Livelihood in a Climate-change Vulnerable Region: Evidence from the Sundarbans in India[1]

Miki Khanh Doan, Aleskandr Michuda, Heng Zhu, Anubhab Gupta, Binoy Majumder

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

[1] Miki Khanh Doan (Email: mkdoan@ucdavis.edu) is a graduate student at the Department of Agricultural and Resource Economics at University of California, Davis. Aleksandr Michuda (am2497@cornell.edu) is an assistant research professor at Cornell University's Center for Data Science for Enterprise and Society. Heng Zhu (Email: hwzhu@ucdavis.edu) is a researcher. Anubhab Gupta is an assistant professor at the Department of Agricultural and Applied Economics at Virginia Tech (Email: anubhab@vt.edu). Binoy Majumder is a consultant at The Researcher, West Bengal, India (Email: this.is.binoy@gmail.com).

Livelihood in a Climate-change Vulnerable Region: Evidence from the Sundarbans in India

Miki Khanh Doan, Aleskandr Michuda, Heng Zhu, Anubhab Gupta, Binoy Majumder

5-18-2022

Abstract

Climate change has caused a disproportionate burden on the poor in developing countries,who are primarily reliant on subsistence and small-scale agriculture for their livelihood. This paper contributes to our understanding of the livelihood diversification strategies of a vulnerable coastal population living in the Sundarbans region of India, an archipelago in the Bay of Bengal in India and Bangladesh that has been predicted to be on the brink of the largest exodus of the human populace. Using novel high-frequency weekly data collected over a year on a representative sample of households, this paper utilizes k-means clustering and unsupervised time-series clustering techniques to provide new insights into how poor households adapt to climate change. Our analysis highlights that households rely on multiple income sources and cope from week to week through mechanisms such as local borrowing, spousal income, or changes in expenditure. Households that rear livestock have volatile yet relatively high female income.

1 Introduction

Climate change has led to an increased occurrence of rising sea tides, frequent coastal cyclones, and other natural disasters, which in turn threaten people's livelihoods and displace people from their homes. Recent studies such as Callaghan et al. (2021) find evidence that up to 85% of the global population has been impacted by climate change. However, the poor in developing countries bear a disproportionate burden of climate change without the resources to cope with these adverse shocks (Füssel (2010); Sperling (2003)). Using high-frequency weekly household-level data collected over a year, we discuss the livelihood strategies of poor households from a vulnerable coastal population living in the world's largest delta, the Sundarbans. Increased intensity of coastal cyclones and continuing rising sea levels have already pushed individuals and at times the entire household to migrate out of the delta region. Those remaining are struggling with existing agriculture amid high salinity in agricultural lands and making ends meet with fishing, casual labor, and small businesses. Thus, it is important to understand these divergent livelihood choices which reflect the trade-off poor households have to make to recover from and adapt to recurrent climatic shocks,

We use a novel survey design with high-frequency household-level data to study their adaptation strategies against the changing circumstances caused by climate change. Our sample contains 305 households in 10 villages spanning five administrative blocks in the Sundarbans region of India. Over the course of one year, we collected financial diaries that capture 52 weekly data points for each household on consumption, expenditures, remittances, livelihood activities, and informal borrowing and lending. We first use a cross-sectional k-means clustering algorithm to categorize households into different livelihood strategies. We then leverage the time-series nature of our panel data to conduct a supervised classification based on the variation of household finance within a given livelihood strategy. This technique highlights the critical part of the year for each livelihood strategy and thus allows us to use time variation to explain the divergence of livelihood choices. While k-means clustering has been popular in evaluating livelihood strategies in the past, to the best of our knowledge, our paper is the first to leverage time-series clustering techniques to provide new insights into how poor households adapt to climate change.

The paper is organized as follows. In Section 2, we introduce the Sundarbans region with a description of the location and geography, highlighting the importance of the region. Section 3 describes the study sample and the data collection, particularly how we collected weekly financial information from the sampled households. Section 4 presents results and discussions from the cross-sectional k-means and supervised clustering analysis. The last section concludes.

2 The Sundarbans region

The Sundarbans region, named after the Bengali word Shundorbon for 'beautiful forest', is a delta formed by the confluence of Ganges, Brahmaputra, and Meghna rivers, leading into the Bay of Bengal. The geography largely consists of mangrove forests and mudflats, intersected by streams and channels forming a nutrient-rich region ripe with different varieties of flora and fauna. The Indian Sundarbans covers 40% of the total area of the Sundarbans spanning 4,200 squared kilometers and 54 inhabited islands. Over 4.4 million people reside in the region which has one of the highest rates of poverty in India. According to a report done by World Bank (2014), half of the population lives below the poverty line and nearly 80% of the households primarily engage in inefficient production methods in agriculture, fishing, and aquaculture.

The people in the Sundarbans region continuously and increasingly face a threat to their livelihoods due to soil erosion from rising sea levels and natural calamities in the form of coastal cyclones (Zhu et al. (2018); Mistri (2013)). Due to having a low average elevation of only one meter above sea level, the coastline is estimated to encroach on land at 200 meters annually (Cornforth et al. (2013)). In the last decade or so, this region has witnessed extremely severe tropical cyclones like Sidr (2007), Aila (2009), Fani (2019), Bulbul (2019), and Amphan (May 2020).

More rapid onset events such as cyclones pass periodically through the region, depositing salt on agricultural land in addition to destroying infrastructure. As an adaptation to rapidly declining farmland, those living within the region have sent migrant workers to nearby urban centers in order to support their families. In many of these migrant families, remittances are the primary source of income, and temporary migration acts as an important coping strategy when climate shocks render agricultural work infeasible. Some households located in villages heavily hit by cyclones have responded by uprooting their entire family and moving away altogether, a group often referred to as climate refugees. As a region predicted to be on the brink of having the largest exodus of the human populace (CBS News (2015)), Sundarbans may be seen as an accelerated version of what could potentially happen to other coastal developing countries should events such as rising sea levels fail to be contained.

3 Data and methods

3.1 Data collection

A multi-staged randomization process was utilized in collection of survey data to ensure a representative sample of the population in the region. First, the research team, with assistance from the field team, randomly selected five administrative blocks spanning the two districts (24 Parganas, North and South) in the Sundarbans region in India. Then two villages from each administrative block were chosen to for the survey in such a way that one of the two villages were closer to the coastal rivers and the other farther away. Figure 1 shows the 10 selected villages in the Indian Sundarbans. In the last stage, about 25-30 households from each of the villages were again randomized from the village rosters for the baseline interview, which yields a final sample of 305 households.

We implemented the weekly survey instrument in November 2018 for a full year after collecting a wave of baseline data through a standard survey. During the baseline interview, enumerators trained the sampled households to fill out the financial diaries, including information on their weekly household income, remittance, borrowing, lending, and expenditure on consumption and non-consumption items. The households received two more rounds of training and instructions in the following two weeks before they started independently recording their financial activities in pre-printed diaries handed out to them. The diaries form contained the phone numbers of the enumerator and the field team leader so that the respondent could contact them in case of any follow-up questions on filling up the forms. Additionally, the field team routinely made phone calls to each household to remind them about filling the diaries. At the end of each month, the field team collected the forms from

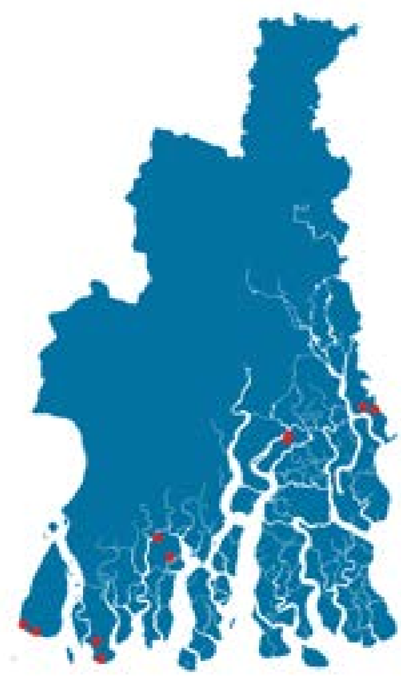


Figure 1: Map of Randomly Selected Villages for Data Collection

the respondent and gave households additional forms to fill up for the following month.

One of the main challenges we faced was incentivizing households to fill out their financial diaries weekly for one year. We approached this issue in two ways. First, we asked the enumerators to take the time to explain the benefits of the exercise and build a good rapport with the households over the study period. The direct benefit includes having a financial record where households could see the inflow and outflow of money within each week. The field team reported that many households were excited to have their school-going children help out with filling out the forms. We also explained to the households how we would use the income and expenditure patterns to understand the financial patterns of communities living in the Sundarbans and better inform research and policies. Second, the field team gave out small gifts (such as emergency lights and umbrellas) to all the households in the sample every 12 weeks.

3.2 Household characteristics

Table 1 summarizes the demographics of households in our sample collected in November 2018. About 91% of the selected households are male-headed. The household head is around 50 years old with 5 years of education. An average household has four to five members with one child. Two-thirds of the households earn income from farming with an average land owned of 32 kathas. About 65% of households have at least one migrant member, and 12% own a business.

Table 1: Summary of Household Demographics

Variables	Mean	*Standard Deviation*
HH head age	49.74	*13.05*
HH head years of education	5.33	*3.77*
Female headed household	0.09	*0.28*
HH size	4.53	*1.83*
# Children	1.08	*1.02*
Prop of HH with migrant	0.65	*0.48*
Farming HH	0.66	*0.48*
HH with business	0.12	*0.33*
Land owned (kathas)	31.72	*50.37*
N	305	

4 Results and discussion

4.1 Typology of household livelihood strategies

In this section, we employ a k-means clustering algorithm over a range of reported household income generating activities. Information on household activities come from both the baseline survey, administered at the beginning of the diaries, and information from the diaries itself. Choice of the k-means approach instead of wards/hierarchical approach was made based on the intended research objective of classifying income generating strategies as being either similar or dissimilar to each other. Conceptually, the k-means approach is simple, and performance of the algorithm is similar to other clustering methods (Kaushik and Mathur (2014)).

A full list of income generating activities is collected and converted to binary indicator variables, indexing whether a particular household engages in said activity. Table 2 below lists the variables used in the k-means algorithm.

Employment activity is dis-aggregated into local employment, or work within the village; and outside employment, which is defined as work inside the Sundarbans, but outside their respective village. Employment includes both full-time employment and casual labor income. Fisherman are defined as those who engage in fishing activities directly, owning their own boats and fishing equipment to catch fish locally in rivers; while fishery workers are those who either work as a deckhand in open ocean vessels, or in the fish processing sector. Some degree of work is provided in the Sundarbans locally through the National Rural Employment Guarantee Act (NREGA). Business indicates whether a household owns or operates a small business while remit, migrant and farming separately indicate whether the household receives

Table 2: Variables Used in K-means Algorithm

Indicator Variable	Details
Local employment	At least one member working in the same village
Outside employment	At least one member working inside Sundarbans, outside village
Fisherman	At least one member does fishing activity
Fishery	At least one member works as deckhand/inside fishery
NREGA	National Rural Employment Guarantee Act work
Business	Household owns and operates a small business
Remit	Household receives remittances
Migrant	At least one member is a migrant, outsie Sundarbans
Farming	Household engages in Farming Activity

remittances, has at least one migrant member (outside Sundarbans), and/or engages in farming, respectively.

The optimal number of clusters was determined by applying the k-means algorithm to a range of possible cluster numbers, in this case 2-12 and finding the elbow point for the Within-Sum-of-Squares (WSS) and where Percentage reduction in error (PRE) is largest, right before a decline. This method indicates that 6 clusters is optimal, table 3 below presents the share of each household cluster that participates in a particular income generating activity.

At the bottom of the table, we name the clusters for ease of reference and calculate an index for the diversification of livelihood activities for that cluster. Recognizing that employment inside the village and in nearby villages can be similar, we take the maximum value of those two variables and add it to the sum of the shares for each of the clusters via

Table 3: Share of Households Participating in Each Income Generating Activity by Clusters

		Cluster ID					
	sample	1	2	3	4	5	6
Local EMP	0.79	0.94	0	0.71	1	0	1
Out EMP	0.36	0	0	1	0.27	0.82	0.03
Fisherman	0.09	0.07	0.36	0.05	0.11	0.09	0.06
Fishery	0.08	0.02	0.50	0.01	0.14	0	0.14
NREGA	0.07	0.09	0	0.05	0.06	0.09	0.09
Business	0.13	0.19	0	0.08	0.11	0.05	0.26
Remit	0.47	0.18	0.43	0.68	0.72	0.73	0.26
Migrant	0.65	0.39	0.07	1	1	0.95	0
Farming	0.65	1	0.14	1	0.42	0	0
N	305	96	14	75	64	22	35
Diversity Index	2.94	2.88	1.50	3.88	3.56	2.73	1.80
Classification		Local work & Farming	Fisherman	Migration & Farming	Local work & Migration	Remittance & Outside work	Local work & Business

the following formula:

$$DiversityIndex = max(LocalEMP, OutEMP) + sum(allotheractivities)$$

Compared across clusters and against the sample average (first column), households classified in the first cluster, “Local work & Farming”, have high shares of local employment and farming, with 94% of households having some wage income from the same village and all households participating in farming. Interestingly, this cluster also has no wage income from other nearby villages nearby, which could indicate relative labor immobility. Close to a third of households in our sample are classified as ”Local work & Farming”.

The second cluster, “Fisherman”, primarily captures households whose livelihood strategies revolve around either fishing or fishery-related work. Asset specificity for fishing equipment is high, and we observe a lower level of diversification for households in this cluster.

Both the third cluster (Migration & Farming) and the fourth cluster (Local work & Migration) have migrants. However, households classified into the Migration & Farming cluster have farming activities, while households in the local work migration cluster rely on employment income from their locale and other regions of India (e.g., the nearby city center of Kolkata).

The fifth cluster, “Remittance & Outside work”, is similar to the “Local work migration” cluster, except most households do not work inside the same village, but rather in other regions of the Sundarbans. The final cluster, ”Local work & Business”, has high shares of local employment and is more likely to own or operate a business, given that ownership of businesses is still relatively low at around 26%.

Diversification of livelihood activities is lowest for fisherman (1.5) and the local work business clusters (1.8); it is highest for the migration farming and local work and migration clusters at 3.88 and 3.56, respectively.

4.2 Classification of Household Behavior across Time

The financial diaries that we collected give us a rich set of high-frequency time series to study. The time series aspect of our dataset allows us to complement our classification of households into livelihood strategies by also looking at their dynamic behavior over the course of the year. The variation within a livelihood strategy sheds light on how households cope from week to week, through mechanisms such as borrowing behavior, spousal income, or changes in expenditure around special events and festivals.

To explore the time series variables, we use a k-means clustering algorithm with a modified loss function called dynamic time warping Cuturi and Blondel (2017). Dynamic time warping is a loss function used in cluster analysis of temporal sequences, where events may occur at different rates or at slightly different times (i.e., when time series are phase-shifted). For instance, dynamic time warping will be able to group households that increase borrowing behavior before planting season, if they increase this behavior before or after each other.[1]

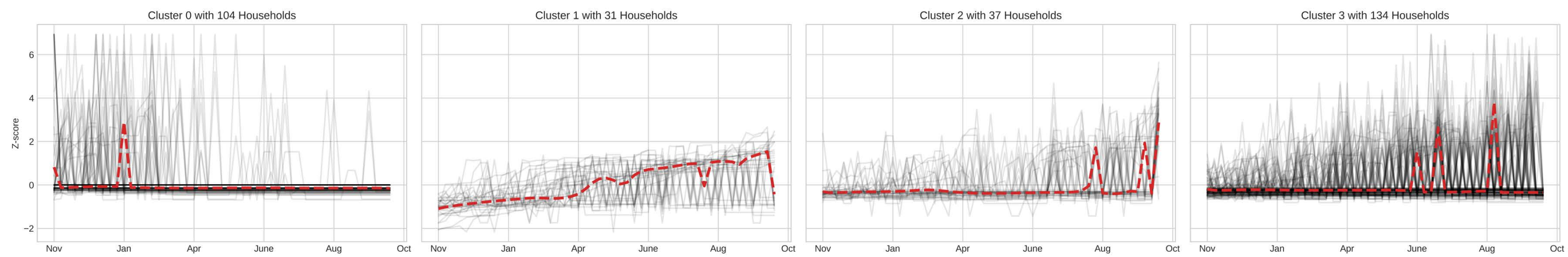
We conduct a time series k-means analysis on every time series variable in the dataset and calculated its silhouette score for 2 to 6 clusters.[2] We present the results of two of these

[1]This is in contrast to the standard k-means context, where the Euclidean distance is used as the loss function. If a practitioner were to use the standard loss function in a time series setting, performance would suffer, as even small changes in the phase of different time series would be construed as dissimilarity. See Rousseeuw (1987) for more information.

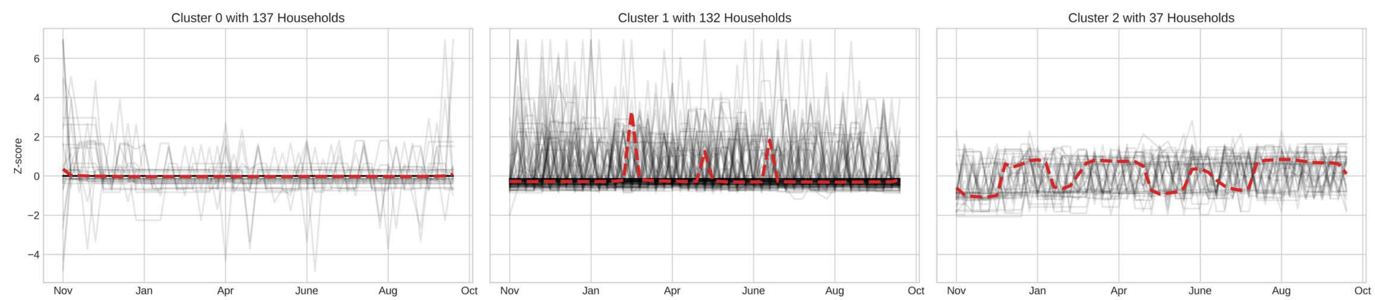
[2]A silhouette score is a typical metric for evaluating how well a clustering algorithm is able to separate the data into groups. This number ranges from -1 to 1, with a score of -1 denoting misclassification of points

variables: cumulative net debt and female income which can be seen in Figure 2. Note that all clustering variables were z-score normalized for comparability.

into clusters, 0 denoting poor separation and 1 denoting strong separation of the data into clusters.



(a) Dynamic Time Warp Clustering of Cumulative Net Debt



(b) Dynamic Time Warp Clustering of Female Income

Figure 2: Dynamic Time Warp Clustering Results

Note: Clustering conducted with k-means using a dynamic time warping loss function. Number of clusters were chosen based on the highest silhouette score chosen from a range of 2-6 clusters. This yielded a silhouette score of 0.18 for cumulative net debt and 0.41 for female income. Cluster variables were z-score normalized for comparability across time series.

Looking at Figure 2a first, we see the results for cumulative net debt, which has four clusters. We can see here that clusters 0, 2, and 3 are mostly characterized by short-time specific increases in net debt that then go back to 0. These are possibly due to small loans taken out before planting season or in preparation for impending festivals and holidays. Cluster 1, however, shows a different pattern, with a steady increase in borrowing across the year and a drop towards the end of the year in October. The increase in borrowing is likely a cyclical process, with households incrementally accumulating loans and then paying them off. We explore possible determinants of this behavior in Figure 4. Cluster 1 is a particularly important cluster to investigate because unlike the other clusters in the data, households in Cluster 1 do not seem to pay off loans and find themselves paying back loans for at least half the year.

For female income, we characterize households by three distinct clusters: cluster 0, where there is a stable flow of female income or no income, cluster 1 where there are short-term increases across the year, and then cluster 2, where there are fluctuations across the year. Cluster 2 is the smallest group amongst the three, but may characterize women in households that are taking part in employment with volatile earnings throughout the year. Cluster 2 is particularly interesting because it seems to suggest a more active female role in the household. We explore the determinants of these clusters in Figure 5. If we now un-normalize female income, we can see how it compares to male income across clusters. Figure 3 shows this relationship. We can see that while male income is higher than female income across all clusters, female income in cluster 2 tends to be higher than female income in the other two clusters. In contrast, although male income for cluster 2 households is more volatile, it isn't significantly different than male income for males in clusters 0 and 1.

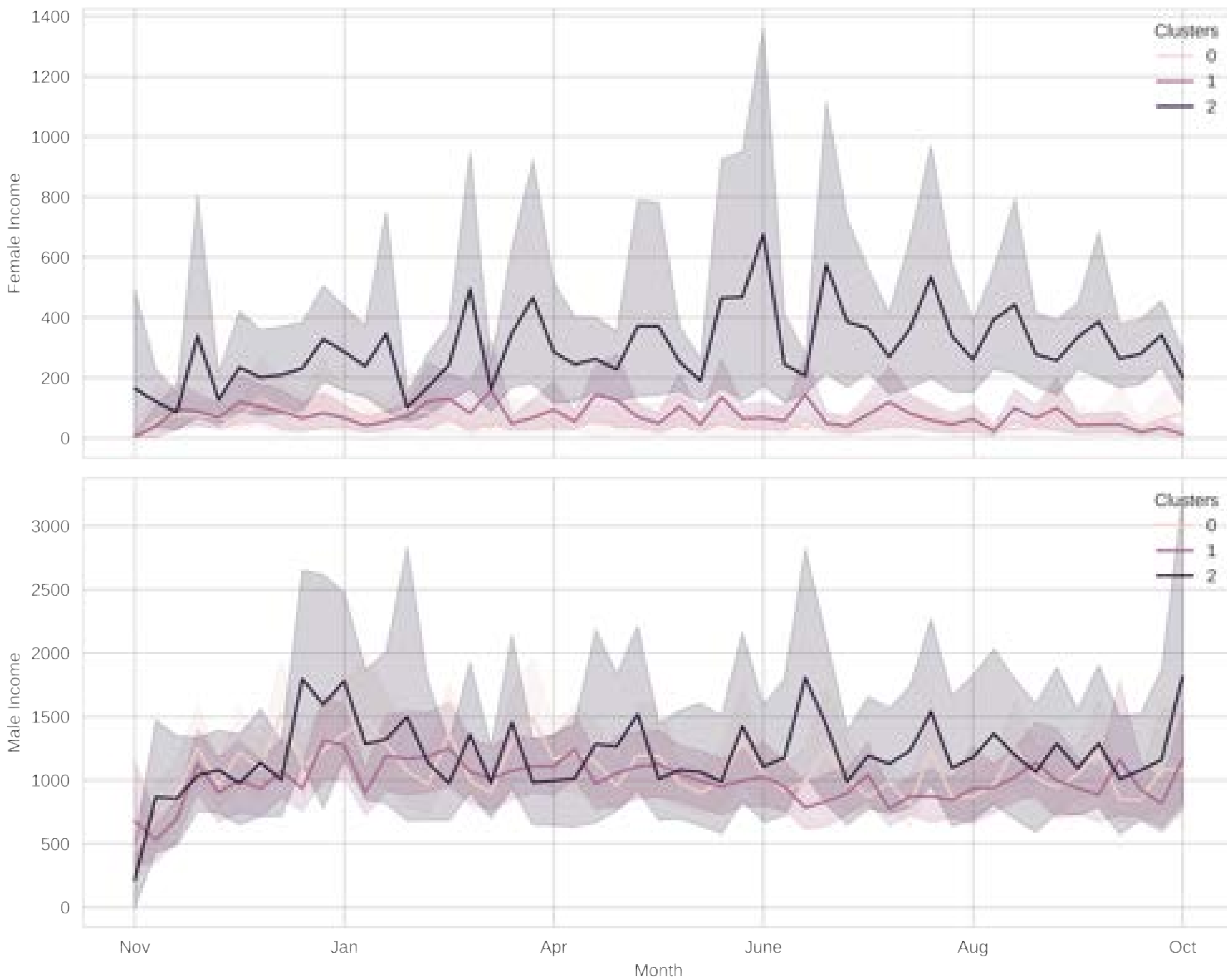


Figure 3: Comparison of Male and Female Income for Female Income Clusters

After running the time series clustering, we conduct a multinomial logit analysis using the time series clusters as dependent variables. We then calculate the marginal effects of some key household variables to understand what drives membership into those clusters. Figure 4 shows marginal effects for the cumulative net debt clusters. We will focus our discussion on trying to understand the determinants of cluster 1. We can see from the household characteristics do not seem to be driving cluster membership, except in whether the household head is married and is a female. Having a female household head and being married decreases the probability of being cluster 1. In terms of education, having a university education decrease the probability of being in cluster 1, while finishing secondary school marginally increases the probability of being in that cluster. Interestingly, having someone in the household who takes part in service work for an NGO or the government increases the probability of being in cluster 1.

We now turn our attention to female income. Having a married household head increases the chances of being in cluster 2 by about 25%. Being a female household head has a positive but insignificant effect on being in cluster 2. Having a migrant in the household increases the probability of being in cluster 2 by about 5%. When considering education, it seems that both a secondary school and university education of the household head leads to a decrease in being in cluster 2. In terms of occupation, if the household has someone working in transportation-related work, the probability of being in cluster 2 increases. Most jarringly, livestock workers and households utilizing the NREGA program decrease the probability of being in cluster 2 precipitously.

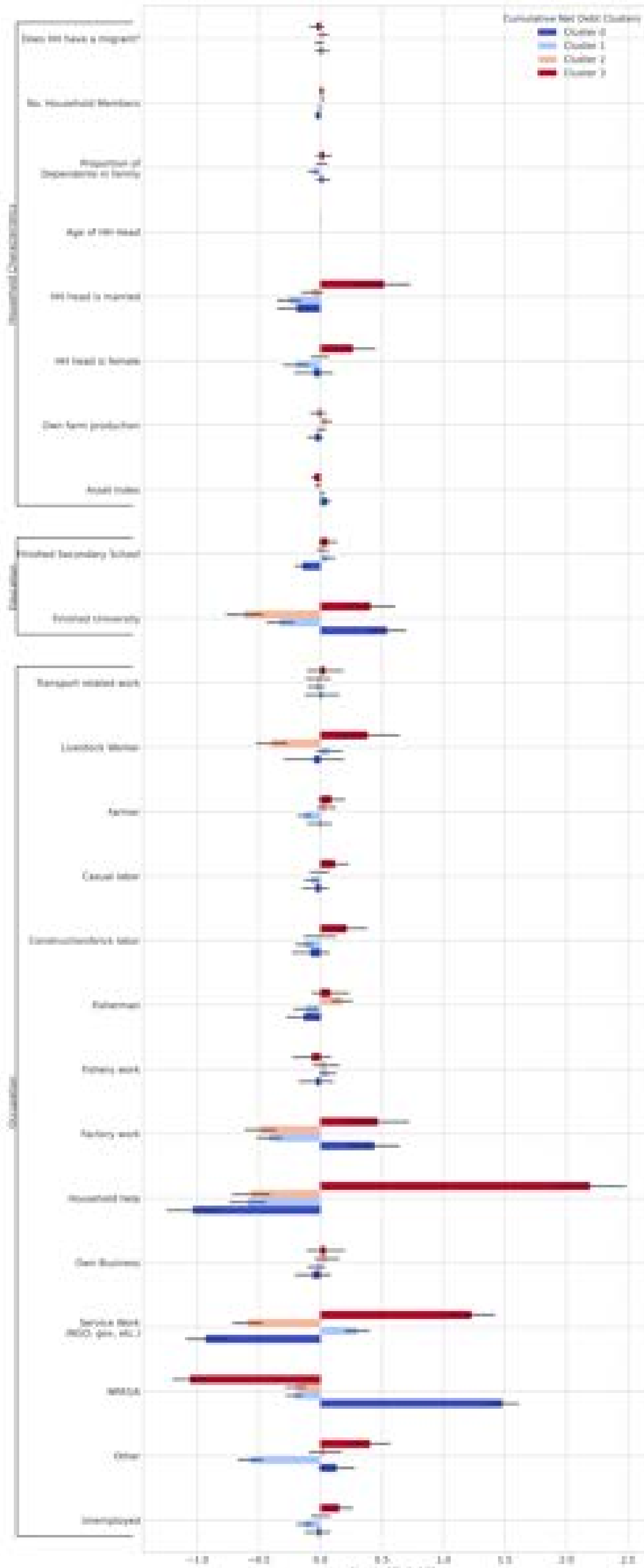


Figure 4: Marginal Effects for Cumulative Net Debt from Multinomial Logit Regression

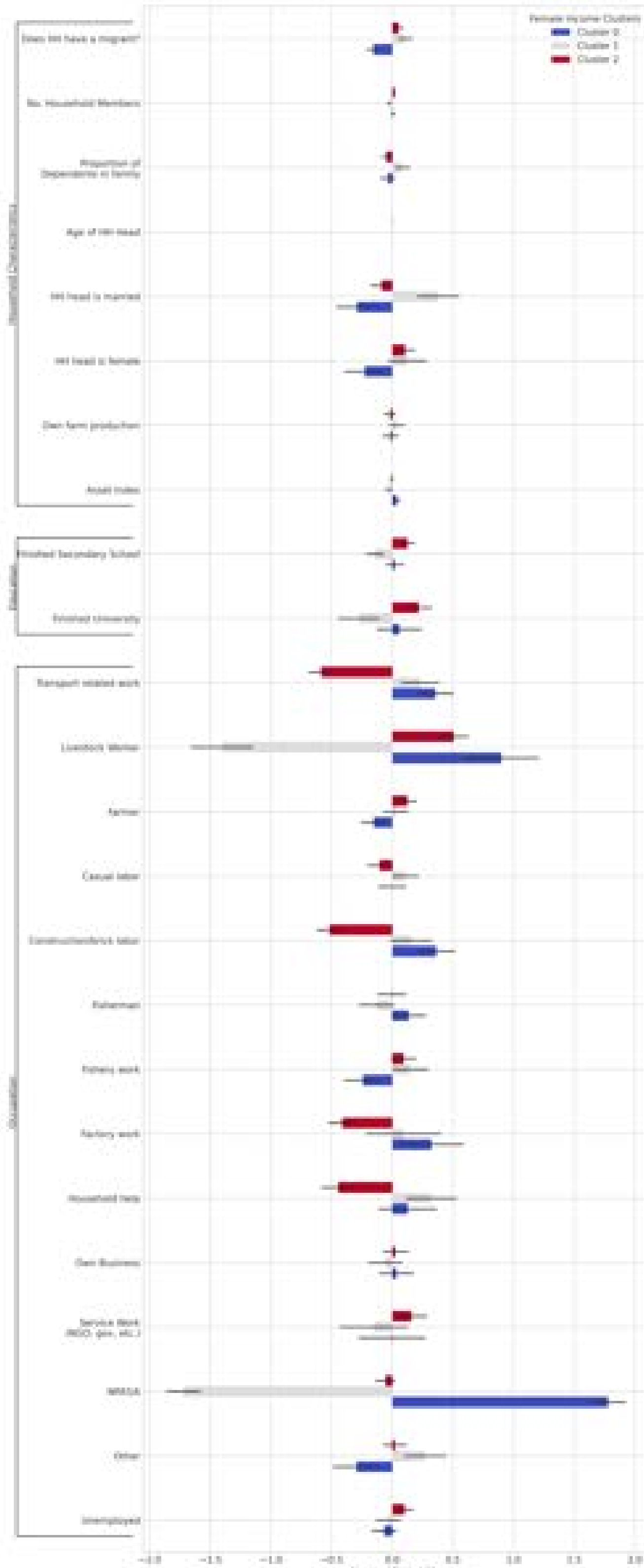


Figure 5: Marginal Effects for Female Income from Multinomial Logit Regression

5 Conclusion

In this paper, we use a high-frequency research design to gain insights into the week-to-week behavior of households and get a glimpse into crucial economic factors, such as expenditures and incomes for all members of the households. As agriculture becomes less viable with a high soil salinity level, households turn to migration and informal borrowing networks.

We leverage unsupervised clustering techniques and explore the livelihood strategies of these vulnerable households in two key ways. First, we use demographic and baseline survey information to cluster households into livelihood strategy groups. The k-means cluster analysis highlights that the sampled households rely on multiple income sources, each with different exposure to climate shocks. Second, we use time series clustering to understand the intra-household income dynamics and debt behavior in these households. We pinpoint a set of households that are particularly prone to indebtedness for the majority of the year. Furthermore, there is some suggestive evidence that households with volatile yet relatively high female income are associated with income from livestock.

Although our focus is on the Sundarbans region, our research design and methods can be applied in other settings to explore a more dynamic aspect of livelihood strategies. Moreover, the livelihood strategies in the Sundarbans are illustrative of a wider phenomenon of household coping mechanism to climate change. Future work will involve expanding the scope of this project, both within and outside India, to get a more externally demonstrative analysis.

References

Callaghan, M., Schleussner, C. F., Nath, S., Lejeune, Q., Knutson, T. R., Reichstein, M., Hansen, G., Theokritoff, E., Andrijevic, M., Brecha, R. J., Hegarty, M., Jones, C., Lee, K., Lucas, A., van Maanen, N., Menke, I., Pfleiderer, P., Yesil, B., and Minx, J. C. (2021). Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature Climate Change*, 11(11):966–972.

CBS News (2015). Sundarbans islands disappear into sea, India, Bangladesh unprepared for 13 million-person exodus.

Cornforth, W. A., Fatoyinbo, T. E., Freemantle, T. P., and Pettorelli, N. (2013). Advanced Land Observing Satellite Phased Array Type L-Band SAR (ALOS PALSAR) to Inform the Conservation of Mangroves: Sundarbans as a Case Study. *Remote Sensing 2013, Vol. 5, Pages 224-237*, 5(1):224–237.

Cuturi, M. and Blondel, M. (2017). Soft-DTW: a Differentiable Loss Function for Time-Series. *Proceedings of the 34th International Conference on Machine Learning*, PMLR 70:894–903.

Füssel, H. M. (2010). How inequitable is the global distribution of responsibility, capability, and vulnerability to climate change: A comprehensive indicator-based assessment. *Global Environmental Change*, 20(4):597–611.

Kaushik, M. and Mathur, B. (2014). Comparative Study of K-Means and Hierarchical Clustering Techniques. *International Journal of Software & Hardware Research in Engineering*, 2(6).

Mistri, A. (2013). Migration and Sustainable Livelihoods: A Study from Sundarban Biosphere Reserve. *Asia Pacific Journal of Social Sciences*, (2):76–102.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C):53–65.

Sperling, F. (2003). Poverty and Climate Change-Reducing the Vulnerability of the Poor through Adaptation, OECD, UNDP, UNEP and the World Bank.

World Bank (2014). Building Resilience for Sustainable Development of the Sundarbans: Strategy Report. Technical report, Washington D.C.

Zhu, H., Gupta, A., Majumder, B., and Steinbach, S. (2018). Short-term effects of India's demonetization on the rural poor. *Economics Letters*, 170:117–121.