Improving the Design of Agricultural Index Insurance

The purpose of this project is to improve the design of the Bank of Thailand's index insurance program. More specifically, I will be working on improving the underlying prediction model used in the program. In index insurance, easily observable quantities (such as rainfall) are used to predict agricultural loss. If the model predicts a large enough loss, the insurer automatically issues out payments. Technically, the problem consists of classifying a time series of remote sensing observations (e.g. rainfall and soil moisture) into different types of losses. One of the biggest problems with index insurance is basis risk, which is the risk that the insured party suffers a loss, but no payment is issued. Traditionally, data scarcity has necessitated the use of simple prediction models with moderate prediction quality. The data in this setting offers an opportunity to use more sophisticated machine learning methods that have achieved state of the art performance in many related tasks. This project could provide valuable information for the design and scaling up of index insurance programs.

Index insurance is a popular way of providing agricultural insurance in developing countries. Index insurance programs have been implemented in a variety of countries (e.g. India, Mexico, Tanzania) and it is estimated that tens of millions of farmers worldwide are covered by such products Greatrex et al. (2015). While there has been a lot of research on the effects of index insurance (see Casaburi and Willis (2018); Karlan et al. (2014); Cai, De Janvry, and Sadoulet (2020)), there has been relatively little research on the design of index insurance. Chantarat et al. (2013) develops a methodology for the design of an index insurance program, and this methodology is what is most commonly used in academic publications discussing the desing of index insurance programs (see Jensen et al. (2019); Flatnes, Carter, and Mercovich (2018)). The government of Thailand has an extraordinarly rich dataset on agricultural losses. This might make it possible to use state of the

art prediction models that were previously infeasible to train due to lack of data. The result of this study can thus be helpful in informing the investment decisions of other countries when designing their own agricultural index insurance programs.

I have been in contact with Dr. Sommarat Chantarat, Research Director at the Bank of Thailand, and with Surasak Choedpasuporn, a Senior Analyst at the Bank of Thailand working on their index insurance program. I have had several meetings with them to gain a deeper understanding of the context. In meetings with Mr. Choedpasuporn, we have reviewed the current methods used for the prediction model, as well as the shortcomings of the model. For example, the model is good at predicting cases of total loss, or of no loss, but is ineffective at predicting cases of moderate loss. The model also performs poorly when there is significant cloud cover.

The Bank of Thailand has access to insurance claims from the government's current agricultural insurance program. This data contains information on the size of the loss, the cause of the loss (e.g. drought, flood, pest), and the date of the loss. Additionally, the data has the geospatial location of 47% of all registered plots. The governemnt of Thailand requires that all farmers register their plots, and according to the researchers I spoke to, registration seems to be nearly universal. The last data source for this project is publicly available remote sensing data. There is data available on Normalized Difference Vegetation Index (NDVI), rainfall, soil moisture, and temperature. Overall, there are around 6 million observations per year in the data. Technically, the problem in this scenario is to use a time series of remote sensing data to predict the extent of the loss suffered by a farmer. I will evaluate the performance of different convolutional neural network architectures (CNNs) on this task. CNNs have been shown to achieve competitive performance on many time series classification tasks (Wang, Yan, and Oates (2017)). We will be testing the architectures proposed by Bai, Kolter, and Koltun (2018), Cui, W. Chen, and Y. Chen (2016), and Wang, Yan, and Oates (2017). The main objective of the trip would be to access the data and to gain a deeper understanding of the context, especially issues pertaining to the deployment of the model. I would be working closely with researchers at the Bank of Thailand, and I believe this will greatly inform how my research can be more useful in practice.