Modeling Weather Vulnerability Dynamically: Applications of Multiple Linear Regression to Weather Index Microinsurance[1]

By Sophie Wu

Supervised by Ricardas Zitikis

[1] This is the final project within my Western Undergraduate Student Research Internship (USRI) at the Faculty of Science under the title "Artificial Intelligence and Microcredit: Algorithmic Improvement of Financial Systems in Industrialising Economies."

Acknowledgements

I owe the existence of this research paper to Dr. Ricardas Zitikis, who has been unrelenting in his support of my research interests since our first correspondence. Coming from an unconventional background into statistical research has certainly come with its own set of hurdles, but thanks to my supervisor being a supportive and firm believer in interdisciplinary research, I have never doubted that I have something unique to contribute in my work.

I would also like to give a special thank you to the School for Advanced Studies in the Arts and Humanities, and Dr. Barbara Bruce, experiential learning coordinator at SASAH, who read through my reflections and offered me emotional support throughout the process of producing this paper as well. I am exceptionally grateful to be a part of such an encouraging community as a part of my undergraduate experience.

Abstract

This paper offers a broad overview of the philanthropic goals of microinsurance — namely, to provide vulnerable populations with more self-sufficient and sustainable methods of coping with risk — and through this lens, analyses the applications of multiple linear regression in developing dynamic models for microinsurance. We explain the foundations of MLR (multiple linear regression), and then give two examples for how a simple multiple linear regression model can be adapted with a novel outcome variable (famine) and dependent variables (climate change related costs). Overall, a better understanding of MLR can lend to a better understanding of how microinsurance can scale its practices to new regions. Since this is an overview of the general practice of microinsurance, and not on any particular region or case study, we draw some insights on the practice of microinsurance modeling from some specific regions, such as the Bihar region of India, and illustrate generally how these insights can be used to improve microinsurance broadly.

Introduction

A gap currently exists in microfinance research. There exists a vast pool of statistical research on microfinance, which often focuses on solving specific problems in financial modeling which are related to microfinance. There is also plenty of critical research concerning the practical efficacy of microfinance as a philanthropic tool. Both of these forms of research provide microfinance with a more robust framework for long-term success, but the existence of this distinction speaks to a lack of collaborative and interdisciplinary work currently happening within microfinance. Closing this is essential to solving the dynamic issues within microfinance. Especially given the limited quantitative and qualitative information available to microfinance practitioners (compared to practitioners working in conventional financial markets), there is an essential need to extract substantial relevant insights from the imperfect data available.

Fundamentally, this body of work is a deep-dive into the statistical challenges posed in weather index microinsurance, and hopes to be useful to both statisticians looking to understand the consequences of their work in microinsurance, and microinsurance practitioners looking to understand the relevance of statistical modeling to their work. Although this paper focuses exclusively on a particular form of microfinance, this paper hopes to offer its reader some universal intuition into the statistical practices of microfinance as a whole, especially in regard to the relationship between limited datasets and the social goals of microfinance.

A core overview of microfinance

Microfinance, in offering financial services to individuals who typically lack access to conventional financial services, promises a bipartisan combination of philanthropy and economic development. Microfinance institutions aim to give their clients — typically individuals and households in emerging markets— more flexibility in managing their own finances. Ideally, these services offer microfinance clients both the agency to improve their lives in the short-term and the ability to gain self-sufficiency in the long-term. While microfinance is most associated with microcredit — small low interest loans — credit alone cannot alleviate the gap in financial service availability in developing markets. The world's most vulnerable populations are also the most prone to risk, especially as industrialising countries are especially prone to economic, political, and geographic turmoil. Unmanaged risk can lead to individuals falling into "poverty traps", where in the event of a disastrous event, they may forego decisions that are good for them in the long-term (by selling off valuable assets, taking their children out of school, etc.) in order to cope with the aftermath. (Ahsan, 2009)

Microinsurance

Microinsurance seeks to give households an alternative and more self-sufficient method for coping with these risks. As the name suggests, it offers smaller coverage than a conventional insurance scheme, but at an affordable rate so that poorer clients can access an increased number of risk management tools. Although microinsurance, like all microfinance practices, currently faces a stern critical eye from the philanthropic community as a novel approach to poverty reduction, microinsurance has been shown to reduce the aforementioned unhealthy coping habits which households typically use to adapt to crisis (Carter et al., 2013). Churchill (2006) describes the characteristics of a microinsurance scheme as follows:

(a) transactions are low-cost (and reflect members' willingness to pay);

(b) clients are essentially low-net-worth individuals and households (but not necessarily uniformly poor);

(c) communities are involved in the important phases of the process (such as package design and rationing of benefits); and

(d) the essential role of the network of microinsurance units is to enhance risk management of the members of the entire pool of microinsurance units over and above what each can do when operating as a stand-alone entity.

This list is aspirational — not all microinsurance schemes match all four of these characteristics. Notably, microinsurance schemes are frequently started with profit motivations and some may take in more community feedback than others. But it suffices in showing that the desired impacts of microinsurance are, Ideally, however, in line with the needs and desires of the community.

Weather index microinsurance

Typically, insurance claims are processed on a case-by-case basis, with households or individuals filing claims when they are impacted by a particular unfortunate circumstance covered by their insurance policy. In microinsurance, this poses two distinct problems. Firstly, the cost of processing claims and the time it takes for each claim to be processed mean that customers are paying more for microinsurance, and waiting longer to receive their premiums. It is already paramount to build trust within the clientele, and this can bite away at the perceived value of purchasing microinsurance for future clients. Secondly, it presents the problem of moral hazard — of customers being willing to take on riskier behavior because they are insured (in the case of agricultural insurance, not properly investing in the care of their crops).

Weather index insurance — commonly abbreviated to WII — bypasses these problems by using weather indexes (typically rainfall indexes, although this paper will also explore alternative data sources) to determine payouts. This can result in quick payouts (which are essential in preventing the financial spiral that many individuals face in situations of crisis) and a more "objective" metric for loss that reduces moral hazard, since the actions of any individual client cannot influence the overall outcome of the index.

Index microinsurance also shows significant promise in the field of famine prevention, because of the direct relationship between drought and famine in many of the countries in which microinsurance operates. Historically "most famines in poor economies are associated with the impact of extreme weather (and) the worst famines have been the product of back-to-back shortfalls of the staple crop" (O Grada 2007, p. 7). In countries where this is relevant, having a robust weather index insurance infrastructure can result in much faster and more sustainable recuperation of the agricultural industry, which in turn would result in preventing the widespread famine and food shortages which typically accompany severe weather events.

Designing a basic weather index insurance model

How can we design a basic weather index insurance product? Let's begin by looking at a simplest possible model for index microinsurance - one that is based only on rainfall, where it is assumed that lower rainfall leads to lower production for farmers. It is assumed that an observable relationship must exist between y, the outcome variable of interest (profits for the farmer), and $x_1,\ x_2,\ x_3,\ \ldots,\ x_{\mathrm{i}}$ weather variables. This relationship can be represented through f (some function on $x_1,\ x_2,\ x_3,\ \ldots,\ x$) and ϵ as an error term.

$$y\ =\ f(x_1,\ x_2,\ x_3,\ \ldots,\ x_{\mathrm{i}})\ +\ \epsilon \qquad (1)$$

Assuming that f() is invertible, we can find a "trigger" level (in this context, the highest possible point of rainfall at which payments would begin) which would initiate payments. Payments can be calculated using three parameters — a trigger, exit, and maximum payout. For this example, if rainfall during a particular phase is more than trigger the , no payout will occur for that phase, and if it less than the exit (a level of rainfall so low at which it would d, the maximum payout will be rewarded. If rainfall is between trigger and exit, the payout is linearly interpolated between the zero level payout at the trigger and the maximum payout at the exit (which should be a value corresponding to the amount of expected possible revenue for farmers during the season if weather conditions are optimal) (Osgood et al., 2007)

$$Payout\ =\ ((1\ -\ RainfallSum\ -\ Exit)/(Trigger\ -\ Exit))\ *\ MaxPayou \qquad (2)$$

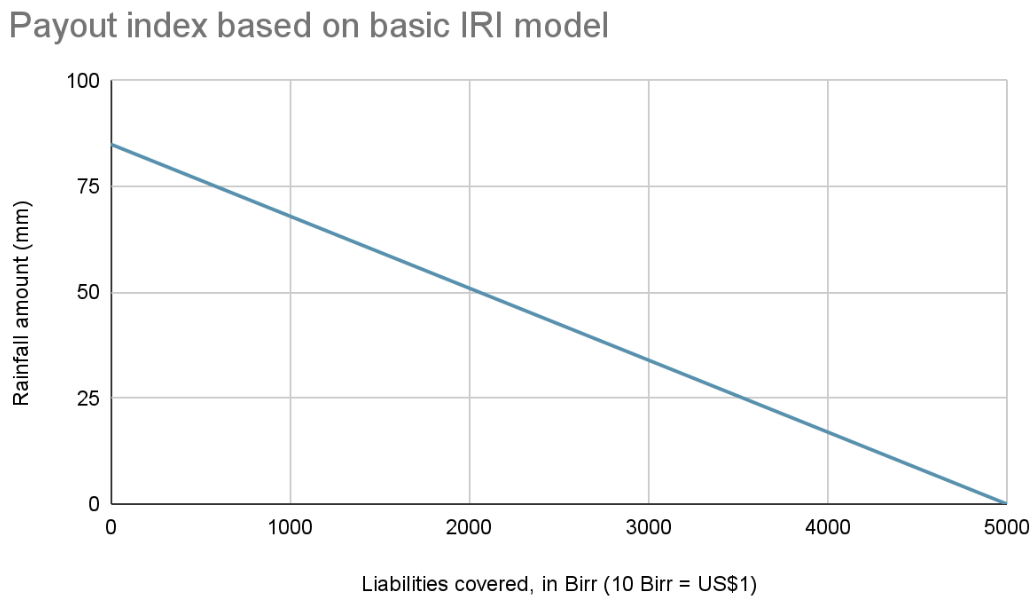

Figure 1. Example payout index based on IRI model

We use a basic contract design by Columbia University's International Research Institute for Climate and Society to illustrate what an example payment index could look like, using 0mm rainfall as an exit, 5000 Birr ($500 USD) as a maximum payout, and 85 mm of rainfall as a trigger level.

This is obviously an especially simple model — other complications can arise through a poor understanding of the relationship between the index and the index-related risk posed. As a trivial example, the previous model will not take into account the fact that too much rain could also impact the Or, what if different irrigation systems are different in different individual households? Basis risk (difference between actual losses incurred to the client and losses that are calculated based on the index) poses one of the largest challenges to index microinsurance. If the client is not paid enough to cover the losses they have felt, the microinsurance scheme has failed to provide this client with a successful risk management tool, and the client's (and the client's community's) trust in microinsurance can falter tremendously. Penetration of index microinsurance is still quite low, and improving the evaluation of basis risk is incredibly important to establishing the trustworthiness of microinsurance. Ideally, through establishing better statistical methods within microinsurance, practitioners can produce both an improved product for their clients, as well as understand better the factors affecting the profits of local agriculture practices and recommend more robust practices to improve local infrastructures.

Improving models using multiple linear regression

All models are wrong, but some are useful.

— George Box

For the victim of a catastrophic and unfortunate event, the cost of such an event may be immeasurable. And even if their cost is measurable, it may also be only ever truly understood by the victim themselves. It is important to the microinsurer to create a model which can dependably and fairly support the insured in the case of insurable events, but it is worth remembering here that any model produced is not meant to represent the exact and individual scale of . It is not the goal of the insurer to cover the exact cost of every particular cost associated with an insurable incident, nor is this a feasible and scalable approach. It is also essential that the core offering of weather insurance — to protect its clients from climate-related costs to the business — is communicated clearly so that there is no confusion as to what qualifies as insurable.

Yet it is crucial for index microinsurance to provide a real and positive value to their clients lives. Actuaries will typically assume, in conventional insurance markets, that more risk averse individuals would be more likely to purchase insurance. After all, insurance protects its clients in the case of a possible unwanted circumstance (which in many cases, could be disastrous to the uninsured individual). However, would-be microinsurance clients in developing markets are actually shown to be *less* likely to purchase microinsurance if they are more risk-averse (Gine, Townsend & Vickery, 2008). Distrust in large and foreign institutions as well as a lack of financial literacy can impact the penetration of a microinsurance product. These factors should not evade the consideration of microinsurance practitioners as they attempt to scale their offerings to a wider audience. But most importantly, if microinsurance can prove to its clients that it can offer protection from risk to its clients at a reasonable price, the significant value of microinsurance will make it a sensible purchase for its client base.

Through emphasizing these values, this chapter will offer a basic introduction to multiple linear regression. It will also give some examples of how multiple linear regression models can be adapted to a variety of relevant circumstances, noting that many of the key relationships in new microinsurance markets remain to be explored due to the recent nature of the industry.

Multiple linear regression

Multiple linear regression expresses the relationship between many independent variables and a dependent variable as such (with i being the number of independent variables):

$$f(x_1, x_2, x_3, \ldots, x_i) = C + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i \quad (3)$$

Where β_i represents a regression coefficient for the independent variable x_i, y represents the dependent variable being measured, and C is the constant of the regression equation. The following paper will use the ordinary least squares method when estimating regression coefficients.

Most commonly, the coefficients are chosen to minimize RSS (residual sum of squares), which represents the sum of squared errors between the model's results and the actual data.

$$RSS = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + \ldots + (y_i - f(x_i))^2 \quad (4)$$

Note that our first model, which only used rainfall as an input and payouts as an output, could be represented as a linear model with only one variable. A key advantage of multiple linear regression is that it is a straightforward method for modeling a dependent variable using more than one independent variable, and the accuracy of a regression model can be improved with the addition of novel independent variables which are relevant to the problem which the model attempts to solve.

Adapting microinsurance using MLR models

This chapter will review two examples in adjusting the application of multiple linear regression in microinsurance. The first will discuss adjusting the outcome variable in the MLR model for a different indicator rather than agricultural output — famine. The second discusses how adding additional independent variables to adjust for climate change can help reduce basis risk, as well as . Both of these examples provide an illustration of how multiple linear regression can be adapted along with the continued practice of microinsurance.

Adjusting the outcome variable for famine

Mid-upper arm circumference is a point of data which is significantly less costly to collect than weight-for-height, and has been found in several studies to be a far better predictor of child mortality than weight-for-height as well (Alam, Wojtyniac, and Rahaman 1989).

As noted before, there is a distinctly high correlation between local famine and the agricultural shock which affects developing economies under climate catastrophes. and then if there is a wide discrepancy between payouts and MUAC, the model can then be revised. If famine were to be measured alongside agricultural output, the proportion of children aged 6-59 months in a community who suffer a mid-upper arm circumference (MUAC) less than a certain threshold, indicating malnourishment, can be used in a separate model to compare with the original model to offer insight into the accuracy of the original model.

This has been a proposed model (Chantarat, et al. 2007) with no attempted implementation yet. This example is highlighted to offer an idea of the flexibility behind a MLR model — when a specific data point becomes relevant to the problem and easy to collect, separate models using the same input variables and separate outcome variables can be developed with ease and cross referenced with the original model to provide payments when appropriate. Obviously, tying additional outcome variables to payments puts the insurer at increased risk. The implementation of such an index implies a strong enough correlation exists between the original outcome variable and the novel outcome variable, and there is also a relationship between the novel outcome variable and the intention of the insurer (in this case, the philanthropic intention of reducing local famine through increasing local food security).

Adjusting the dependent variables for climate change

We can improve our previous model with an understanding that lack of rainfall is not the only meteorological factor which can impact the revenue of a farmer. As a trivial counterexample, excess rainfall can be just as damaging to an agricultural business as shortened rainfall. But there are a variety of other meteorological factors which could hypothetically impact the business of the farmer. Recently, climate change has caused increasing instability in climate predictability. There has also been an increasing recognition by international bodies and academics of the "polluter pays" principle — meaning that local agricultural producers in developing countries should not have to pay for increasing climate-change related costs (Stevens, 1994). Ideally, index insurance payments will reflect more than the initial cost of recuperation which farmers face when dealing with particularly unfavorable conditions, and will also incorporate the cost of developing the infrastructure to cope with future risks. Climate Cost of Cultivation, which has been developed to evaluate both climatic and non-climatic parameters in understanding the additional costs which come to farmers from, aims to address these issues by incorporating a wide variety of timed data points related to global warming.

The following section will review this paper:

"Climate Cost of Cultivation": A New Crop Index Method to Quantify Farmers' Cost of Climate Change Exemplified in Rural India. *Geneva Pap Risk Insur Issues Pract* 41, 280–306 (2016).

We will evaluate the effectiveness of the index developed in this paper, with a focus on how multiple linear regression can offer a model to scale the Climate Cost of Cultivation (CCC) index to different regions outside of Bihar, India (the region examined in the original paper).

Notably, the areas where microinsurance usually operate — remote areas within developing countries — often lack comprehensive data infrastructures. Rain gauge networks have formed the basis of most pilot index insurance programs, which are relatively inexpensive to implement, and can help us form a simple model like the one originally shown in this paper. But since the CCC index requires a significant amount of regional data (as shown in the following table), it would require significantly more initial set-up costs to replicate this exact model.

Category	***Parameter***	***Data type***	***Range***	***Source***
Climatic	Precipitation	Gridded data (0.25° × 0.25°), daily	1979/80–2012/13	Indian Meteorological Department
	Maximum temperature	Gridded data (1° × 1°), daily	1979/80–2012/13	Indian Meteorological Department
	Minimum temperature	Gridded data (1° × 1°), daily	1979/80–2012/13	Indian Meteorological Department
	Relative humidity	Satellite gridded data (approx. 1.9° × 1.9°), daily	1979/80–2012/13	NCEP Reanalysis II
	Wind speed	Satellite gridded data (approx. 1.9° × 1.9°), daily	1979/80–2012/13	NCEP Reanalysis II
	Solar radiation	Satellite gridded data (approx. 1.9° × 1.9°), daily	1979/80–2012/13	NCEP Reanalysis II
	CO2 concentration	CO2 concentration data at the Mauna Loa Observatory, annual	1979/80–2012/13	NOAA-ESRL, ftp://ftp.cmdl.noaa.gov/ccc/co2/trends/co2_annmeaa_mlo.txt
Non-climatic	Soil type	Raster image	2013	National Bureau of Soil Survey and Land Use Planning, India
	Groundwater depth	Groundwater tables, post-monsoon	2005–2013	Central Groundwater Board, India
Data for result validation	Cost of cultivation data	Crop yield and irrigation costs, seasonal	2000/01–2012/13	Directorate of Economics and Statistics, Dept. of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India

Table 1. Data needed to create Climate Cost of Cultivation using Bihar model

Regression can be performed on the climatic and non-climatic factors (as independent variables) to create a model to estimate the climate cost of cultivation (as the dependent variable, which would be equivalent to the payouts) to obtain the coefficients given in the following table:

Dependent variable:	***Plot level net farm income loss***		***R 2***	***0.0091***
Independent variables	***Coefficients***	***Standard error***	***t Stat***	***P-value***
Intercept	393.15	119.97	3.28	0.001052
Seasonal rainfall (mm)	−0.10	0.03	−3.27	0.001087
Pre-season rainfall (mm)	−0.02	0.00	−4.58	0.000005
Seasonal max temp. (°C)	8.40	2.33	3.61	0.000311
Groundwater depth (m)	−1.53	1.72	−0.89	0.373857
AWC (m3/m3)	0.08	0.07	1.15	0.250319
Infiltration (mm)	−6.99	1.91	−3.65	0.000259
CO2 (ppm)	−0.93	0.25	−3.68	0.000236

Table 2. Regression results of climate cost of cultivation using data from table 1

The obtained coefficients tell us that seasonal rainfall, pre-season rainfall, seasonal maximum temperature, and infiltration rate of soil have a statistically significant impact on the cost of cultivation (plot level net farm income loss). The model produced has significantly less basis risk than a typical index insurance product in the region.

It must be emphasized here that climate change can affect a variety of weather-related factors, many of which may not yet be observable — after all, climate change has already been reported to modify the efficacy of irrigation systems (Valipour, 2017), cause distinct precipitation and temperature patterns (Hong, 2016), and decrease agricultural water availability (Valipour, 2016). There are a variety of climate change related factors which are still to be observed which could impact the cost of cultivation for farmers. Additionally, many of the data points in the Bihar dataset may only be relevant to the Bihar region, given the significant variability of different local geographies. As such, it is important for microinsurance practitioners to be able to adapt a model of this form to different regions.

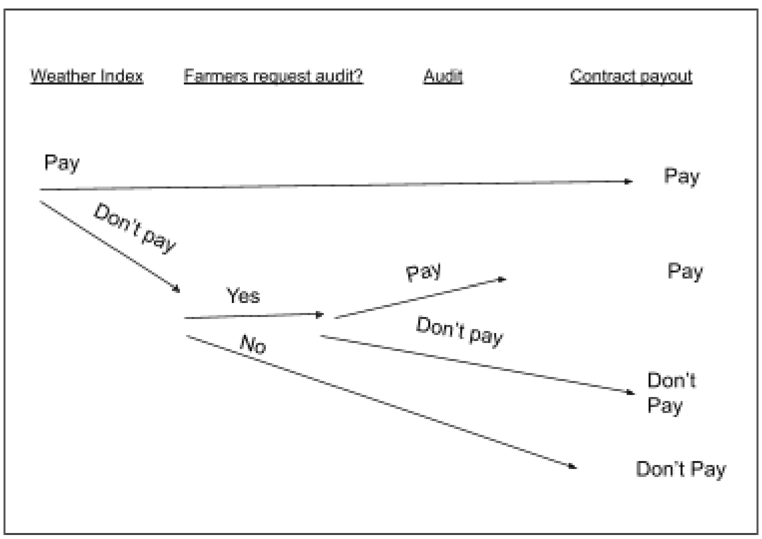

Figure 2. Structure of index based conditional audit contract

The figure above shows how a weather index based conditional audit contract could offer microinsurance practitioners flexibility in modifying their model in the process of auditing farmers. While we may hope that our initial model is as accurate as possible in assessing cost of cultivation, we can also collect cost of cultivation data dynamically in the process of correcting for errors in the model.

The model provided is easily adjustable. If new climatic or non-climatic factors are discovered to be significantly relevant to the evaluation of cost of cultivation in a particular region, they can be added as new independent variables to the model, and proper sensors and measuring processes can be set up after in the region. This ensures that any measurement costs which are established by the microinsurance practitioner are crucial to the actual risk assessment of the insurer, and eliminates the chances of irrelevant data being added to the model.

Conclusion

Although the scarcity of data available in developing markets poses significant challenges to microinsurers, practitioners should not hold their breath. Microinsurance has the potential to offer significant benefits to both its clients and the local communities in which these clients exist, and even if initial modeling may not be perfect, there are many ways to encourage the development of dynamic modeling in the process of building a multiple linear regression model for weather index microinsurance payments. Through a better understanding of these models, the collection of further data points, and the flexibility to incorporate novel understandings of risk and vulnerability in developing markets, microinsurers can build more solvent models over time.

Bibliography

Alam, A., B. Wojtyniac, and M. Rahaman. 1989. "Anthropometric Indicators and Risk of Death." American Journal of Clinical Nutrition 49:884-88.

O Grada, C. 2007. “Making Famine History.” Journal of Economic Literature 95:5-38

Valipour, Mohammad. 2016. "How do different factors impact agricultural water management?" Open Agriculture, vol. 1, no. 1, pp. 89-111. https://doi.org/10.1515/opag-2016-0014

Valipour, Mohammad. 2017. "Global experience on irrigation management under different scenarios" Journal of Water and Land Development, no. 32, pp. 95-102. https://doi.org/10.1515/jwld-2017-0011.

Hong, Moon-Hyun, Kim, Jung-Hwan, Jung, Gyung-Ja, & Jeong, Sang-Seom. 2016. “Rainfall Threshold (ID curve) for Landslide Initiation and Prediction Considering Antecedent Rainfall” Journal of the Korean GeoTechnical Society, 32(4), 15–27. https://doi.org/10.7843/KGS.2016.32.4.15

Fuchs, A., & Wolff, H. 2011. Concept and Unintended Consequences of Weather Index Insurance: The Case of Mexico. American Journal of Agricultural Economics, 93(2), 505-511. Retrieved August 3, 2021, from http://www.jstor.org/stable/41240318

Stevens, C. 1994. Interpreting the polluter pays principle in the trade and environment context. Cornell International Law Journal, 27(3), 577-590.

Giné, X., Townsend, R., Vickery, J. 2008. Patterns of Rainfall Insurance Participation in Rural India. World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/4491 License: CC BY-NC-ND 3.0 IGO.

Osgood, D., McLaurin, M., Carriquiry, M., Mishra, A., Fiondella, F., Hansen, J., Peterson, N., and Ward, N. 2007. "Designing weather insurance contracts for farmers in Malawi, Tanzania, and Kenya," International Research Institute for Climate and Society, Earth Institute, Columbia University: 1-100

Matul, M., & Churchill, C. 2012. Protecting the poor: A microinsurance compendium. International Labour Office - ILO.

Carter, M., Janzen, S., 2013. After the Drought: The Impact of Microinsurance on Consumption Smoothing and Asset Protection

Resume

Sophie Wu

swu493@uwo.ca - (647)-525-5168 - London, Ontario

EDUCATION

University of Western Ontario **Expected Graduation: April 2023**
Mathematics and School for Advanced Studies in the Arts and Humanities, 3.92/4.0 GPA

EXPERIENCE

Planned Parenthood Toronto **June 2018 - September 2018**
Digital Marketing Assistant *Toronto, Ontario*

- Coordinated LGBTQ+ youth advisory council, led the creation of an interactive social media campaign to promote allyship among youth in Toronto

Queen's College Mississauga **September 2018 - August 2019**
ESL Teaching Assistant *Toronto, Ontario*

- Designed and supported lesson plans for ESL learners

Kesvn Studio **May 2021 - July 2021**
Graphic Design Intern *Vancouver, BC*

- Created graphics and developed websites for design clients, produced materials for business development and helped source new clients

VOLUNTEERING

- Executive Director for Canadian University Society for Intercollegiate Debate
- Graphic Designer, CAISA Fashion Show for the Children's Health Foundation
- Volunteer, Inclusive Design Research Centre

MISCELLANEOUS

- Finalist at Yale Hackathon, 2020
- Finalist and Third Speaker at North American Women's Debating Championship, 2021

ASK ME ABOUT

- Skills: Python and R for data science, being crafty with Adobe Creative Cloud, typing really quickly (150 wpm), talking to strangers
- Interests: the intersection between quantitative and qualitative storytelling, empathetic design