since we have 20 state clusters, we also run the Wild-Cluster Bootstrap on top of the fixed effects model to get adjusted p-values.

Appendix C

Table 9: Proportion of women according to highest education level attended

		All Women				Never Married Women			
Age Group	**Year**	**No Education**	**Primary**	**Secondary**	**Higher**	**No Education**	**Primary**	**Secondary**	**Higher**
15-19	2005-06	0.1572	0.1390	0.6685	0.0380	0.0966	0.1167	0.7397	0.0469
15-19	2015-16	0.0706	0.0708	0.8159	0.0549	0.0536	0.0637	0.8230	0.0597
15-19	2019-21	0.0441	0.0524	0.8396	0.0638	0.0343	0.0468	0.8511	0.0678
20-24	2005-06	0.2323	0.1247	0.4834	0.1597	0.0556	0.0629	0.5219	0.3596
20-24	2015-16	0.1448	0.1068	0.5189	0.2295	0.0535	0.0524	0.4469	0.4472
20-24	2019-21	0.0894	0.0729	0.5498	0.2879	0.0300	0.0302	0.4388	0.5009
25-29	2005-06	0.3090	0.1311	0.4140	0.1459	0.0589	0.0594	0.4336	0.4480
25-29	2015-16	0.2343	0.1362	0.4664	0.1631	0.0753	0.0545	0.3885	0.4817
25-29	2019-21	0.1581	0.1163	0.5035	0.2221	0.0480	0.0348	0.3509	0.5663
30-34	2005-06	0.3688	0.1398	0.3687	0.1227	0.1075	0.0733	0.4381	0.3811
30-34	2015-16	0.3181	0.1435	0.4276	0.1108	0.1466	0.0807	0.4409	0.3319
30-34	2019-21	0.2438	0.1433	0.4671	0.1458	0.1217	0.0712	0.3823	0.4249
35-39	2005-06	0.4356	0.1483	0.3218	0.0943	0.1583	0.0935	0.4568	0.2914
35-39	2015-16	0.3964	0.1515	0.3709	0.0812	0.2227	0.0979	0.4499	0.2295
35-39	2019-21	0.3298	0.1577	0.4205	0.0919	0.1930	0.1111	0.4485	0.2434
40-44	2005-06	0.4447	0.1697	0.3003	0.0853	0.2422	0.1304	0.4161	0.2112
40-44	2015-16	0.4718	0.1522	0.3130	0.0630	0.3063	0.1130	0.3972	0.1836
40-44	2019-21	0.4152	0.1553	0.3588	0.0707	0.2888	0.1276	0.4198	0.1638
45-49	2005-06	0.4839	0.1692	0.2678	0.0791	0.1979	0.2292	0.3750	0.1979
45-49	2015-16	0.5518	0.1522	0.2512	0.0449	0.3462	0.1340	0.3730	0.1469
45-49	2019-21	0.4975	0.1606	0.2932	0.0487	0.3209	0.1352	0.4000	0.1440

Table 10: Proportion of women by timing of marriage by rural-urban status

Age group	Rural									Urban								
	2005-06			2015-16			2019-21			2005-06			2015-16			2019-21		
	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs
15-19	0.8389	0.1611	0.0000	0.7070	0.2930	0.0000	0.6742	0.3258	0.0000	0.7247	0.2753	0.0000	0.6585	0.3415	0.0000	0.6178	0.3822	0.0000
20-24	0.5809	0.4191	0.0000	0.3938	0.6062	0.0000	0.3636	0.6364	0.0000	0.4263	0.5737	0.0000	0.3039	0.6961	0.0000	0.2904	0.7096	0.0000
25-29	0.5339	0.4460	0.0201	0.4005	0.5724	0.0271	0.3525	0.6125	0.0350	0.3602	0.5994	0.0403	0.2811	0.6637	0.0552	0.2472	0.6850	0.0678
30-34	0.5669	0.3866	0.0465	0.4602	0.4781	0.0617	0.4034	0.5242	0.0724	0.3750	0.5235	0.1014	0.3255	0.5600	0.1145	0.2679	0.5963	0.1359
35-39	0.5705	0.3778	0.0517	0.4683	0.4685	0.0631	0.4460	0.4783	0.0757	0.4235	0.4813	0.0952	0.3505	0.5393	0.1101	0.3203	0.5506	0.1291
40-44	0.5878	0.3670	0.0453	0.4879	0.4539	0.0583	0.4672	0.4714	0.0615	0.4268	0.4957	0.0775	0.3771	0.5272	0.0957	0.3420	0.5487	0.1093
45-49	0.5554	0.3953	0.0493	0.4492	0.4752	0.0756	0.4429	0.4852	0.0719	0.4170	0.5033	0.0797	0.3723	0.5241	0.1035	0.3556	0.5386	0.1058

Table 11: Proportion of women by timing of marriage for women with highest education level attended up to primary schooling

Age group	No Education									Primary Education								
	2005-06			2015-16			2019-21			2005-06			2015-16			2019-21		
	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs
15-19	0.8890	0.1110	0.0000	0.7737	0.2263	0.0000	0.7480	0.2520	0.0000	0.8430	0.1570	0.0000	0.7326	0.2674	0.0000	0.7387	0.2613	0.0000
20-24	0.7295	0.2705	0.0000	0.5342	0.4658	0.0000	0.5093	0.4907	0.0000	0.6304	0.3696	0.0000	0.4929	0.5071	0.0000	0.4818	0.5182	0.0000
25-29	0.6899	0.3051	0.0050	0.5506	0.4385	0.0109	0.5277	0.4594	0.0130	0.5804	0.4111	0.0085	0.5047	0.4829	0.0124	0.4935	0.4928	0.0136
30-34	0.7107	0.2783	0.0110	0.5976	0.3746	0.0278	0.5609	0.4064	0.0327	0.5852	0.3919	0.0229	0.5327	0.4342	0.0331	0.5128	0.4533	0.0338
35-39	0.6943	0.2876	0.0181	0.5756	0.3908	0.0336	0.5721	0.3881	0.0398	0.5705	0.3962	0.0333	0.5130	0.4433	0.0437	0.5117	0.4398	0.0485
40-44	0.6857	0.2951	0.0192	0.5661	0.3971	0.0368	0.5609	0.4030	0.0361	0.5749	0.3935	0.0316	0.5225	0.4341	0.0434	0.5148	0.4386	0.0466
45-49	0.6183	0.3579	0.0239	0.5015	0.4382	0.0603	0.5075	0.4399	0.0526	0.5424	0.4129	0.0447	0.4680	0.4670	0.0650	0.4661	0.4755	0.0584

Table 12: Proportion of women by timing of marriage for women with highest education level attended beyond primary schooling

Age group	Secondary Education									Higher Education								
	2005-06			2015-16			2019-21			2005-06			2015-16			2019-21		
	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs	Before 18yrs	18-25yrs	After 25yrs
15-19	0.7250	0.2750	0.0000	0.6860	0.3140	0.0000	0.6607	0.3393	0.0000	0.4242	0.5758	0.0000	0.3484	0.6516	0.0000	0.3016	0.6984	0.0000
20-24	0.3986	0.6014	0.0000	0.3390	0.6610	0.0000	0.3505	0.6495	0.0000	0.0638	0.9362	0.0000	0.1075	0.8925	0.0000	0.0973	0.9027	0.0000
25-29	0.3339	0.6366	0.0295	0.3089	0.6589	0.0322	0.3070	0.6590	0.0340	0.0387	0.8334	0.1279	0.0660	0.8157	0.1184	0.0664	0.8130	0.1206
30-34	0.3327	0.5788	0.0885	0.3329	0.5835	0.0835	0.3112	0.6089	0.0798	0.0437	0.6747	0.2816	0.0695	0.6625	0.2681	0.0600	0.6586	0.2814
35-39	0.3292	0.5713	0.0994	0.3255	0.5806	0.0939	0.3233	0.5753	0.1013	0.0590	0.6393	0.3017	0.0693	0.6492	0.2815	0.0694	0.6373	0.2933
40-44	0.3411	0.5722	0.0867	0.3297	0.5806	0.0896	0.3223	0.5859	0.0918	0.0594	0.6932	0.2474	0.0802	0.6440	0.2758	0.0670	0.6593	0.2736
45-49	0.3440	0.5589	0.0972	0.3059	0.5849	0.1091	0.3037	0.5878	0.1084	0.0801	0.6823	0.2376	0.0731	0.6447	0.2822	0.0670	0.6534	0.2796

Table 13: Changes in sample of never married women across DHS surveys

Variable	Prop_200506	Prop_201921	Raw_Diff	Proportional_Diff	Std_Error_200506	Std_Error_201516	Num_obs_200506	Num_obs_201516
Highest Education level attended-Primary school	0.1103	0.0573	-0.0530	-0.4805	0.0019	0.0006	26697	169402
Highest Education level attended-Secondary school	0.6261	0.6565	0.0304	0.0486	0.0030	0.0012	26697	169402
Highest Education level attended-High school	0.1449	0.2285	0.0836	0.5769	0.0022	0.0010	26697	169402
Rural/Urban status-Urban	0.4143	0.3812	-0.0331	-0.0799	0.0030	0.0012	26697	169402
Head of household-Female	0.1417	0.1487	0.0070	0.0494	0.0021	0.0009	26697	169402
Wealth Index-Poorer	0.1480	0.1870	0.0390	0.2635	0.0022	0.0009	26697	169402
Wealth Index-Middle	0.1951	0.2026	0.0075	0.0384	0.0024	0.0010	26697	169402
Wealth Index-Richer	0.2367	0.2167	-0.0200	-0.0845	0.0026	0.0010	26697	169402
Wealth Index-Richest	0.3093	0.2354	-0.0739	-0.2389	0.0028	0.0010	26697	169402
Contraceptive Usage-Ever used/using	0.0029	0.0096	0.0067	2.3103	0.0003	0.0002	26697	169402
Contraceptive Intention and use-Using/ intends to use	0.0237	0.0072	-0.0165	-0.6962	0.0009	0.0002	26697	169402
Have Exposure to media	0.8996	0.8836	-0.0160	-0.0178	0.0018	0.0008	26697	169402
Occupational status-working	0.3143	0.1908	-0.1235	-0.3929	0.0028	0.0023	26697	28023
Occupational group-agriculture	0.1605	0.0882	-0.0723	-0.4505	0.0022	0.0017	26695	28023
Occupational group-clerical	0.0138	0.0060	-0.0078	-0.5652	0.0007	0.0005	26695	28023
Occupational group-sales	0.0118	0.0394	0.0276	2.3390	0.0007	0.0012	26695	28023
Occupational group-services hh/domestic	0.0201	0.0136	-0.0065	-0.3234	0.0009	0.0007	26695	28023
Occupational group-skilled/unskilled manual labor	0.1110	0.0228	-0.0882	-0.7946	0.0019	0.0009	26695	28023
Occupational group-professional/technical/managerial	0.0495	0.7716	0.7221	14.5879	0.0013	0.0025	26695	28023
Occupational history-currently employed	0.3144	0.1908	-0.1236	-0.3931	0.0028	0.0023	26697	28023
Occupational history-worked last year	0.0543	0.0493	-0.0050	-0.0921	0.0014	0.0013	26697	28023
Domestic violence justified-Disagree	0.8160	0.8212	0.0052	0.0064	0.0024	0.0023	26671	28023

Table 14: Changes in sample of never married women across DHS surveys - Regression results

Variable	Unadjusted	Adjusted F.E.
Primary education	-0.053***	-0.054***
	(0.002)	(0.009)
Secondary education	0.030***	0.060**
	(0.003)	(0.022)
Higher education	0.084***	0.060***
	(0.003)	(0.009)
Urban	-0.033***	-0.045***
	(0.003)	(0.010)
Household head - Female	0.007**	0.004
	(0.002)	(0.005)
Poorer	0.039***	0.043***
	(0.003)	(0.008)
Middle Income	0.007**	0.012
	(0.003)	(0.010)
Richer	-0.020***	-0.019
	(0.003)	(0.013)
Richest	-0.074***	-0.081***
	(0.001)	(0.002)
Ever used/using contraceptives	0.007***	0.006***
	(0.001)	(0.002)
Contraceptive use and intent	-0.017***	-0.017*
	(0.001)	(0.007)
Have exposure to media	0.000	0.000
	(0.000)	(0.000)
Currently employed	-0.283***	-0.284***
	(0.002)	(0.019)
Employed last year	-0.046***	-0.047***
	(0.001)	(0.006)
DV justified - disagree	-0.680***	-0.679***
	(0.002)	(0.016)

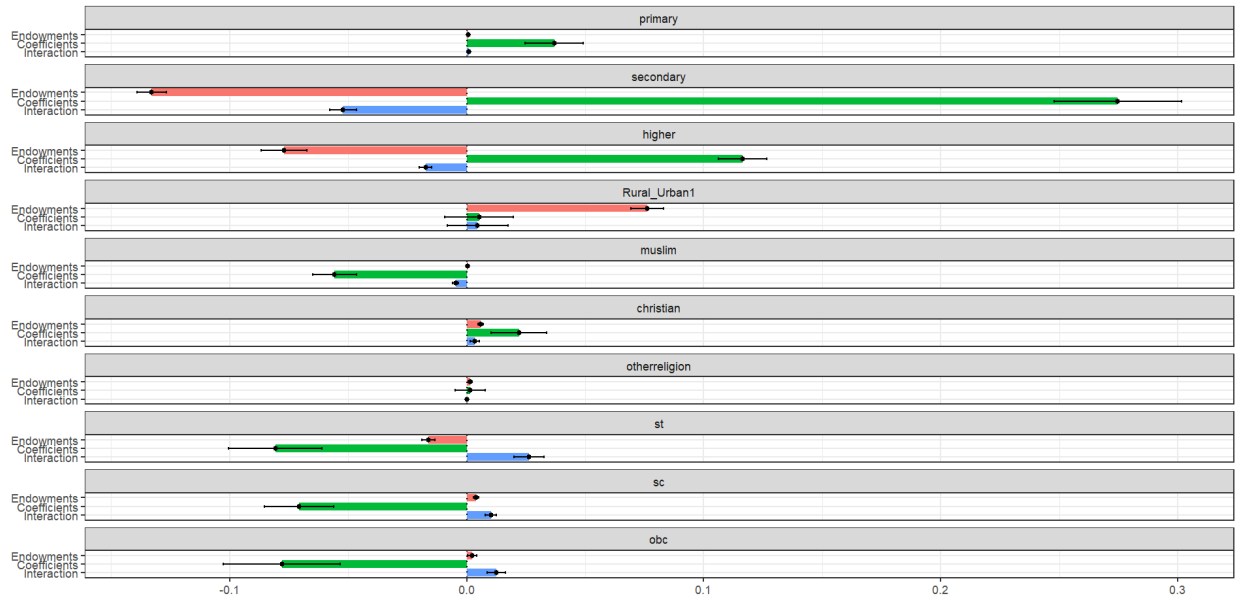

Figure 2: Threefold decomposition analysis using Survey period groups

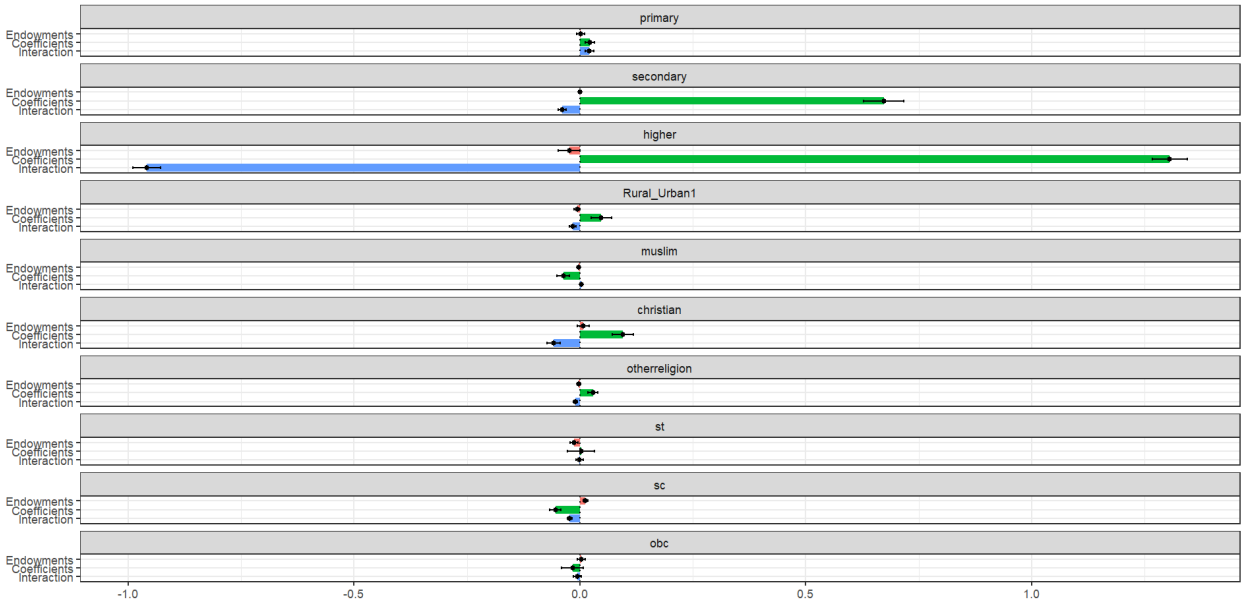

Figure 3: Threefold decomposition analysis using Marriage timing groups

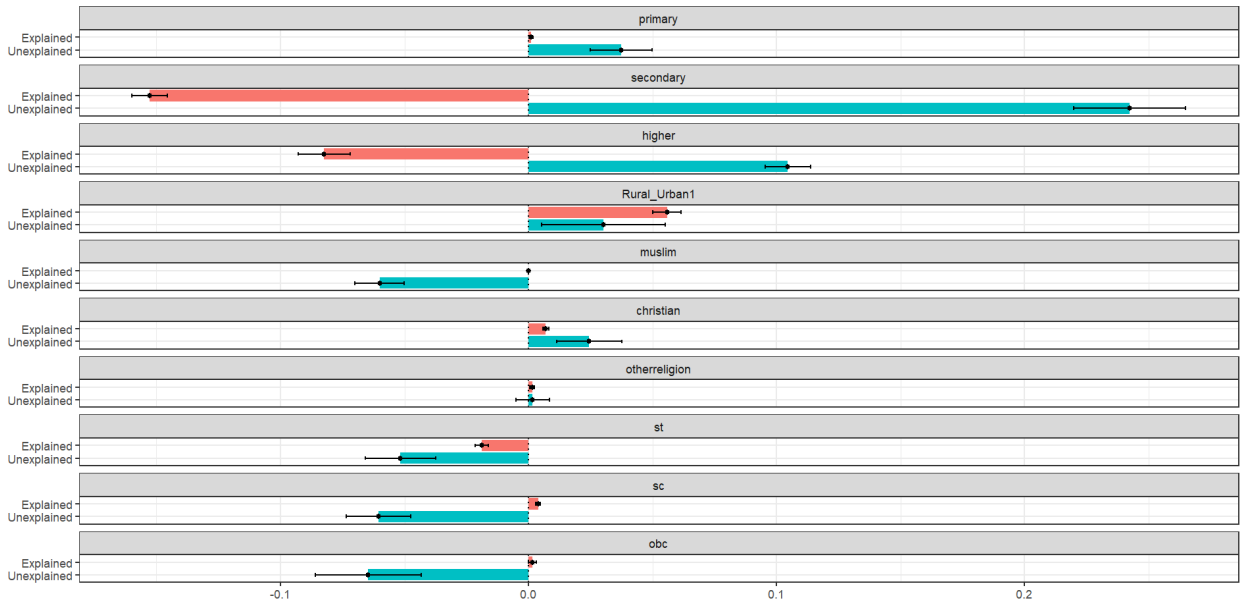

Figure 4: Twofold decomposition analysis using Survey period groups

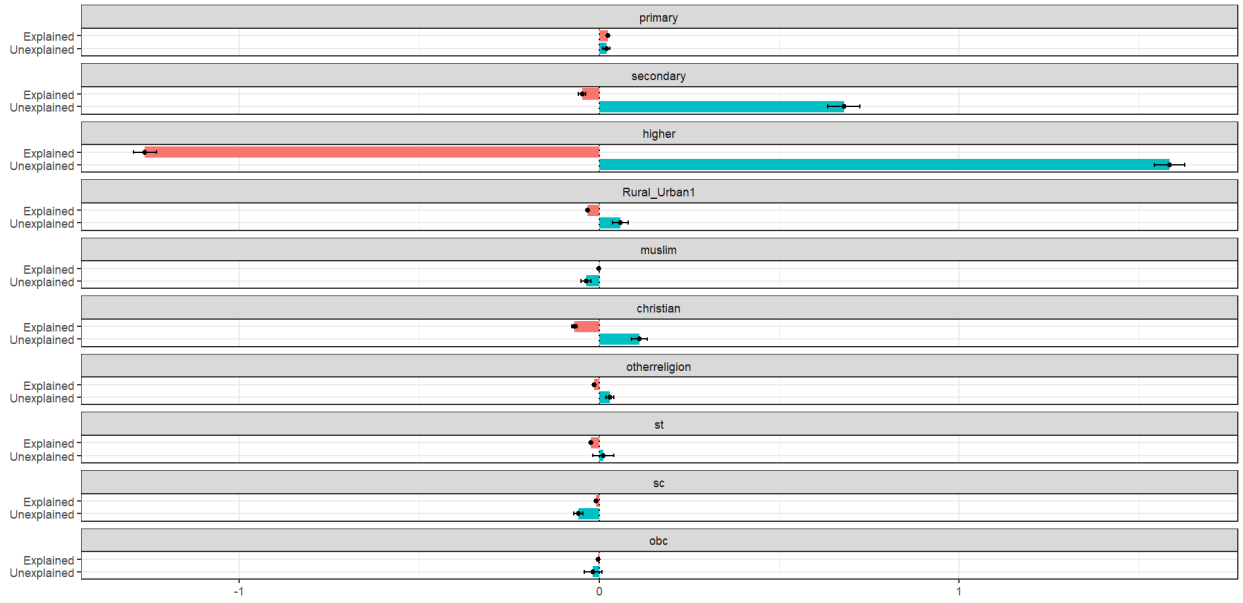

Figure 5: Twofold decomposition analysis using Marriage timing groups

Bibliography

Agarwala, M., Bhattacharjee, S., & Dasgupta, A. (2022): Unintended consequences of Indian groundwater preservation law on crop residue burning. Economics Letters, 214, 110446.

Aggarwal, S., Francis, E., & Robinson, J. (2018): Grain today, gain tomorrow: Evidence from a storage experiment with savings clubs in Kenya. Journal of Development Economics, 134, 1–15.

Amponsah, D., Awunyo-Vitor, D., Wongnaa, C. A., Prah, S., Sunday, O. A., & Acheampong, P. P. (2023): The impact of women groundnut farmers' participation in village savings and loans association (vsla) in northern Ghana. Journal of Agriculture and Food Research, 11.

Anderson, M. L. (2008): Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. Journal of the American Statistical Association, 103(484), 1481–1495.

Annan, J., Bundervoet, T., Seban, J., & Costigan, J. (2013): A randomized impact evaluation of village savings and loans associations and family-based interventions in Burundi. Evaluation Report.

Bass, J., Murray, S., Cole, G., Bolton, P., Poulton, C., Robinette, K., Seban, J., Falb, K., & Annan, J. (2016): Economic, social and mental health impacts of an economic intervention for female sexual violence survivors in Eastern Democratic Republic of Congo. Global Mental Health.

Beaman, L., Karlan, D., & Thuysbaert, B. (2014): Saving for a (not so) rainy day: A randomized evaluation of savings groups in Mali. Technical report, National Bureau of Economic Research.

Becker, G. S. (1973): A theory of marriage: Part i. Journal of Political economy, 81(4), 813–846.

Berg-Cross, L., Scholz, A.-M., Long, J., Grzeszcyk, E., & Roy, A. (2004): Single professional women: A global phenomenon challenges and opportunities. Journal of International Women's Studies, 5(5), 34–59.

Blinder, A. S. (1973): Wage discrimination: Reduced form and structural estimates. Journal of Human resources, 436–455.

Burlando, A., & Canidio, A. (2016): The allocation of capital in rural credit markets. Journal of International Development, 28(8), 1381–1395.

Burlando, A., Goldberg, J., & Etcheverry, L. (2020): Banking the group: Impact of credit and linkages among Ugandan savings groups. University of Oregon Working Paper.

Burlig, F., Jina, A., Kelley, E. M., Lane, G. V., & Sahai, H. (2024): Long-range forecasts as climate adaptation: Experimental evidence from developing-country agriculture. National Bureau of Economic Research.

Chahal, G., Sood, A., Jalota, S., Choudhury, B., & Sharma, P. (2007): Yield, evapotranspiration and water productivity of rice (oryza sativa l.)–wheat (triticum aestivum l.) system in Punjab (India) as influenced by transplanting date of rice and weather parameters. Agricultural water management, 88(1-3), 14–22.

Corno, L., Hildebrandt, N., & Voena, A. (2020): Age of marriage, weather shocks, and the direction of marriage payments. Econometrica, 88(3), 879–915.

Dixon, R. B. (1971): Explaining cross-cultural variations in age at marriage and proportions never marrying. Population studies, 25(2), 215–233.

Dupas, P., & Robinson, J. (2013): Why don't the poor save more? evidence from health savings experiments. American Economic Review, 103(4), 1138–1171.

Ferguson, S. J. (2000): Challenging traditional marriage: Never married Chinese American and Japanese American women. Gender & Society, 14(1), 136–159.

Flynn, J., & Sumberg, J. (2018): Are savings groups a livelihoods game changer for young people in Africa? Development in Practice, 28(1), 51–64.

Frisancho, V., & Martin, V. (2020): Savings groups, risk coping and financial inclusion in rural areas. IDB Publication, Working Papers, (10899).

Gash, M. (2017): Understanding the impact of savings groups. The SEEP Network.

Gash, M., & Odell, K. (2013): The evidence-based story of savings groups: A synthesis of seven randomized control trials. SEEP Network, 3613–3.

Goodman-Bacon, A. (2021): Difference-in-differences with variation in treatment timing. Journal of econometrics, 225(2), 254–277.

Greaney, B. P., Kaboski, J. P., & Van Leemput, E. (2016): Can self-help groups really be "self-help"? The Review of Economic Studies, 83(4), 1614–1644.

Gui, T. (2020): "Leftover women" or single by choice: Gender role negotiation of single professional women in contemporary China. Journal of Family Issues, 41(11), 1956–1978.

Guilmoto, C. Z. (2012): Skewed sex ratios at birth and future marriage squeeze in China and India, 2005–2100. Demography, 49(1), 77–100.

Heath, R. (2014): Women's access to labor market opportunities, control of household resources, and domestic violence: Evidence from Bangladesh. World Development, 57, 32–46.

Hendi, A. S. (2019): Proximate sources of change in trajectories of first marriage in the United States, 1960–2010. Demography, 56(3), 835–862.

Hlavac, M. (2022): Stargazer: Well-formatted regression and summary statistics tables [R package version 5.2.3]. Social Policy Institute. Bratislava, Slovakia. https://CRAN.R-project.org/package=stargazer

Jensen, R. (2012): Do labor market opportunities affect young women's work and family decisions? experimental evidence from India. The Quarterly Journal of Economics, 127(2), 753–792.

Jensen, R., & Thornton, R. (2003): Early female marriage in the developing world. Gender & Development, 11(2), 9–19.

Kant, Y., Chauhan, P., Natwariya, A., Kannaujiya, S., & Mitra, D. (2022): Long term influence of groundwater preservation policy on stubble burning and air pollution over North-West India. Scientific Reports, 12(1), 2090.

Karlan, D., Savonitto, B., Thuysbaert, B., & Udry, C. (2017): Impact of savings groups on the lives of the poor. Proceedings of the National Academy of Sciences, 114(12), 3079–3084.

Kiiza, B., & George, O. (2021): The impact of savings and credit cooperatives on household welfare: Evidence from Uganda. Journal of Economics and Public Finance, 7(3).

Kirdar, M. G., Dayıoğlu, M., & Koç, İ. (2009): The impact of schooling on the timing of marriage and fertility: Evidence from a change in compulsory schooling law.

Kritz, M. M., & Gurak, D. T. (1989): Women's status, education and family formation in subsaharan africa. International family planning perspectives, 100–105.

Ksoll, C., Lilleør, H. B., Lønborg, J. H., & Rasmussen, O. D. (2016): Impact of village savings and loan associations: Evidence from a cluster randomized trial. Journal of Development Economics, 120, 70–85.

Lamb, S. (2018): Being single in india: Gendered identities, class mobilities, and personhoods in flux. Ethos, 46(1), 49–69.

Loughran, D. S., Zissimopoulos, J. M., et al. (2004): Are there gains to delaying marriage?: The effect of age at first marriage on career development and wages. Rand Labor and Population Santa Monica, CA.

Marguerie, A., & Premand, P. (2023): Savings facilitation or capital injection? impacts and spillovers of livelihood interventions in post-conflict Côte D'Ivoire. Policy Research Working Papers, (10563).

McDonald, A. J., Srivastava, A. K., & Gerard, B. (2019): Tradeoffs between groundwater conservation and air pollution from agricultural fires in northwest India. Nature Sustainability, 2(7), 580–583.

Ministry of Agriculture & Farmers Welfare, G. o. I. (2020): Cost of cultivation surveys (Survey). https://eands.da.gov.in/Cost_of_Cultivation.htm

Nawn, N. (2013): Using cost of cultivation survey data: Changing challenges for researchers. Economic and Political Weekly, 139–147.

Oaxaca, R. (1973): Male-female wage differentials in urban labor markets. International economic review, 693–709.

Pai, D., Rajeevan, M., Sreejith, O., Mukhopadhyay, B., & Satbha, N. (2014): Development of a new high spatial resolution (0.25$\times$ 0.25) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. Mausam, 65(1), 1–18.

Government of Haryana. (2009): THE HARYANA PRESERVATION OF SUBSOIL WATER ACT, 2009 [Accessed: (04.14.2025)].

Government of Punjab. (2009): THE PUNJAB PRESERVATION OF SUBSOIL WATER ACT, 2009 [Accessed: (04.14.2025)].

Rosenzweig, M. R., & Udry, C. R. (2019): Assessing the benefits of long-run weather forecasting for the rural poor: Farmer investments and worker migration in a dynamic equilibrium model. National Bureau of Economic Research.

Shah, F., Coulter, J. A., Ye, C., & Wu, W. (2020): Yield penalty due to delayed sowing of winter wheat and the mitigatory role of increased seeding rate. European Journal of Agronomy, 119, 126120.

Singh, K. (2009): Act to save groundwater in Punjab: Its impact on water table, electricity subsidy and environment. Agricultural Economics Research Review, 22, 365–386.

Srivastava, S., Chand, R., Singh, J., Kaur, A. P., Jain, R., Kingsly, I., & Raju, S. (2017): Revisiting groundwater depletion and its implications on farm economics in Punjab, India. Current Science, 422–429.

Tesfaye, W., & Gebremariam, G. (2020): Consumption smoothing and price enhancement motives for grain storage: Empirical perspectives from rural Ethiopia. Agricultural and Food Economics, 8(1), 25.

To, S. (2013): Understanding sheng nu ("leftover women"): The phenomenon of late marriage among Chinese professional women. Symbolic Interaction, 36(1), 1–20.

Tripathi, A., Mishra, A. K., & Verma, G. (2016): Impact of preservation of subsoil water act on groundwater depletion: The case of Punjab, India. Environmental management, 58(1), 48–59.

Wooldridge, J. M. (2021): Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. Available at SSRN 3906345.

Yamaguchi, T. (2006): 'Loser Dogs' and 'Demon Hags': Single women in Japan and the declining birth rate. Social Science Japan Journal, 9(1), 109–114.