University of São Paulo
"Luiz de Queiroz" College of Agriculture

Weather index insurance design: a novel approach for crop insurance in Brazil

Daniel Lima Miquelluti

Thesis presented to obtain the degree of Doctor in Science. Area: Applied Economics

Piracicaba
2019

Daniel Lima Miquelluti
Agronomist

Weather index insurance design: a novel approach for crop insurance in Brazil

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **VITOR AUGUSTO OZAKI**

Thesis presented to obtain the degree of Doctor in Science. Area: Applied Economics

Piracicaba
2019

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Miquelluti, Daniel Lima

Weather index insurance design: a novel approach for crop insurance in Brazil / Daniel Lima Miquelluti. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2019.

53 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Seguro agrícola 2. Risco Sistêmico 3. GWQLASSO I. Título

ACKNOWLEDGEMENTS

To Prof. Dr. Vitor Augusto Ozaki, my advisor, for his support and attention during this work. His friendly, demanding and critical form, was a fundamental contribution in my growth as a researcher.

To my parents, David and Jucléia, for the support, advice and understanding along the way, as well as for the solid formation that was given to me.

To my girlfriend Simone for her support and understanding all along this path, I do not think it was easy to put up with me at the most tense moments.

Thank you to all the doctoral professors, friends and employees of the department of Applied Economics of ESALQ/USP, for the assistance and attention during this period.

Thanks to GESER and PhD colleagues.

CAPES, the Coordination for the Improvement of Higher Education Personnel, which granted me a scholarship during this PhD, financial assistance that contributed to the feasibility of this work.

Finally, I would like to express my sincere gratitude to all the people who, directly or indirectly, contributed to the accomplishment of this work.

TABLE OF CONTENTS

RESUMO

Design de seguro de índice climático: uma nova abordagem para o seguro agrícola no Brasil

O seguro agrícola é reconhecido como um dos mecanismos mais eficientes de proteção de renda na agricultura, transferindo o risco da fazenda para outros agentes e setores econômicos. O seguro tende a estimular o aumento da área cultivada e o uso de tecnologia, principalmente por atuar como garantia adicional de acesso ao crédito. No Brasil, no entanto, a massificação do seguro rural é limitada devido ao orçamento restrito para financiar o programa de subvenção governamental. Além disso, a falta de previsibilidade e garantia de recursos impede o planejamento de investimentos de longo prazo pelo setor privado, impõe custos aos beneficiários e gera insatisfação do público alvo. Esta tese visa contribuir para a expansão do seguro agrícola no Brasil por meio da pesquisa de seguro de índice climático, que possui menores custos administrativos e regulatórios quando comparado ao seguro tradicional. A ausência de validação de sinistro in loco e monitoramento de risco moral reduz os custos administrativos desse tipo de seguro, permitindo um seguro agrícola sem subsídio. No primeiro de dois artigos, exploramos a disponibilidade e a qualidade de bancos de dados públicos para produtividade de soja e precipitação diária no estado do Paraná, no Brasil, a fim de verificar a viabilidade de um produto de seguro de índice climático. Usamos a imputação múltipla por equações encadeadas (MICE) para preencher valores ausentes no conjunto de dados de precipitação e estudar a existência de padrões espaciais e temporais nos dados por meio de agrupamento hierárquico. Nossos resultados indicam que o Paraná preenche os requisitos de dados para um seguro de índice climático escalável com o uso do método MICE, e o agrupamento hierárquico é uma ferramenta eficaz no pré-processamento de dados. O segundo artigo estuda a eficiência de uma nova abordagem de regressão, a regressão quantílica LASSO ponderada geograficamente (GWQLASSO) na modelagem da relação entre o índice climático e a produtividade de soja. O GWQLASSO permite que os coeficientes de regressão variem espacialmente, enquanto utiliza a informação proveniente dos locais vizinhos de modo a obter estimativas robustas. O componente LASSO do modelo facilita a seleção de variáveis explicativas relevantes. Um produto de seguro de índice climático (WII) é desenvolvido com base em um índice de precipitação normalizado (intervalo de 1 mês) derivado de um conjunto de dados diários de precipitação para 41 estações meteorológicas (uma por município) no Estado do Paraná no período de 1979 a 2015. Os dados de rendimento da soja também são obtidos para estes 41 municípios de 1980 a 2015. A eficácia do produto GWQLASSO é avaliada em comparação com uma abordagem de regressão quantílica clássica e um produto tradicional de seguro de rendimento utilizando-se a medida de risco espectral (SRM) e o semi-desvio médio. Embora o GWQLASSO tenha se mostrado tão eficaz quanto a regressão quantílica, ele superou o produto de seguro de rendimento, provando assim ser uma alternativa ao mercado de seguro agrícola no Brasil e em outros locais com dados limitados.

Palavras-chave: Seguro agrícola; Risco sistêmico; GWQLASSO

ABSTRACT

Weather index insurance design: a novel approach for crop insurance in Brazil

Crop insurance is recognized as one of the most efficient mechanisms of income protection in agriculture, transferring risk from agriculture to other agents and economic sectors. Insurance tends to stimulate the increase of cultivated area and the use of technology, especially as it acts as an additional guarantee for access to credit. In Brazil, however, the massification of rural insurance is limited due to the restricted budget to fund government subsidization. Also, the lack of predictability and guarantee of resources prevents the long-term planning of investments by the private sector, imposes costs on the beneficiaries and generates dissatisfaction of the target public. This thesis aims to contribute to the expansion of crop insurance in Brazil through the research of index insurance, which has lower administrative and claim adjustment costs when compared to traditional insurance. The absence of in situ claim adjustment and moral hazard monitoring reduces the administrative costs of this type of insurance, permitting a subsidy free crop insurance. In the first of two articles, we explore the availability and quality of public databases for soybean yields and daily rainfall in the state of Paraná in Brazil in order to verify the feasibility of an index insurance product. We use multiple imputation by chained equations (MICE) to fill missing values in the rainfall dataset and study the existence of spatial and temporal patterns in the data by means of hierarchical clustering. Our results indicate that Paraná fulfills data requirements for a scalable weather index insurance with MICE and hierarchical clustering being effective tools in the pre-processing of data. The second article studies the efficiency of a novel regression approach, the geographically weighted quantile LASSO (GWQLASSO) in the modelling of yield-index relationship for weather index insurance products. GWQLASSO allows regression coefficients to vary spatially, while using the information from neighboring locations to derive robust estimates. The LASSO component of the model facilitates the selection of relevant explanatory variables. A weather index insurance (WII) product is developed based on 1-month SPI derived from a daily precipitation dataset for 41 weather stations in the State of Paraná (Brazil) for the period of 1979 through 2015. Soybean yield data are also used for the 41 municipalities from 1980 through 2015. The effectiveness of the GWQLASSO product is evaluated against a classic quantile regression approach and a traditional yield insurance product using the Spectral Risk Measure (SRM) and the Mean Semi-deviation. While GWQLASSO proved as effective as quantile regression it outperformed the yield insurance product, thus proving an alternative to the crop insurance market in Brazil and other locations with limited data.

Keywords: Crop insurance; Systemic risk; GWQLASSO

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

One of the flagships in the recent agricultural policy in Brazil, crop insurance has been advertised as one of the pillars of the 2016/2017 and 2017/2018 Agricultural and Livestock Plan (Ministério da Agricultura, Pecuária e Abastecimento, 2016; Ministério da Agricultura, Pecuária e Abastecimento, 2017). However, since its development in Brazil, this type of insurance has not achieved its intended endings with the protected area under 10% of the agricultural land (Ozaki, 2013). The low uptake is credited to the government insufficient investments in subsidies for the crop insurance program, however as noted by Oñate et al. (2016) one of the most subsidized crop insurance programs in Brazil, Proagro Mais, has failed to reduce uncertainty and risks. Also, as historic yields are not always available, insurers tend to use data provided by the Brazilian Institute of Geography and Statistics (IBGE), which are aggregated at the municipality level, thus pushing away high yield farmers and attracting the ones with low yields[1] (when compared to the munipality average yield).

Relying on subsidies to increase crop insurance uptake seems not to be a good alternative as tax payers' and several countries' perception of farm subsidies worsens (Edwards, 2018). The benefits of this type of subsidy have shown to favor only the ones receiving it and not the entire community (Drabenstott, 2015; Babcock, 2015; Kirwan & Roberts, 2016). Therefore, subsidy free alternatives should be sought in order to improve the financial security of farmers.

This does not mean the government should end all crop insurance programs, but improve their self-sustainability. In this sense, one promising product is parametric insurance, which has lower premium costs when compared to traditional insurance. The absence of in situ claim adjustment and moral hazard monitoring greatly reduces the administrative costs of this type of insurance, permitting a subsidy free crop insurance (Jensen & Barrett, 2017). Another advantage of index insurance products is the rapid and payment of indemnities, also due to the non-existence of local loss assessment.

The basis of index insurance development is systemic risk, one of the factors halting conventional crop insurance expansion. The correlation of losses among policyholders causes significant increase in the indemnities, renting conventional crop insurance infeasible in the long run. Given that crops are exposed to a series of widespread risks, such as drought, floods and windstorms it is clear that traditional crop insurance will not provide the adequate protection.

In Brazil, crop insurance uptake through the Crop Insurance Subsidy Program (PSR) is almost limited to four states, Rio Grande do Sul, Paraná, Santa Catarina e São Paulo, which respond to 83,6% of the insurance policies sold (21,4%, 40,4%, 6,5% and 15,3% respectively). States in the northeast, which are subject to continuous drought events, and thus increased systemic risk only correspond to 0,7% (Ministério da Agricultura, Pecuária e Abastecimento, 2019). These statistics, in conjunction with the low area coverage, show that crop insurance is confined to the southern portion of Brazil.

Aiming to contribute to the expansion of a sustainable crop insurance market in Brazil our main objective is to develop an weather index insurance (WII) design for the state of Paraná. This goal is divided in two essays, in the first one, entitled "Identifying potential regions for a precipitation index insurance product in Paraná – Brazil: a hierarchical clustering approach", we evaluate if the state attends the minumum data requirements for a WII and search for precipitation and risk spatial patterns. In the second essay, entitled "An application of geographically

[1] This problem is known in the insurance literature as adverse selection.

weighted quantile LASSO to weather index insurance design in Paraná – Brazil", we develop an WII design using a novel approach, the geographically weighted quantile least absolute shrinkage and selection operator (GWQLASSO), and compare it to a traditional quantile regression and a yield insurance product.

This thesis complements the existing literature not only by exploring methods to deal with missing data and detecting spatial patterns, but mainly through the use of a completely novel method to model the yield-index relationship. Also, we reinforce the effectiveness of parametric insurance in Paraná (Ozaki & Shirota, 2005), showing that a weather index insurance is preferable to a traditional yield insurance.

References

Babcock, Bruce A. "The concentration of US agricultural subsidies." *Iowa Ag Review* 7, no. 4 (2015): 4.

Drabenstott, Mark. "Do farm payments promote rural economic growth?" *Ag Decision Maker Newsletter* 9, no. 6 (2015): 2.

Edwards, Chris. "Agricultural subsidies." (2018).

Jensen, Nathaniel, and Christopher Barrett. "Agricultural index insurance for development." *Applied Economic Perspectives and Policy* 39, no. 2 (2017): 199-219.

Kirwan, Barrett E., and Michael J. Roberts. "Who really benefits from agricultural subsidies? Evidence from field-level data." *American Journal of Agricultural Economics* 98, no. 4 (2016): 1095-1113.

Ministério Da Agricultura Pecuária E Abastecimento (MAPA). "PLANO AGRÍCOLA E PECUÁRIO 2016/2017." MAPA Indicadores. Accessed August 07, 2018. http:// http://www.agricultura.gov.br/assuntos/sustentabilidade/plano-abc/arquivo-publicacoes-plano-abc/PAP1617.pdf.

———. "PLANO AGRÍCOLA E PECUÁRIO 2017/2018." MAPA Indicadores. Accessed August 07, 2018. http://www.agricultura.gov.br/assuntos/sustentabilidade/plano-agricola-e-pecuario/arquivos-pap/folder-pap-2017-18.

———. "Atlas do Seguro Rural" MAPA Indicadores. Accessed January 07, 2019. http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm.

Oñate, Carlos Andrés, Vitor Augusto Ozaki, and Boris Bravo-Ureta. "Impact Evaluation of the Brazilian crop insurance public program "Proagro Mais"." In *2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts*, no. 236096. Agricultural and Applied Economics Association, 2016.

Ozaki, Vitor Augusto, and Ricardo Shirota. "Um estudo da viabilidade de um programa de seguro agrícola baseado em um índice de produtividade regional em Castro (PR)." *Revista de Economia e Sociologia Rural* 43, no. 3 (2005): 485-503.

Ozaki, Vitor Augusto. "Qual o custo governamental do seguro agrícola?" *Revista de Economia e Sociologia Rural* 51, no. 1 (2013): 123-136.

2. IDENTIFYING POTENTIAL REGIONS FOR A PRECIPITATION INDEX INSURANCE PRODUCT IN PARANÁ – BRAZIL: A HIERARCHICAL CLUSTERING APPROACH

Daniel Lima Miquelluti
Departament of Economics, Management and Sociology
ESALQ-USP
danielmiq@usp.br
Vitor Ozaki
Departament of Economics, Management and Sociology
ESALQ-USP
vitorozaki@usp.br

Abstract

In this article we explore the availability and quality of public databases for soybean yields and daily rainfall in the state of Paraná in Brazil in order to verify the feasibility of an index insurance product. We use multiple imputation by chained equations (MICE) to fill missing values in the rainfall dataset and study the existence of spatial and temporal patterns in the data by means of hierarchical clustering. Our results indicate that Paraná fulfills data requirements for a scalable weather index insurance with MICE and hierarchical clustering being effective tools in the pre-processing of data.

Keywords: Index-insurance; Hierarchical clustering; MICE

2.1. Introduction

One of the key aspects of parametric insurance design is data, especially of high quality and from a sustainable source. In the context of index insurance, high quality means a long, consistent and unbiased historical record. However, as noted in Collier et al. (2009) the data needs for a weather index insurance (WII) depends on the characteristics of the weather event insured. In this aspect, the behavior of the risk spatially and temporally will reflect on the necessary spatial and temporal resolution of data.

Opposed to the traditional lines of insurance, parametric insurance relies on the spatial correlation of risks (systemic risks), so one of the first steps when designing this type of insurance is to determine the area affected by the event as this will indicate the necessary spatial resolution (Collier et al., 2010). Each event presents a spatial behavior, so the topography of the target region must be carefully studied, as a rough terrain alters weather patterns. Regions with a rough terrain generally indicate a low spatial correlation of weather events, thus, WII is not the best option of insurance in this case.

Spatial correlation is not the only form of correlation one must account when searching for data. Temporal correlation is also important, as weather events tend to follow a pattern in time. Such phenomena are observed in South America with the occurrence of El Niño and La Niña, or in Asia with the monsoons. Data must have the proper temporal resolution to capture these seasonal patterns.

Just as important as historical weather data are historical records of loss and their cause, which will provide information of the impacts of different levels of the weather risk thus enabling the determination of an index trigger. Ideally, when developing a WII, one should be able to estimate the probability distribution function and correlations (presumably high) of each of these variables.

A general benchmark for the minimum length of climatic data is 30 years (Collier et al., 2010). This value is used in many areas and represents the minimum number of observations necessary to estimate the central tendency and variance of a variable with acceptable accuracy.

When it comes to loss distribution, the researcher is interested in the conditional distribution of losses in respect to each level of the index. The loss is treated as a random variable because the weather event does not produce the same outcome to all, the individual loss is conditional on factors such as the distance from the weather station, business diversification and adoption of risk mitigating procedures. This variation of outcomes is reflected in the dispersion of the loss distribution and is called the basis risk. A central point in WII design is to reduce basis risk by the correct derivation of an index and cause of loss.

In Brazil, parametric insurance was introduced in 2017 by Swiss Re for a single large producer of corn, cotton and soybean in the states of Bahia, Mato Grosso and Minas Gerais. However, the literature in the subject is still inexistent, even the Brazilian literature in crop insurance is poor. This is due, in part, to the data scarcity which was mitigated in 2016 by the release of a Crop Insurance Atlas by the Brazilian Ministry of Agriculture, Livestock and Supply.

Therefore, aiming to contribute to the expansion of parametric insurance in Brazil, our objective is to assess if the state of Paraná is suitable for this type of product, regarding the data needs and existence of yield and rainfall spatial patterns. We focus on soybean as Paraná is the second largest producer in Brazil with a total of 19,073,706 tons produced in 2017, being also the second in average yields (3,663 kg/ha in 2017). Also, this state is where crop insurance has its higher penetration with 14,7% of the total crop area insured via the Rural Insurance Premium Subsidization Program (PSR), thus a natural choice to study the feasibility of novel crop insurance products (MAPA, 2017).

We first evaluate rainfall and soybean yield data length and quality, then use multiple imputation by chained equations (MICE) to fill missing values in the rainfall dataset and finally calculate the three-month standardized precipitation index (SPI) from the monthly totals. As for the yield data we first remove time trends, finally we use hierarchical clustering in order to identify spatial patterns.

2.2. Materials and Methods

Daily precipitation data in Brazil are available from the National Water Agency (ANA) and the National Institute of Meteorology (INMET), being that the former presents a more comprehensive distribution of weather stations in the state of Paraná. Therefore, we collected precipitation data only from ANA, spanning from 01/06/1973 through 31/12/2015 for a total of 1163 weather stations. This series was later aggregated in monthly totals.

Also, the series of annual soybean yields for each of the 399 municipalities in the state of Paraná, from 1980 through 2016, were obtained from the National Institute of Geography and Statistics (IBGE).

2.2.1. Data cleaning and yield detrending

From the initial set of 1163 weather stations and 399 municipalities we filtered the ones with 15% or less of missing data (Collier et al., 2009), resulting in 78 stations and 174 municipalities. Values of precipitation were capped at 150mm to account for measurement errors in the weather stations.

Crop yield data are subject to changes in practices and technology, which are not of interest for this study, therefore we detrended yields. A linear regression was adjusted to the yield data with time as the explanatory variable, then the last observed yield was corrected using the model residuals for each year (Gallagher, 1987; Duarte et al., 2018). The detrended yields are defined by the following equation:

$$\tilde{y_t} = \widehat{y_{2016}}\left(1 + \frac{\widehat{e_t}}{\widehat{y_t}}\right)$$

where $\tilde{y_t}, \widehat{y_t}$ *and* $\widehat{e_t}$ are, respectively, the corrected yield, the fitted yield and the residual for year t, $\widehat{y_{2016}}$ is the fitted yield for 2016.

2.2.2. Imputation for precipitation data

Given the existence of missing data we applied Multiple Imputation by Chained Equations (MICE), a method that combines imputation for multivariate data (Rubin, 1987) and Fully Conditional Specification, which was developed under several names, being chained equations the one implemented here using the R software (Van Buuren, 2000; 2010).

While multiple imputation considers a single imputation model for each variable with missing values, the chained equations technique permits the use of separate and univariate imputation models for each of these variables (Bartlett et al. 2014). In this way, hundreds of variables may be imputed with a high degree of flexibility (He et al. 2010). Continuous variables may be modeled through linear regression and binary variables through logistic regression for example (Chevret et al. 2015). However, MICE does not have the same theoretical basis as other methods such as multivariate normal imputation, what does not seem to be an issue (White et al. 2011).

A natural question when using imputation methods is whether the missing rate may be too high to use multiple imputation methods such as MICE. Research shows that these methodologies are unbiased when data is missing at no higher than 50%, being unstable for higher percentages, especially if the data distribution is asymmetrical (Lee & Carlin 2012; Haji-Maghsoudi et al. 2013). However, this does not imply that multiple imputation should be discarded as it exhibits superior performance to other methods even for a 75% data loss, despite biased estimates (Marshall et al. 2010).

For a partially observed random sample of the multivariate distribution $\mathrm{P(Y|\theta)}$, completely specified by the vector of k unknown parameters θ and representing the complete data Y, the posterior distribution of θ and then the predictive distribution of Y are obtained through a Gibbs sampler of the form:

$$\theta_1^{\star(t)} \sim \mathrm{P}\left(\theta_1 \middle| Y_1^{(obs)}, Y_2^{(t-1)}, \dots, Y_k^{(t-1)}\right)$$
$$Y_1^{\star(t)} \sim P\left(Y_1 \middle| Y_1^{(obs)}, Y_2^{(t-1)}, \dots, Y_k^{(t-1)}, \theta_1^{\star(t)}\right)$$
$$\vdots$$
$$\theta_k^{\star(t)} \sim \mathrm{P}\left(\theta_k \middle| Y_k^{(obs)}, Y_1^{(t-1)}, \dots, Y_{k-1}^{(t-1)}\right)$$
$$Y_k^{\star(t)} \sim P\left(Y_k \middle| Y_k^{(obs)}, Y_1^{(t-1)}, \dots, Y_{k-1}^{(t-1)}, \theta_k^{\star(t)}\right)$$

where $Y_j^t = (Y_j^{(obs)}, Y_j^{*(t)})$ is the jth imputed variable at iteration t and $Y^{(obs)}$ is the portion of Y that is observed.

We chose predictive mean matching (PMM) as the imputation method within MICE given that precipitation is generally skewed, thus not normally distributed. Nevertheless, simulations have shown that normal imputation models do work with non-normal data (Graham & Schafer 1999). Imputations made through PMM better resemble the observed values than methods based on the normal distribution (White et al. 2011). This follows from the way PMM work as it uses the predicted value for a given missing value to identify similar observations. These identified observations are used to create a matching set is containing *q* matches, from which PMM then draws a random observation. Therefore, PMM uses the real observed values to fill the missing data and thus preventing extrapolation beyond the range of the data (Little 1988).

As covariates we chose latitude, longitude, with month and year binaries to capture seasonal changes. With this specification, the MICE procedure assumes Y being normally distributed[2] and estimates a linear multiple regression. This yields a $\widehat{\boldsymbol{\beta}}$ vector of parameters (of length *k*), with an estimated covariance matrix $\boldsymbol{V}$ and root mean-squared error $\widehat{\sigma}$, from fitting this model to $Y^{(obs)}$.

The next step is to draw the imputation parameters $\sigma^{\star}$, $\boldsymbol{\beta}^{\star}$ from the exact joint posterior distribution of σ, $\boldsymbol{\beta}$. The parameter $\sigma^{\star}$is drawn as $\sigma^{\star} = \widehat{\sigma}\sqrt{(n_{\mathrm{obs}} - k)/g}$, where n_{obs} if the number of observed values, *g* is a random draw from a χ^2 distribution with $n_{obs} - k$ degrees of freedom. Then, $\boldsymbol{\beta}^{\star}$ is drawn as $\boldsymbol{\beta}^{\star} = \widehat{\boldsymbol{\beta}} + \frac{\sigma^{\star}}{\widehat{\sigma}}\boldsymbol{u}_1\mathbf{V}^{1/2}$, where $\boldsymbol{u}_1$ is a vector of *k* independent random draws from a standard Normal distribution and $\mathbf{V}^{1/2}$ is the Cholesky decomposition of $\mathbf{V}$.

For each missing value Y_i with covariates $\boldsymbol{X}_i$ PMM identifies the *q* individuals with the smallest values of $\left|\widehat{\boldsymbol{\beta}}\boldsymbol{X}_o - \boldsymbol{\beta}^{\star}\boldsymbol{X}_i\right|$ $(o = 1, \ldots, n_{obs})$. Of these *q* closest individuals, one is chosen at random ($Y_{h'}$), and the imputed value of Y_i is $Y_{h'}$. Thus, the imputed value is an observed value of *Y* whose prediction is closely matched by the perturbed prediction.

The size of the matching set is chosen by the researcher with values like $q = 1$ in leading to estimated standard errors that are too low and t-statistics that are too large (Morris et al. 2014). Whereas values ranging from $q = 3$ over $q = 10$ showed a small advantage (Schenker & Taylor 1996; Morris et al. 2014). The size of the matching set is dependent on sample size and may have poor performance in small samples as the difference between similar observations is increased.

PMM has shown similar performance to correctly specified parametric models and better than poorly specified ones characterized by non-normality (Morris et al. 2014; Schenker & Taylor 1996) and skewness (Marshall et al. 2010) considering that the method does not have a strong theoretical backing (Kenward & Carpenter 2007).

Finally, for this analysis the number of repeated imputations was *m*=5, *q*=5 and the number of iterations was *t*=20. The quality of the imputations was checked using the Kolmogorov-Smirnov test in order to check departures from the original distribution of data (Raghunathan & Bondarenko, 2007).

The application of MICE has been successful in several areas, including precipitation data imputation in Brazil by de Carvalho et al (2017).

2 This assumption does not affect the quality of the imputations as this regression is simply a metric for matching (Little, 1988).

2.2.3. Clustering procedures

Prior to the application of hierarchical clustering, precipitation data was aggregated monthly and we calculated the standardized precipitation index (SPI) with a three-month scale, thus capturing drought events during the crop season (Mckee et al, 1993). We chose the Ward's clustering method with an Euclidean distance matrix since it has already proved successful in defining homogenous precipitation regions in Brazil (Keller Filho, 2005). The optimal number of clusters was obtained through the majority rule of 30 indices, an algorithm implemented in Charrad et al (2014).

2.3. Results and Discussion

2.3.1. Weather station and yield data spatial distribution

Observing the spatial distribution of weather stations with less than 15% of missing data and plotting a 50km halo there is only a portion of the state without coverage, mainly around the city of Londrina (Figure 1). This result indicates an aptitude for parametric insurance at the meso and/or macro scales, targeted to cooperatives and other larger risk aggregators (Collier, 2010). At larger scales, weather index insurance permits the identification of large events and decreases the impact of basis risk. Microinsurance is possible for the municipalities with a weather station (78) and surrounding locations up to 15 km, however this greatly reduces the scalability of WII in Paraná.

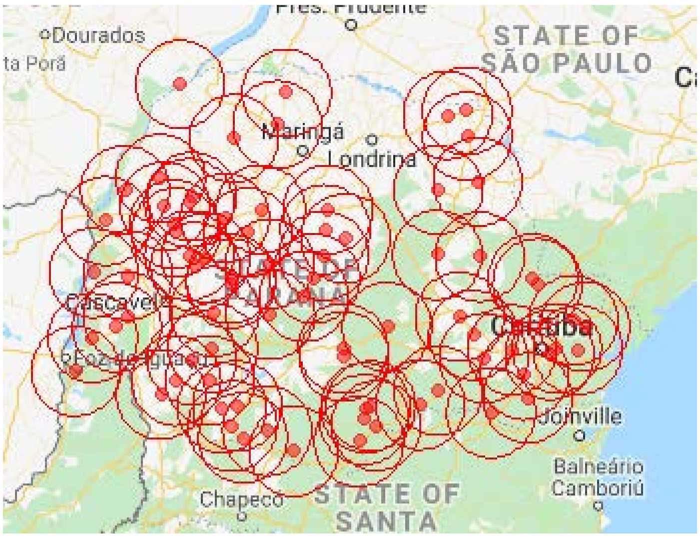

Figure 1. Weather stations spatial distribution (with a 50km halo)

The yield data coverage is more disperse with some gaps, especially in the northwest and east portion of the state (Figure 2). In the northwest this lack of data reflects the characteristics of the region, with sandy soils and warm climate, being thus restrictive to the growth of soybean. Another reason for the low presence of soybean is the predominance of ranching in this region (Franchini et al, 2016). This author also notes that in the east the presence of soybean is limited. Nevertheless, the available data represents the bulk of soybean producers in the state with approximately 70% of the state total production in 2016 (IBGE, 2018).

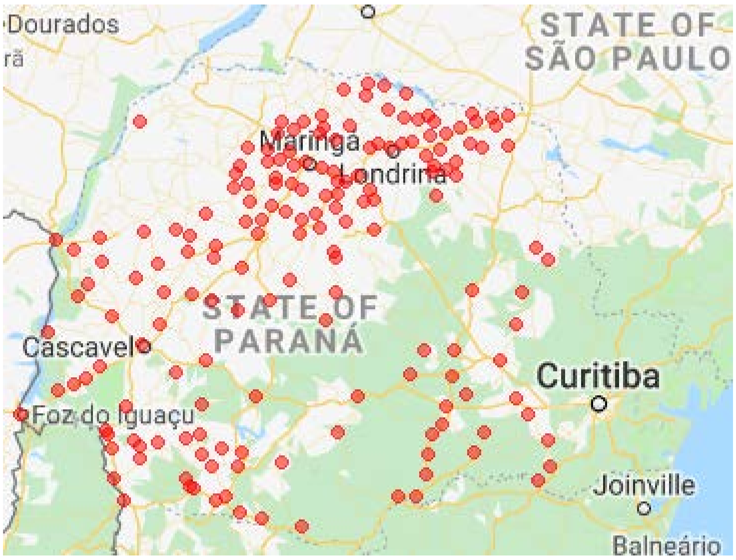

Figure 2. Yield data spatial distribution

2.3.2. Imputations

According to the Kolmogorov-Smirnov test the distribution of the imputed precipitation does not differ from the original dataset (D = 0.005964, p-value = 0.1016), therefore the procedure did not alter the underlying structure of the data. This result reinforces the use of MICE as a valid imputation procedure for precipitation data in Brazil (de Carvalho et al, 2017). It must be noted that albeit its effectiveness, MICE should be used with caution in datasets with 50% or more of missing values. Also, the specification of the correct imputation model and quality of predictors plays a large role in the quality of the imputations (White et al, 2011).

2.3.3. Precipitation clusters

According to the majority rule, the optimal number of clusters was two, with nine votes, followed by three clusters with six votes (Table 1). Given that we used a different approach to the clustering methodology than Keller Filho et al (2005), whereas he used several statistical parameters calculated from five-day accumulated precipitation we only used the three-month SPI, our results do not completely match but are very similar regarding the characteristics of the clusters. Cluster 1 represents areas in the west, center and north of the state, with higher total precipitation in the year aggregate but greater variability among years. Whereas cluster 2 represents the center and east of Paraná, with a lower total precipitation but with less variability (Figure 3).

Table 1. Ideal number of clusters for each variable

Variable		Ideal number of clusters*								
		0	1	2	3	5	6	7	8	10
Votes	Rainfall	2	1	9	6	0	1	1	2	1
	Yield	2	0	11	8	2	0	0	1	2

Source: Authors
*Only numbers with at least one vote are presented

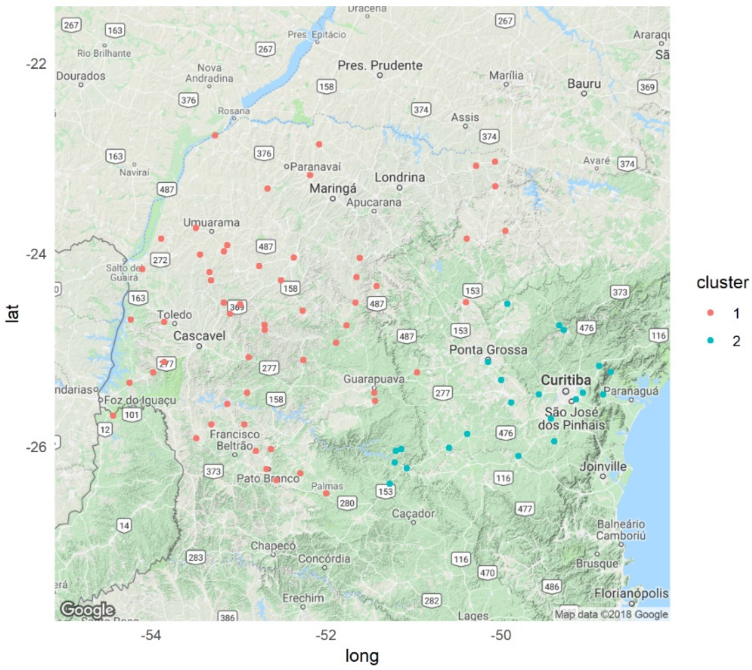

Figure 3. Precipitation clusters spatial distribution

Regarding the SPI values for each month and cluster, it is interesting to observe that there is little difference in the median of monthly SPI, albeit statistically significant according to our cluster analysis (Figure 4). When carefully analyzed, it can be observed that cluster 1 has a greater number of observations in the lower ranges of SPI, indicating the occurrence of moderate and severe droughts. This can be explained by the greater variability in precipitation, and the occurrence of droughts in the north and northeast of the state as identified by Fritzsons et al (2011).

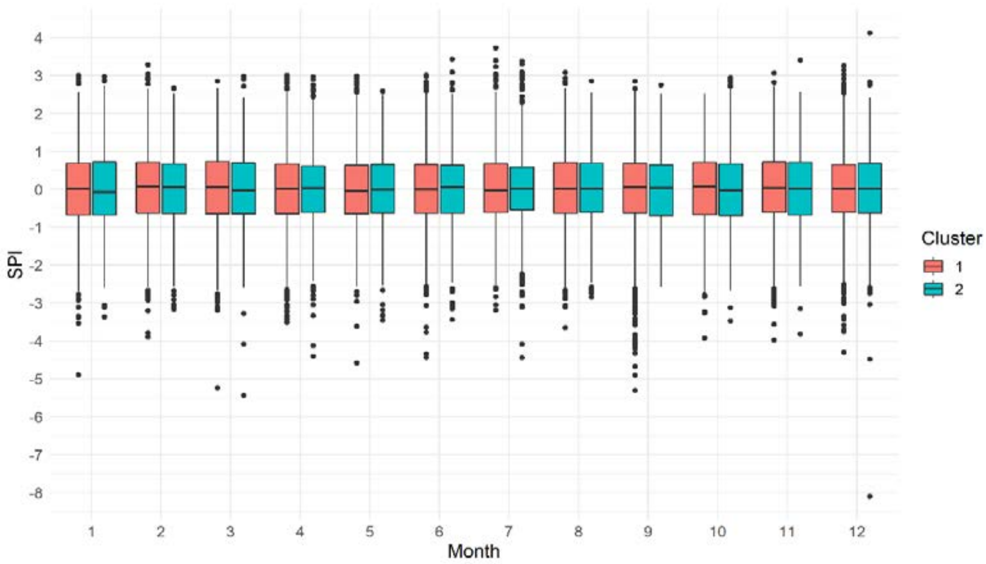

Figure 4. Monthly SPI boxplot, per cluster.

When analyzing only the period in which soybeans are grown in the state, October through March, cluster 1 presents variable conditions, as there is a surplus in precipitation during the growth and reproductive stages with a decrease in precipitation in the end of the growth period (Figure 5). However, there must be caution with the occasional occurrence of drought, which can be mitigated using irrigation or risk management products such as crop insurance. Despite the decrease in precipitation from January through April/May, the total precipitation in this period is sufficient for cultivars ranging from 450 to 700 mm of water requirements.

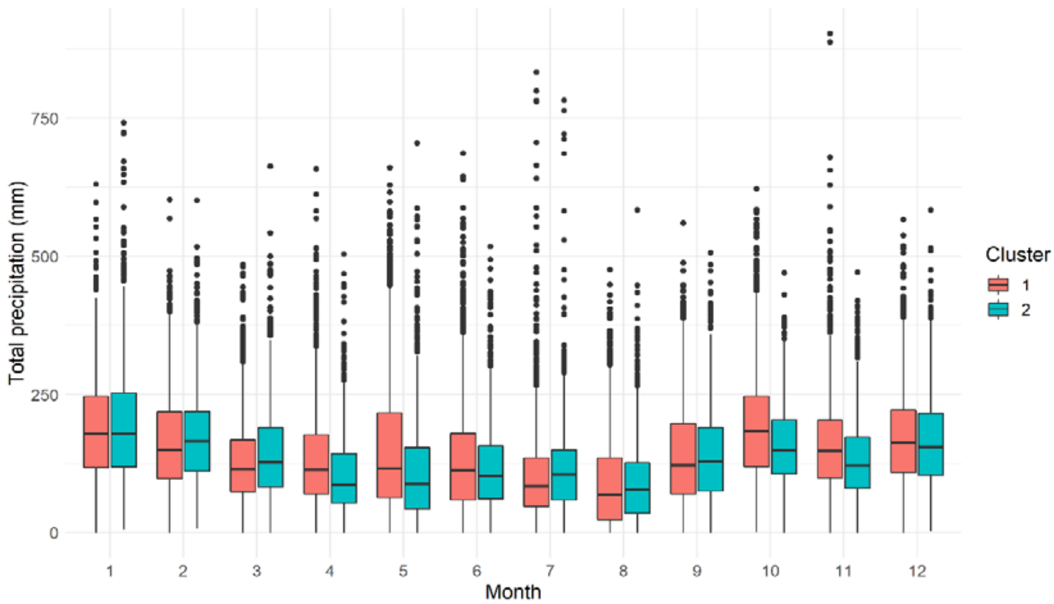

Figure 5. Monthly precipitation (total) boxplot, per cluster.

For Cluster 2, the opposite is observed with lower levels of precipitation from October through December and higher levels in January and February. However, in these areas, there is a steeper descent in precipitation levels, being the region adequate for cultivars requiring from 450 to 650 mm of water. It must be noted that areas represented in cluster 2 have a lower variability, thus, it suffers less from drought and excessive rain periods (Figures 5 and 6).

2.3.4. Yield clusters

According to the majority rule, the optimal number of clusters was two, with eleven votes, followed by three clusters with eight votes (Table 1). Similar to the precipitation clusters we have cluster 1 representing the west and northwest of the state while cluster 2 comprehends the south, center and east of Paraná (Figure 6). Thus, the only difference from the rainfall clusters is that the yield cluster 1 has less presence in the center and south of Paraná.

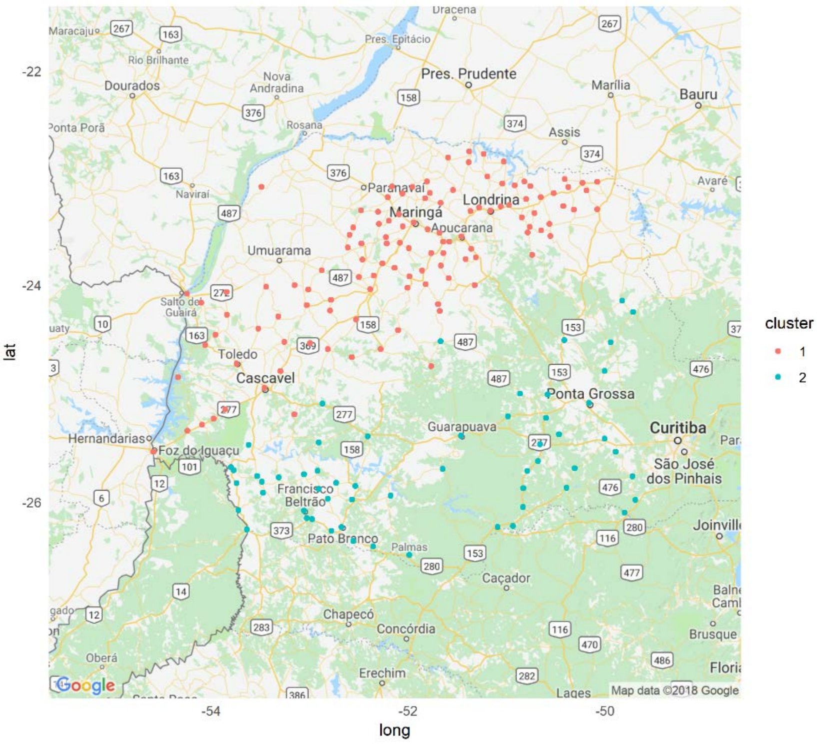

Figure 6. Yield clusters spatial distribution.

Both clusters present a similar yield level from the beginning of the series through 1990 and from 2001 onwards, however, in the period comprised between 1991 and 2000 cluster 1 has lower yields (Figure 7). Also, in years where losses occurred (1986, 1991, 1992, 2005, 2009, 2012), cluster 1 municipalities suffered greater losses, increasing cluster variability and decreasing the mean and median of the whole period (Table 2). The latter can be explained by the presence of municipalities in the northern portion of the state in cluster 1, as said in the previous section this region has sandy soils and higher temperatures, being more susceptible to drought (Franchini, 2016). Other researchers such as Pavan (2013) and Felema et al. (2016) also study the spatial behavior of soybean yields in Paraná. However, while our results agree to some measure, we refrain from making comparisons as both studies used only two years of data against our 37-year time series.

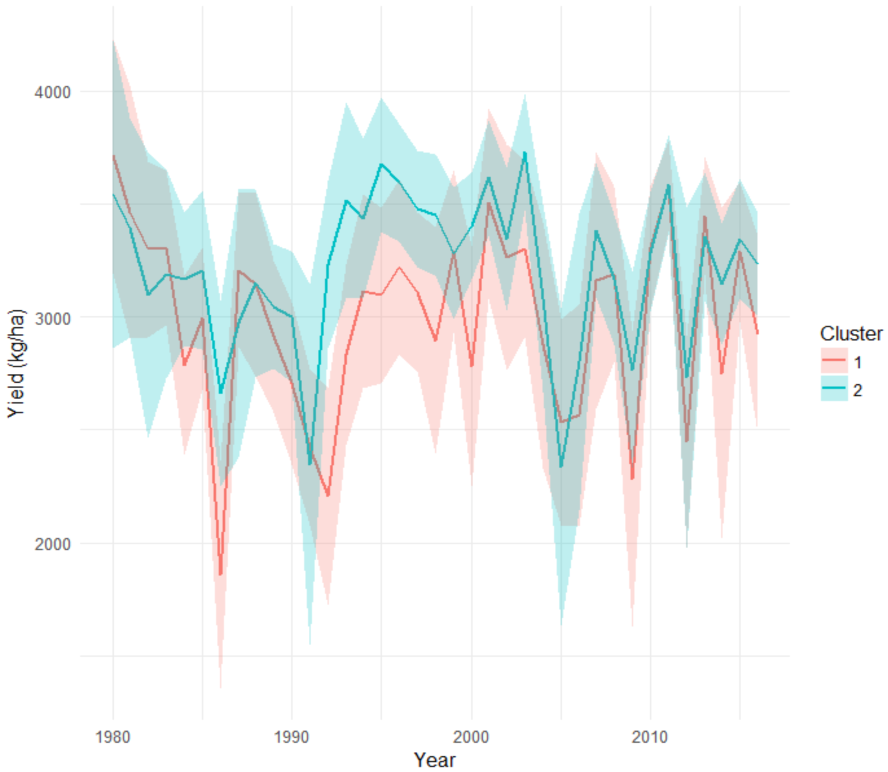

Figure 7. Soybean yields (kg/ha) time series with 95% confidence intervals, per cluster.

Table 2. Yield clusters descriptive statistics.

Cluster	Mean	Median	Standard deviation	Coefficient of variation (%)
1	2993,46	3051,21	597,05	19,95
2	3209,30	3282,11	533,26	16,62

Source: Authors

2.3.5. Yield and precipitation clusters relationship

When comparing with the results found for the precipitation clusters, the need to consider other environmental variables is exemplified. Regardless of the precipitation cluster 2 having lower precipitation levels, other factors such as soil type and temperature lead to greater yields in this region. The southwest of Paraná is the only region with high precipitation and high yields. Intersecting the clusters would lead to a further separation, with three separate regions, the southeast with lower precipitation levels and high yields, the west and center with good precipitation levels (but with higher variability) and lower yields and the southwest as described above. These "new clusters" could present separate regions for the design of a weather index insurance products, with each region having a fine-tuned product.

This analysis does not encompass soil and other weather variables, which are also important in the determination of the suitability of cultivars for each region. The northwest of Paraná presents sandy soils and higher temperatures, therefore, farms in this region suffer more from drought periods as these soils have a lower water holding capacity and the increase in temperature leads to a higher evapotranspiration. On the contrary, for the south portion of the state, soils are rich in clay, altitudes are higher and temperatures lower, this coupled with a low variability in precipitation results in a lower risk of drought related yield losses (Lima et al., 2012; Franchini et al., 2016). Consequently, when choosing adequate risk management strategies and in the design of crop insurance products, such as weather index insurance, these variables must be taken in account.

2.4. Conclusion

Verifying the availability and quality of data sources is one of the first steps when designing a weather index insurance product. This step is particularly difficult in large developing countries such as Brazil, where the weather agencies do not have the necessary funds to maintain a large net of weather stations. Given this lack of resources, the existing stations also suffer from missing data, a problem that generally implies in pricier insurance. In this paper we evaluate the quality of precipitation and yield data in Paraná-Brazil and present a proven method to deal with missing data.

Despite the variability of soil and temperature conditions we find that the state of Paraná presents a great opportunity for index insurance based on precipitation data. There is a good coverage of suitable weather stations and the clusters found indicate the scalability of WII and the existence of spatially correlated weather events. The sharp decrease in weather stations from the original set to the filtered one is due to the lack of historical data in many of the stations, as the number of operational stations is around 900, thus the weather station coverage should improve with time.

We also find that MICE proved a reliable method to fill gaps in precipitation data with up to 15% of missing observations, therefore it should be considered by insurers as an alternative to the practice of loading insurance premium in cases where data is not complete. This would provide a more attractive product without losing precision in the pure risk estimates as the method does not change the probability distribution of data.

Our article presents a beginning of the exploration of weather index insurance design in the Brazilian literature, as we did not verify the economic viability of index insurance, focusing only in the technical aspects required for its operation. Thus, additional studies are required to determine if WII is a viable option to the crop insurance market in Paraná and how it compares to existing crop insurance products.

References

Bartlett, Jonathan W., Shaun R. Seaman, Ian R. White, James R. Carpenter, and Alzheimer's Disease Neuroimaging Initiative*. "Multiple imputation of covariates by fully conditional specification: accommodating the substantive model." *Statistical methods in medical research* 24, no. 4 (2015): 462-487.

Charrad, Malika, et al. "Package 'NbClust'." *Journal of Statistical Software* 61 (2014): 1-36.

Chevret, S., Shaun Seaman, and M. Resche-Rigon. "Multiple imputation: a mature approach to dealing with missing data." *Intensive care medicine* 41, no. 2 (2015): 348-350.

Collier, B., J. Skees, e B. Barnett, 2009 Weather index insurance and climate change: opportunities and challenges in lower income countries. The Geneva Papers on Risk and Insurance Issues and Practice **34**: 401–424.

Collier, B., B. Barnett, e J. Skees. "State of knowledge report–data requirements for the design of weather index insurance." (2010).

de Carvalho, José Ruy Porto, et al. "Model for Multiple Imputation to Estimate Daily Rainfall Data and Filling of Faults." *Revista Brasileira de Meteorologia* 32.4 (2017): 575-583.

Duarte, Gislaine V., Altemir Braga, Daniel L. Miquelluti, and Vitor A. Ozaki. "Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance." *Journal of Applied Statistics* 45, no. 11 (2018): 1920-1937.

Felema, João, Edson Ramos de Medeiros, Carlos Roberto Ferreira, Marcia Regina Gabardo da Camara, and Sidnei Pereira do Nascimento. "Um estudo da produtividade do feijão, do milho e da soja na agricultura paranaense, nos anos de 2000 e 2010: uma análise espacial." *Ensaios FEE* 36, no. 4 (2016): 817-842.

Franchini, Julio Cezar, Alvadi Antonio Balbinot, Jr., Pablo Ricardo Nitsche, Henrique Debiasi, and Ivani De Oliveira Negrão Lopes. "Variabilidade Espacial E Temporal Da Produção De Soja No Paraná E Definição De Ambientes De Produção." (2016). Accessed September 22, 2018. https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1052786

Fritzsons, Elenice, et al. "Análise da pluviometria para definição de zonas homogêneas no Estado do Paraná." *Raega-O Espaço Geográfico em Análise* 23 (2011).

Gallagher, Paul. "US soybean yields: Estimation and forecasting with nonsymmetric disturbances." *American Journal of Agricultural Economics* 69.4 (1987): 796-803.

Graham, John W., and Joseph L. Schafer. "On the performance of multiple imputation for multivariate data with small sample size." *Statistical strategies for small sample research* 50 (1999): 1-27.

Haji-Maghsoudi, Saiedeh, Ali-akbar Haghdoost, Azam Rastegari, and Mohammad Reza Baneshi. "Influence of pattern of missing data on performance of imputation methods: an example using national data on drug injection in prisons." *International journal of health policy and management* 1, no. 1 (2013): 69.

He, Yulei, Alan M. Zaslavsky, M. B. Landrum, D. P. Harrington, and P. Catalano. "Multiple imputation in a large-scale complex survey: a practical guide." *Statistical methods in medical research* 19, no. 6 (2010): 653-670.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA (IBGE). Sistema IBGE de Recuperação Automática. 2018. Available in: https://sidra.ibge.gov.br/home/ipp/brasil. Accessed: dec. 2018.

Keller Filho, T.; Lima, P.R.S.R.; Assad, E.D. Regiões pluviometricamente homogêneas no Brasil. Pesq. Agropec. Bras., v. 40, n. 4, p. 311-322, 2005.

Kenward, Michael G., and James Carpenter. "Multiple imputation: current perspectives." *Statistical methods in medical research* 16, no. 3 (2007): 199-218.

Lee, Katherine J., and John B. Carlin. "Recovery of information from multiple imputation: a simulation study." *Emerging themes in epidemiology* 9, no. 1 (2012): 3.

Lima, V. C., M. R. Lima, and V. F. Melo. "Conhecendo os principais solos do Paraná: abordagem para professores do ensino fundamental e médio." *Sociedade Brasileira de Ciência do Solo (Eds.). Núcleo Estadual Paraná, Brasil* (2012): 18.

Little, Roderick JA. "Missing-data adjustments in large surveys." *Journal of Business & Economic Statistics* 6, no. 3 (1988): 287-296.

Marshall, Andrea, Douglas G. Altman, and Roger L. Holder. "Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study." *BMC medical research methodology* 10, no. 1 (2010): 112.

McKee, Thomas B., Nolan J. Doesken, and John Kleist. "The relationship of drought frequency and duration to time scales." *Proceedings of the 8th Conference on Applied Climatology*. Vol. 17. No. 22. Boston, MA: American Meteorological Society, 1993.

McKee, Thomas B. "Drought monitoring with multiple time scales." *Proceedings of 9th Conference on Applied Climatology, Boston, 1995*. 1995.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. Departamento De Gestão De Riscos. *Relatório Geral 2017 – Programa de Subvenção ao Prêmio do Seguro Rural (PSR)*. 2017.

Morris, Tim P., Ian R. White, and Patrick Royston. "Tuning multiple imputation by predictive mean matching and local residual draws." *BMC medical research methodology* 14, no. 1 (2014): 75.

Pavan, Lucca Simeoni. "Os determinantes da produtividade agrícola dos municípios paranaenses: uma análise de dados espaciais." Master's thesis, Universidade Estadual de Maringá, 2013.

Raghunathan, Trivellore E., et al. "A multivariate technique for multiply imputing missing values using a sequence of regression models." *Survey methodology* 27.1 (2001): 85-96.

Raghunathan, Trivellore, and Irina Bondarenko. "Diagnostics for multiple imputations." (2007).

Rubin, Donald B. *Multiple imputation for nonresponse in surveys*. No. 04; HA31. 2, R8. 1987.

Schenker, Nathaniel, and Jeremy MG Taylor. "Partially parametric techniques for multiple imputation." *Computational statistics & data analysis* 22, no. 4 (1996): 425-446.

Van Buuren, Stef, Hendriek C. Boshuizen, and Dick L. Knook. "Multiple imputation of missing blood pressure covariates in survival analysis." *Statistics in medicine* 18.6 (1999): 681-694.

Van Buuren, Stef. *Multivariate imputation by chained equations: MICE V1. 0 user's manual*. Leiden: TNO, 2000.

Van Buuren, Stef. "Multiple imputation of discrete and continuous data by fully conditional specification." *Statistical methods in medical research* 16, no. 3 (2007): 219-242.

Van Buuren, Stef, and Karin Groothuis-Oudshoorn. "mice: Multivariate imputation by chained equations in R." *Journal of statistical software* (2010): 1-68.

White, Ian R., Patrick Royston, and Angela M. Wood. "Multiple imputation using chained equations: issues and guidance for practice." *Statistics in medicine* 30, no. 4 (2011): 377-399.

3. AN APPLICATION OF GEOGRAPHICALLY WEIGHTED QUANTILE LASSO TO WEATHER INDEX INSURANCE DESIGN IN PARANÁ – BRAZIL

Daniel Miquelluti
Departament of Economics, Management and Sociology
ESALQ-USP
danielmiq@usp.br
Vitor Ozaki
Departament of Economics, Management and Sociology
ESALQ-USP
vitorozaki@usp.br

Abstract

This article studies the efficiency of a novel regression approach, the geographically weighted quantile LASSO (GWQLASSO) in the modelling of yield-index relationship for weather index insurance products. GWQLASSO allows regression coefficients to vary spatially, while using the information from neighboring locations to derive robust estimates. The LASSO component of the model facilitates the selection of relevant explanatory variables. A weather index insurance (WII) product is developed based on 1-month SPI derived from a daily precipitation dataset for 41 weather stations in the State of Paraná (Brazil) for the period of 1979 through 2015. Soybean yield data are also used for the 41 municipalities from 1980 through 2015. The effectiveness of the GWQLASSO product is evaluated against a classic quantile regression approach and a traditional yield insurance product using the Spectral Risk Measure (SRM) and the Mean Semi-deviation. While GWQLASSO proved as effective as quantile regression it outperformed the yield insurance product, thus proving an alternative to the crop insurance market in Brazil and other locations with limited data.

Keywords: GWQLASSO; Index-insurance; Systemic risk

3.1. Introduction

The unpredictability of climatic variations is the principal risk factor in soybean cultivation on the south of Brazil. Reports on indemnities paid by government risk management programs, the Program for the Guarantee of Agricultural and Livestock Activity (Proagro)[3] and Rural Insurance Premium Subsidization Program (PSR) (MAPA, 2015; BACEN, 2018), shows that the occurrence of droughts are the main event of loss (85% of the insured sum), followed by excessive rain (7.6% of the insured sum) and hail (4.2% of the insured sum). In addition, losses due to strong wind, excessive temperature fluctuation and flood are also mentioned.

Crop insurance is recognized as one of the most efficient mechanisms of income protection in agriculture, transferring risk from agriculture to other agents and economic sectors. Insurance tends to stimulate the increase of cultivated area and the use of technology, especially as it acts as an additional guarantee for access to credit (Goodwin et al, 2004). In this sense, it not only contributes to the achievement of lower interest rates (Cai, 2016) by

3 Created with the objective of exempting the rural producer from the fulfillment of financial obligations in rural credit operations in case of income losses motivated by climatic adversities.

the rural producer, since the reduction of agricultural risk translates into lower credit risk, but also contributes to the development of financial, insurance and capital markets. As a result, it minimizes the pressure for subsidized credit and ex-post government financial bailout, reducing the recurring pressure for renegotiations of rural debts.

However, the degree of penetration of agricultural insurance, considering the size and relevance of Brazilian agribusiness, is still insignificant. One of the reasons for the restriction of the subsidized crop insurance program and the massification of rural insurance in the country is the limited availability of budgetary resources to fund the policies. Also, the lack of predictability and guarantee of resources prevents the long-term planning of investments by the private sector, imposes costs on the beneficiaries and generates dissatisfaction of the target public (MAPA, 2017).

The Proagro risk management program also faces difficulties, according to Oñate et al. (2016) there was no increase in welfare for participating farmers. Considering the fact that the pricing of Proagro does not take into account regional differences, only crop type and cultural management practices such as the use of irrigation (BACEN, 2018), we believe different approaches must be sought by the government.

A possible alternative to overcome these issues is parametric insurance, which has lower administrative and regulation costs when compared to traditional insurance. The absence of in situ claim adjustment and moral hazard monitoring greatly reduces the administrative costs of this type of insurance, permitting a subsidy free crop insurance (Jensen e Barrett, 2017). Another advantage of parametric insurance is the rapid payment of indemnities.

Parametric insurance first appeared in the pioneering written by Chakravarti (1920). After more than a decade studying the subject, the author developed an insurance product based on rainfall levels for Chitradurga in India. Indemnities were paid if total rainfall measures in the beginning of the agricultural year were 35% below normal. The payouts were divided in two periods, from January through July and from January through October, according to the production cycle. The author noted that the area should be as uniform as possible, in respect to rainfall, for the insurance to work properly. The premiums were calculated to be as close as possible to land tax value, with both premiums and indemnities depending on the land's quality. In order to keep the farmers enrolled in the insurance scheme, contracts would be ranging from 5 to 10 years, so that each farmer would receive at least one indemnity and thus perceive the value of crop insurance (Mishra, 1995; Rao, 2011).

Halcrow (1949) devised a different form of index insurance, based in the area-yields. The main idea was to develop an insurance product where indemnities would be due when the mean-yield of a uniform area fell below a pre-defined level (which could be defined as a proportion of the expected mean-yield). The size of the area could vary as long as the homogeneity of yields was maintained, and the insured farmer would select a percentage of the expected yield for the area.

The main advantage of this type of insurance over the traditional crop-insurance products is the reduction of moral hazard[4]. Since the farmer could not significantly alter the area-yield, risk increasing measures are not economically viable. This would also lead to a reduction in deductibles and coverage levels limitation by the insurers (Miranda, 1991). This author also notes that adverse selection[5], which is caused by information asymmetry, is reduced in area-yield insurance as this information is available to the general public. Adding to the advantages of this

4 When the insured incur in risk increasing activities or stop taking risk-mitigating actions.

5 The inability to correctly measure farmer risk lead insurers to price the insurance incorrectly and in consequence to a greater proportion of high-risk farmers in their portfolio. This will ultimately lead to a market collapse.

type of insurance are the reduced administrative costs since an index-based insurance does not require individual assessment of yields, a major cost for traditional crop-insurance.

Two years after the work published by Miranda (1991) a yield-based index-insurance was developed by the Federal Crop Insurance Corporation (FCIC) in conjunction with Skees et al. (1997). The product named Group Risk Plan (GRP) was expanded in 1994 and reached 70% of market participation in 1997, considering the seven major crops and excluding forage. An additional feature of GRP was the possibility to scale the protection (the product of expected yield and expected price) up to 150%. This option was intended to increase protection since farm and county yields are not perfectly correlated. The difference between the county-yields, the index, and the value of individual yields, is called basis-risk, a problem that is always present in index-based insurance. In this way, GRP was design to reduce basis-risk by using double exponential smoothing to forecast the central tendency of yields, scaling the protection and paying indemnities based on the percentage reduction of yields rather than the weight/volume reduction. Since yield data provided by the National Agricultural Statistics Service (NASS) are available only at the county level, it wasn't possible to change the area in order to increase homogeneity of yields.

The GRP insurance was later expanded in 1999 to cover price variations and the index turned into a revenue index, named the Group Risk Income Protection (GRIP). The expected price was calculated individually for each crop and region. Both GRP and GRIP were replaced by the Area Risk Protection Insurance Policy (ARPI) in 2013. This new policy is formed of three insurance plans, Area Revenue Protection (ARP), Area Revenue Protection with Harvest Price Exclusion (ARPwHPE) and Area Yield Protection (AYP). The ARP and ARPwHPE are similar to the GRIP and the AYP is similar to the GRP, with the harvest price exclusion option meaning the amount protected will not rise if harvest prices rise (Schnitkey, 2014).

Weather based index products were to be operated only from 2006 with the approval of flood insurance by the Peruvian government (Khalil et al., 2007). Following that, several studies and pilots were launched, mostly in developing countries (Skees et al., 2001, 2007; Giné et al., 2010; Leblois et al., 2014; Maestro et al., 2016).

Parametric insurance in Brazil is quite limited, with only one insurer offering tailored weather index insurance products (Swiss Re) as of 2018. Past initiatives include a yield index product, commercialized by AgroBrasil (Carter et al., 2015) in the state of Rio Grande do Sul and a hypothetical yield index insurance for Castro in Paraná (Ozaki, 2005).

Therefore, aiming to contribute for the expansion of parametric insurance in Brazil, we intend to assess if the Paraná state presents a suitable environment for this type of product. This study specifically targets soybean in Paraná, the second largest soybean producer in Brazil with a total of 19,073,706 tons produced in 2017, being also the second in average yields (3,663 kg/ha in 2017). We develop a weather index product based on the Standardized Precipitation Index (SPI) and analyze its hedging effectiveness against a common yield insurance.

We also extend the work of Conradt (2015), who proposed the use of quantile regression to model the yield-index relationship, by applying the Geographically Weighted Quantile least absolute shrinkage and selection operator (GWQLASSO) (Wang, 2018) framework. Our hypothesis is that the spatial component, captured by the latter, plays an important role in the determination of the yield-index relationship. Also, this methodology is less data intensive, as it borrows information from neighboring locations. The effectiveness of our model is compared to the traditional yield insurance and the quantile regression approach by means of two risk measures, the Spectral Risk Measure (SMR) and the Mean-semideviation model.

This article is organized as follows: in the empirical framework section, we present in detail the different methodologies utilized throughout the article, then in empirical application we give some context in our data base and the proposed index insurance product for Paraná. Our findings and discussion are found in results and discussion and we finish with conclusions.

3.2. Empirical Framework

This section outlines the conceptual framework strategy used in the article. We present an overview of the methods used to model the yield-index relationship and to evaluate the proposed index insurance contract.

3.2.1. Quantile Regression

Quantile regression models the causal effects of covariates in different quantiles of the cumulative distribution function of the response variable, and therefore are an alternative approach to the usual linear regression methodology. That is, while the classical models are limited to the analysis of conditional mean, the quantile regression allows analysis throughout the conditional distribution of the response variable in the covariates.

The quantile regression models emerged as a generalization of the absolute residual minimization method developed in the early 19th century. The quantile regression has long stumbled on the difficulty of estimating the parameters that, unlike the usual linear regression models, have no analytical formula. However, with the advent of computers, as well as the development of linear programming techniques, the methodology has been gaining more space in empirical studies and academic research.

Let $Y_i, i = 1, \ldots, n$, random variables and $\boldsymbol{x}_i \in \mathbb{R}^p$ the observed vector of covariates. Consider that the variables Y_i are conditionally independent given $\boldsymbol{x}_i$; $\forall i = 1, \ldots, n$.

Whereas the usual regression is limited to describing the relationship of Y_i with the covariates of the study under the terms of conditional methods, quantile regression is a statistical modeling technique which allows analyzing this relationship in any quantile τ of interest, $\tau \in [0,1]$. In other words, it is a methodology capable of describing the function $f(\cdot, \tau)$ such that

$$Q_{Y_i|\boldsymbol{x}_i}(\tau) = f(\boldsymbol{x}_i, \tau) \tag{1}$$

for all $\tau \in [0,1]$. The function $f(\cdot, \tau)$ is the systematic part of the regression model. Note that $f(\cdot, \tau)$ may be different for each τ.

An intuitive way of understanding quantile regression, which is usually presented in the area literature, is an analogy to classical regression models (see, for example, Koenker (2005)).

In this case, each observed value of the response variable of the study is given by the sum of a systematic part, which is the quantile of order τ of Y_i, $f(\boldsymbol{x}_i, \tau)$ and a random error ϵ_i. This is:

$$y_i = f(\boldsymbol{x}_i, \tau) + \epsilon_i,$$

with independent and identically distributed ϵ_i. Assuming that the order τ of ϵ_i, conditional to $\boldsymbol{x}_i$, is equal to zero, note that the function to be modeled can be expressed as presented in (1).

As discussed in Koenker (2005), for example, the assumption of errors identically distributed is not a necessary condition for adjusting the quantile regression. Unlike the classical regression methodology, the quantile regression models can incorporate heteroscedasticity information from independent random errors.

Once the concept of quantile regression has been defined, it is necessary to understand how to interpret the model coefficients. Consider, for example, that $\mathrm{f}(\boldsymbol{x}_i, \tau) = \boldsymbol{x}_i^T \boldsymbol{\beta}(\tau)$ for τ fixed. In this case, the interpretation of the parameters $\beta(\tau)$ is essentially the same as the linear model, being the rate of change. That is, the coefficient $\beta_j(\tau), j = 1, \dots, p$, can be interpreted as the rate of change in the τ quantile of the variable Y by varying in a unit the value of the jth covariate while maintaining the values of other variables fixed. This is, $\beta_j(\tau) = \partial Q_{Y|\boldsymbol{x}}(\tau)/\,\partial x_j$.

3.2.2. Geographically Weighted Regression

Geographically Weighted Regression (GWR) is an extension of the classical linear regression for the analysis of non-stationary spatial data. The model allows its parameters to vary spatially, without limiting the form of this variation. The idea of GWR is to make a local adjustment for each point in the study region based on the closest observations. Thus, a continuous function $\beta_j(u_i, v_i)$ is created for each parameter, where (u_i, v_i) are the spatial coordinates of the *i*th point. The objective of GWR is to provide non-parametric estimates of these continuous surfaces using the kernel function.

The GWR model (Fotheringham et al., 2002) is presented below:

$$y_i = \beta_0(u_i, v_i) + \sum_j \beta_j(u_i, v_i) x_{ij} + \epsilon_i \tag{2}$$

$$\epsilon_i \sim N(0, \sigma^2).$$

Note that the assumptions of the classical regression model (Normal, homoscedastic and uncorrelated errors) remain. However, by allowing spatial variation for the parameters, the problems of autocorrelation and heteroscedasticity are reduced. The persistent limitation is normality, so this model is not yet the most suitable for treating spatial counting data, for example. It is interesting to note that classical regression is a special case of GWR. This simplification occurs when there is no spatial variation in the parameters.

Mathematically, $\beta_j(u_i, v_i)$ is estimated in matrix form by:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^T \boldsymbol{W}(u_i, v_i) \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_i, v_i) \boldsymbol{y} \tag{3}$$

where $\hat{\boldsymbol{\beta}}$ represents an estimate for $\boldsymbol{\beta}$, and $\boldsymbol{W}(u_i, v_i)$ is an $n \times n$ matrix with elements outside the diagonal equal to zero and diagonal elements representing the geographical weight of each observation at point *i*. Briefly, and defining (u_i, v_i) by (i) the parameters in each row of the matrix of Equation (3) are estimated by:

$$\hat{\beta}(i) = (\boldsymbol{X}^T \boldsymbol{W}(i) \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}(i) \boldsymbol{y}$$

where *i* represents the matrix line of Equation (2) and $\boldsymbol{W}(i)$ is a diagonal matrix of spatial weights $n \times n$ of the form:

$$\boldsymbol{W}(i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix},$$

where w_{in} is the weight given to point *n* in the calibration of the model for point *i*.

The estimator of Equation (3) is a weighted least square estimator but does not use a constant weight matrix. The weights in GWR, the values of the weighting matrix **W**, are calculated for each location *i*. In this way, each locality receives a different weight in the estimation in *i*, that is, a calibration is made for each point of interest. In this sense, the idea is that the weights are a measure of proximity of the observation to the point of estimation *i*.

The key point of this technique is the definition of the "circle of inclusion" of observations around point *i*, or more generally, of the spatial structure. The specified circle has a radius of size *h*. If *h* is too large, then almost all data will be included in the estimation of $\beta_j(u_i, v_i)$, making estimates close to the standard linear regression. If *h* is too small, few observations will be included in the calibration, resulting in $\beta_j(u_i, v_i)$ estimates with large standard errors. Finding the best *h* size is therefore extremely important in finding the best GWR fit.

The weight characteristic is also relevant in the adjustment, since it can be done in a discrete or continuous way, as discussed by Brunsdon et al. (1998). In the discrete case, to perform the calibration some points are excluded according to some criterion, for example an inclusion circle with radius *h*, that is, for a given locality *i*, the weight w_{ik} given to locality *k* can be:

$$w_{ik} = \begin{cases} 1, \text{if } d_{ik} < h. \\ 0, \quad else. \end{cases}$$

where d_{ik} is the distance between *i* and *k*. Or another possibility:

$$w_{ik} = \begin{cases} 1, if\ k\ is\ one\ of\ the\ N\ closest\ neighbours\ of\ i, \\ 0, \qquad\qquad\qquad else. \end{cases}$$

The continuous case considers that the *k* localities closest to the locality *i* have more weight in the estimation than more distant localities, in addition the continuous form can follow diverse distributions. In the Gaussian case, the w_{ik} weight can be represented by:

$$w_{ik} = exp\left(\frac{-d_{ik}^2}{2h^2}\right).$$

In this situation, the weight value gradually decreases with distance and can be written:

$$w_{ik} = \begin{cases} [1 - (d_{ik}/h)^2]^2, \text{if } d_{ik} < h. \\ \qquad 0, \qquad\qquad else. \end{cases}$$

These functions are known as "Kernel functions" or Kernels and are denoted by the letter *K* such as: $w_{ik} = K(d_{ik})$. Note that *h* also defines the degree of influence of each observation. The problem, then, is to estimate the constant *h*, sometimes referred to as Kernel bandwidth or smoothing parameter, which also functions as a variability factor of the weight curve.

It is known that the results of GWR are relatively indifferent to the choice of Kernel function but are highly sensitive to the smoothing parameter of the Kernel function used (Fotheringham et al., 2002). In the more general case, a constant smoothing parameter for all points is efficient if the points are equally spaced. However, where data are not equally spaced (spatially dispersed or when areas have different sizes), a constant smoothing parameter might prove suitable for some, but not all, locations. This is because the estimated parameters may have large standard errors due to the few points used in the calibration, or in extreme cases, the estimation would not be possible due to the lack of variability. Thus, to reduce these problems, it is possible to use a variable smoothing parameter, which allows a large smoothing parameter, where the data is scattered, and a small smoothing parameter, where the data is more abundant.

A solution to determine the smoothing parameter is cross-validation (CV), which was suggested by Cleveland (1979), for the local regression of the form:

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h)]^2 , \quad (4)$$

where $\hat{y}_{\neq i}(h)$ is the adjusted value for point y_i, omitting the observation *i*. When *h* becomes the smallest possible, the model is calibrated only in samples near *i* and not *i*. The value that minimizes Equation (4) is the optimal smoothing parameter of the cross-validation method.

It is important to note that the weighted least squares method for the GWR produces biased estimates for the parameters. The bias arises because the model adjusts local regressions assuming that the surface of the parameters is approximately flat in the vicinity of the analyzed regression point, when in fact the parameters probably vary continuously in the space. On the other hand, considering that there is no spatial stationarity, the estimates of the global regression model will be even more biased, since it assumes that the parameter is constant in every study region.

The bias of the GWR estimates, as well as the variance, will depend on the smoothing method. The choice of a very large smoothing parameter gives us an accurate estimate (with less variance) for the parameter, however, when considering more distant points in the calibration of the model, we are introducing bias in this estimation. The other extreme produces opposing results, that is, a small smoothing parameter produces an unbiased estimate, but with more variance, since it is based on a smaller sample size.

However, Staniswalis (1989) shows that under certain conditions (such as limited log-likelihood functions with first, second, and third derivatives also limited, and $b \to 0$ when $n \to \infty$), estimators that maximize the local likelihood, in this case $\widehat{\beta_j}(u_i, v_i)$, are asymptotically normal, non-biased and consistent.

3.2.3. Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO model, originally proposed by Tibshirani (1996), aims to shrink the parameters of a regression model allowing some of them to assume null value. Thus, the technique simultaneously produces the selection of the relevant variables in the model and the estimation of their respective coefficients. The estimates are obtained by a model of minimization of the error in the data subject to a penalty in the norm $L1$ of the coefficients:

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \underset{\beta_0,\beta_1,\ldots,\beta_p}{arg\ min} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ji} \right)^2 subject\ to \sum_{j=1}^{p} |\beta_j| \leq a \quad (4)$$

where *a* is the adjustment parameter that determines the intensity of the shrinkage. Written in the form of a Lagrangian, equation (4) takes the following form:

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \underset{\beta_0,\beta_1,\ldots,\beta_p}{arg\ min} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ji} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

where parameter $\lambda \geq 0$ becomes the intensity of the shrinkage. Put in matrix format, the model is such that:

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \underset{\boldsymbol{\hat{\beta}}}{arg\ min} \parallel \boldsymbol{Y} - \boldsymbol{X\beta} \parallel_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

where $\boldsymbol{\beta}$ is the vector of parameters $p \times 1, \boldsymbol{Y} = (y_i, \ldots, y_n)'$ is the data vector for the dependent variable, $\boldsymbol{X}$ is the matrix $p \times n$ of data from the series of predictors and $\lambda \geq 0$ is the shrinkage parameter.

Thus, the shrinkage parameter λ plays a fundamental role in the model. As λ is reduced to zero or close to zero, λ reaches a value λ_{OLS} such that the regularization term becomes insignificant and the parameters estimated by the LASSO method will be equivalent to those obtained by an OLS model. Otherwise, taking $\widehat{\boldsymbol{\beta}}^{LASSO}$ as the vector $p \times 1$ of estimated parameters obtained through the LASSO and $\widehat{\boldsymbol{\beta}}^{OLS}$ as the vector $p \times 1$ of estimated parameters obtained through the OLS regression, when $\lambda = \lambda_{OLS}$, $\widehat{\boldsymbol{\beta}}^{LASSO} = \widehat{\boldsymbol{\beta}}^{OLS}$. As the value of λ increases, the regression parameters are shrunk to the case where only the intercept remains in the model, i.e. all other parameters are shrunk to zero.

3.2.4. Geographically Weighted Quantile LASSO

A natural extension to the GWR is the geographically weighted quantile regression (GWQR) model, which has the following form:

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_\tau(u_i, v_i) + \epsilon_{\tau,i}$$

where $\epsilon_{\tau,i}$ is the random error term, τ is the quantile of interest, Y_i and $\boldsymbol{X}_i^T = [X_{i1}, \dots, X_{ip}]$ are respectively, the response variable Y and the explanatory variables $\boldsymbol{X}_1, \dots, \boldsymbol{X}_p$ at the geographical location $(u_i, v_i)(i = 1, \dots, n)$.

If $\rho_\tau(z) = z(\tau - I(z < 0))$ is the check loss function at quantile $\tau \in (0, 1)$, with $I(\cdot)$ as the indicator function. For a location (u_t, v_t), let $d_{it} = \| (u_i, v_i) - (u_t, v_t) \|$, where $\|\cdot\|$ is the Euclidean norm. According to Chen et al. (2012), the local-linear GWQR estimates of the coefficients, and their partial derivatives, are the ones that minimize the local weighted quantile loss function:

$$\mathcal{L}_h(u_t, v_t) = \Sigma_{i=1}^n \rho_\tau \left\{ Y_i - \boldsymbol{X}_i^T \begin{bmatrix} \boldsymbol{\beta}_\tau(u_t, v_t) + \boldsymbol{\beta}_\tau^{(u)}(u_t, v_t)(u_t - u_i) + \dots \\ \dots + \boldsymbol{\beta}_\tau^{(v)}(u_t, v_t)(v_t - v_i) \end{bmatrix} \right\} K_h(d_{it})$$

with respect to $\beta_\tau(u_t, v_t)$, the partial derivatives of $\boldsymbol{\beta}_\tau(u, v)$; $\boldsymbol{\beta}_\tau^{(u)}(u_t, v_t)$ and $\boldsymbol{\beta}_\tau^{(v)}(u_t, v_t)$ for a specified kernel function $K_h(\cdot) = K(\cdot/h)/h^2$ and bandwidth h. The latter is chosen via a cross validation procedure that is identical to its GWR counterpart, while the only difference is in the check loss function replacing the quadratic loss function.

Applying the aforementioned LASSO method to the GWQR we have:

$$\mathcal{L}_{h,\lambda} = \textstyle\sum_{t=1}^n \mathcal{L}_h(u_t, v_t) + \sum_{j=1}^p \left(\begin{matrix} \lambda_{1j} \, \| \left[\beta_j(u_1, v_1), \dots, \beta_j(u_n, v_n)\right]^T \| + \cdots \\ \dots + \lambda_{2j} \, \| \begin{bmatrix} \beta^{(u)}(u_1, v_1), \dots, \beta^{(u)}(u_n, v_n), \\ \beta^{(v)}(u_1, v_1), \dots, \beta^{(v)}(u_n, v_n) \end{bmatrix}^T \| \end{matrix} \right) \tag{5}$$

where $\boldsymbol{\lambda}_1 = (\lambda_{11}, \dots, \lambda_{1p})^T \in \mathbb{R}^p$ and $\boldsymbol{\lambda}_2 = (\lambda_{21}, \dots, \lambda_{2p})^T \in \mathbb{R}^p$ are the tuning parameters. This combination of the GWQR technique and the lasso method, is named by Wang et al. (2018) the geographically weighted quantile lasso (GWQLASSO).

Given that both the local weighted quantile loss function and the penalty function in (5) are nondifferentiable at the origin, what results in the common derivative-based algorithm being unusable for obtaining the solution of $\mathcal{L}_{h,\lambda}$. Therefore, a quadratic approximation (Hunter and Lange, 2000) is used to approximate the local weighted quantile loss function, while the local quadratic approximation (Fan and Li, 2001) is used to approximate the penalty function and establish the iterative algorithm of the GWQLASSO.

3.2.5. Spectral Risk Measures

Traditional risk measures, such as the value at risk (VaR) and expected shortfall (ES) have some limitations. The two measures of risk do not explicitly consider the degree of risk aversion of the user of the method (Cotter & Dowd, 2010). It is implicit, when using VaR as a risk measure, that the agent has a negative risk aversion, whereas the choice of ES implies risk neutrality (Grootveld & Hallerbach, 2004). In the case of VaR, the negative risk aversion is explicit when it is verified that the agent does not weigh the losses that exceed the VaR. For ES, risk neutrality is illustrated by the fact that the agent weighs losses that exceeds the VaR uniformly. Therefore, Acerbi (2002), Dowd, Cotter and Sorwar (2008) and Cotter and Dowd (2010) argue that VaR and ES are not consistent risk measures when the agent using the technique has risk aversion.

To overcome this limitation, Acerbi (2002) proposed a measure of spectral risk that is consistent when applied to agents with risk aversion. Thus, consider the risk measure defined by:

$$M_{\varphi} = \int_0^1 q_p \, \varphi(p) dp$$

where q_p is the quantile p of the distribution of losses, $\varphi(p)$ is a weight function defined in p, and p is a cumulative probability interval such that $p \in [0,1]$.

The measure of risk M_{φ} satisfies the conditions of coherence if and only if $\varphi(p)$ satisfies the following properties:

- $\varphi(p) > 0$: the weights must always be non-negative.
- $\int_0^1 \varphi(p) dp = 1$: the sum of the weights must be equal to the unit.
- $\varphi'(p) \geq 0$: high losses are associated with weights greater than or equal to losses of smaller magnitude.

Now, one must select a suitable risk aversion function that satisfies the above properties. Here we use the exponential function of risk aversion:

$$\varphi(p) = \frac{ke^{-k(1-p)}}{1 - e^{-k}}$$

where $k > 0$ is the absolute risk aversion coefficient. This measure of spectral risk attributes greater weights to losses in the higher levels of cumulative probability distribution (the worst losses). In addition, for any dp, the weights vary more rapidly the more risk averse the agent is. The growth rate depends on the value of k, that is, the more risk averse the investor, the more the weights will grow.

3.2.6. Mean Semi-deviation

The standard deviation considers both the below and above average values to be equally undesirable, and this may not be consistent with the objectives of the farmers, as the concern is generally about losses, which become more serious in the case of distributions. Alternatively, we can use an indicator that considers only the dispersion of values on the left side of the distribution, that is, the semideviation, given by:

$$\sigma_{ssd} = \sqrt{\frac{\sum_{i=1}^{n}(min[0, w_i - \bar{w}_i])^2}{n}}$$

where σ_{ssd} is the default semideviation of the wealth stream, w_i are the wealth values generated by the Bayesian bootstrap procedure, and $\bar{w}_i$ is the critical point below which the farmer cares, and *n* is the number of observations. The value of $\bar{w}_i$ represents the minimum acceptable return, that is, the point at which the dispersion of the left distribution is measured.

The concept of semideviation is not new, and its applications in the area of finance have emerged with Markowitz (1959), who in his classic book notes that the choice between the two measures depends on the convenience, familiarity, and differences between the portfolios produced by different metrics, among other pertinent characteristics.

An important feature to be emphasized is that the numerical value of the standard deviation is at least equal to the semideviation. The immediate implication is that we cannot make a comparison between the standard deviation and the semi deviation, even though the two have equal units.

Thus, the mean semi-deviation method is expressed by σ_{ssd} and:

$$U_{it} = \begin{cases} W_{it} - E(\boldsymbol{W}_i), \text{if } W_{it} < E(\boldsymbol{W}_i) \\ 0, \quad else \end{cases}$$

where U_{it} is the farmer utility and E is the expectation operator. The exposure to adverse weather conditions relative to the semideviation is then measured by:

$$V_i = E(\boldsymbol{W}_i) - \frac{1}{2}k\sigma_{ssd}$$

where V_i is the revenue risk. A higher value of V_i is conditioned to a lower level of semideviation, thus indicating less exposure to weather risk.

For both risk measures we chose a k value of 0.5 following Conradt et al (2015) and Dowd et al. (2008).

3.3. Empirical Application

3.3.1. Data cleaning and yield detrending

We continue the analysis initiated in chapter 1 using the filtered National Water Agency (ANA) monthly precipitation data set. For this chapter we focused only on municipalities, in the state of Paraná, with an operational weather station. The time series spans from 01/10/1979 through 01/04/2015 for a total of 41 weather stations, one per municipality. We also use the series of annual soybean yields for these 41 municipalities, from 1980 through 2015, obtained from the National Institute of Geography and Statistics (IBGE).

Crop yields were detrended using the following equation (Duarte et al., 2018):

$$\tilde{y}_{t,i} = \hat{y}_{2015,i}\left(1 + \frac{\hat{e}_{t,i}}{\hat{y}_{t,i}}\right)$$

where $\tilde{y}_{t,i}$, $\hat{y}_{t,i}$, $\hat{y}_{2015,i}$ and $\hat{e}_{t,i}$ are, respectively, the corrected yield, the fitted yield, the fitted yield for 2015 and the residual for year t and municipality i.

3.3.2. Data pre-processing and clustering

As detailed in chapter 1 we applied Multiple Imputation by Chained Equations (MICE) using the R software (Van Buuren, 2000). We calculated the standardized precipitation index (SPI) with a three-month scale, thus capturing severe drought events during the crop season (Mckee et al, 1993). We chose the Ward's clustering method with an Euclidean distance matrix since it has already proved successful in defining homogenous precipitation regions in Brazil (Keller Filho, 2005). The optimal number of clusters was obtained through majority vote of 30 indices, an algorithm implemented in Charrad et al (2014).

3.3.3. Weather Index-insurance

The state of Paraná is an important producer of soybean, being the second largest producer in Brazil. In spite of the evolution in crop technology and crop management, yields are highly susceptible to drought in some regions of the state, with as much of 50% of the final yields being dependent on water availability (Farias et al., 2001; Carmello & Sant'anna Neto, 2016).

Our WII hypothetical product is based on the standardized precipitation index (SPI) rather than cumulative rainfall. We chose this approach as there is a weak correlation between monthly precipitation and yields. This is because water availability depends on variables other than rainfall, such as water storage capacity in the soil and evapotranspiration potential, which is greatly influenced by air temperature (van Lier, 2014). The option for a rainfall-based index is also due to the better coverage of rainfall stations in Paraná.

The Standardized Precipitation Index (SPI) is based on the probabilities of overcoming a certain accumulated precipitate volume. Rainfall values are summed over several scales, for example 3, 6, 12 or 24 months, depending on the interest or need of the analyst. For a given month, for example, October, the 7-month SPI (SPI-7) is obtained from the sum of the precipitations over the seven months preceding the reference month.

The series of data, resulting from the sum of the precipitations over the months, is then adjusted to a probability distribution. In the original formulation, McKee (1993) used the Gamma distribution. From the adjustment of the probability distribution, each element of the adjusted series is assigned a probability of non-overrun. Each of these probabilities of non-overrun is finally associated with the corresponding quantile of the standard normal distribution. The quantile value of the $N(0,1)$ associated with the probability calculated in the period of interest is the SPI value for the month.

One of the advantages of using SPI, according to McKee (1993), is that SPI is only a function of probability. Thus, regardless of the probability distribution function to be used, the SPI can be properly calculated.

Other advantages are that SPI is able to characterize both dry and rainy periods, as well as the fact that it is suitable for any hydrological variable. However, the use of this index also has limitations. Mishra and Singh (2010) argue that the main one is the need for long historical records for its consistent calculation, which is not always possible (Weschenfelder et al., 2011).

The relationship between SPI and soybean yields is then modeled using the GWQLASSO framework. We follow Conradt et al. (2015) and use a method based on the inverse function of the estimated regression to determine the triggers and exits of the contract. This approach permits a precise definition of the coverage level and does not require individual tinkering of the product parameters for each location, thus facilitating and streamlining product development. For our study we chose a coverage level of 100% of the expected yield.

In preliminary assessments we found that the 1-month SPI has the highest correlation with soybean yields, thus we only present here the results for this index from October through March, the months that correspond to the planting and harvesting of soybean in most of Paraná[6].

3.3.4. Premium Estimation

The insurance premium is derived from the probability distribution function (pdf) of indemnities, or an approximation of this distribution. In our study, we use the Historical Burn Analysis (HBA) method to approximate the pdf of indemnities. This method is based in actual realizations of the proposed index which are then converted in payouts. The average value of these payouts represents the expected loss.

HBA is the simplest method to estimate an insurance premium, it also does not require assumptions on the pdf parameters, in contrast to other methods such as Historical Distribution Analysis and Monte Carlo based methods (Hess et al., 2005). We refrain from using these latter methods as our data is aggregated at the municipality level and thus it may misrepresent variability at the farm level.

In order to provide a representative data set for the premium estimation we use the first 30 years of data for this part of the analysis, with the remaining six years being used for the evaluation of the methods.

3.3.5. Product Evaluation

For the product evaluation we use the values of the final wealth realizations for a hypothetical farm with an area of 1 ha. The only assets present in such farm are the soybean yield and the proposed weather index insurance contract:

$$W_{it} = (1/60)vy_{it} + I_{it} - P_i,$$

where W_{it} is the final wealth, v is the price paid to the farmer for each 60kg of soybean[7], y_{it} is the corrected yield, I_{it} is the indemnity and P_i the premium, with i being the municipality and *t* is the year. Final wealth realizations are

---

6 Planting and harvesting progress reports are available at the state level, with the months of October and March corresponding to more than 50% of the total crop area planted/harvested. These months are also assumed as planting and harvesting dates in Franchini et al. (2016).

7 Considering we corrected yields we utilized the 2015 average prices of soybean provided by the Department of Rural Economy (DERAL) of the State Secretariat for Agriculture and Food Supply (SEAB) in Paraná.

calculated for farmers without insurance, thus having only the first component of the right-hand side, and for farmers with the WII parameters estimated by the quantile regression and the GWQLASSO.

In order to measure the efficiency of the proposed index insurance to mitigate the risk faced by farmers we use two risk measures, namely the Spectral Risk Measure (SRM) and the Mean Semi-deviation, coupled with a Bayesian bootstrap procedure. The latter is necessary given that we only dispose of only six years of data for the evaluation. In this step a cross-validation (CV) method would be ideal but the computational requirements of GWQLASSO makes the use of CV not feasible in our case.

The Bayesian bootstrap (Rubin, 1981) is very similar to its classical counterpart (Efron, 1979) differing only in how probabilities are attached to each data value. While in the classical bootstrap a $1/n$, being n the sample size, is attributed to all n observations, in the Bayesian bootstrap the probabilities are given by a posterior distribution centered in $1/n$ but varying across replications. The main difference is in the interpretation of the results as the Bayesian bootstrap is a simulation of the posterior distribution of the parameter being estimated, whereas the classical bootstrap simulates the sampling distribution of an estimator for the parameter of interest.

The relative risk reduction (RR) is structured to compare the risk exposure of farmers in three situations: the first one being a farmer with a WII insurance designed with the GWQLASSO method against a farmer without insurance; the second situation is a farmer with WII insurance (designed with GWQLASSO or quantile regression); and a third situation for a WII insurance (designed with GWQLASSO) versus a yield insurance (YI). Thus, the general formula for the RR is:

$$RR_{case\ 1/case\ 2} = \frac{RM(W_{case\ 1}) - RM(W_{case\ 2})}{RM(W_{case\ 2})},$$

where RM stands for the risk measurements previously described and W for the final wealth realizations.

In our evaluations we consider 4000 Bayesian bootstrap replications to provide better estimates of the relative risk reduction. The latter is also tested against a hypothesis of null relative risk reduction by means of a non-parametric Wilcoxon test.

3.4. Results and Discussion

3.4.1. Cluster analysis

As detailed in chapter 1, the optimal number of clusters from the precipitation data was two. These clusters managed to capture the different precipitation regimes identified by Keller Filho et al (2005), with cluster 1 representing areas with higher total precipitation in the year aggregate but greater variability among years and cluster 2 indicating areas with a lower total precipitation but with less variability. For the yield data, the optimal number of clusters was also two, with both clusters presenting a similar yield level from the beginning of the series through 1990 and from 2001 onwards, however, in the period comprised between 1991 and 2000 cluster 1 has lower yields. Cluster 1 contains municipalities in regions prone to drought, and thus presents lower yields and higher variability.

3.4.2. Yield-index modelling

We observe for the cluster representing the western and northern portions of Paraná that the December SPI presents the greatest impact on yields (Figures 8 and 9). Given that we assume, based on state reports[8], soybean planting dates are beginning on October, the crop would be in the reproductive stage in December, thus, highly sensible to water shortage. Therefore, for the premium estimation in cluster 1 we select the December SPI as the index. For the central and eastern portions of the state, however, both December and February SPI are impacting yields, with the February SPI having a slightly higher impact and thus being the one selected as the index for cluster 2.

Note that these coefficients are selected based on their boxplot as GWQLASSO does not account for temporal structure of the data, thus assuming that all observations come from a unique point in time and resulting in 30 coefficients for each location. Albeit this represents a limitation to the modelling, the impact is reduced as we used time-detrended yields and are not interested in the temporal behavior of the yield/index series, just in their intrinsic relationship. Also, the incorporation of the LASSO method permits a better identification of relevant variables, as the ones with little importance to yields rapidly converge to zero.

The possibility to model yields at several locations simultaneously while considering the spatial structure of the relationship between yields and explanatory variables and also having a method to quickly dismiss unimportant variables present a great opportunity to WII scalability. One of the major issues of WII is its low ability to grow at scale as the models developed for one location may prove completely obsolete as you move away from it, however, with GWQLASSO one may inspect both the general significance of the explanatory variables and their coefficient for each location. This permits a faster screening of possible indices, along with their respective triggers when using the methodology detailed here and proposed in Conradt (2015).

Another hindrance to the spread of WII is the absence of long series of yield data. The GWQLASSO method is less affected by this issue, as it uses information from neighboring yields in the estimation process. This characteristic is especially important in developing countries, which in general does not have long series of yield and weather data.

[8] Available in: http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=32

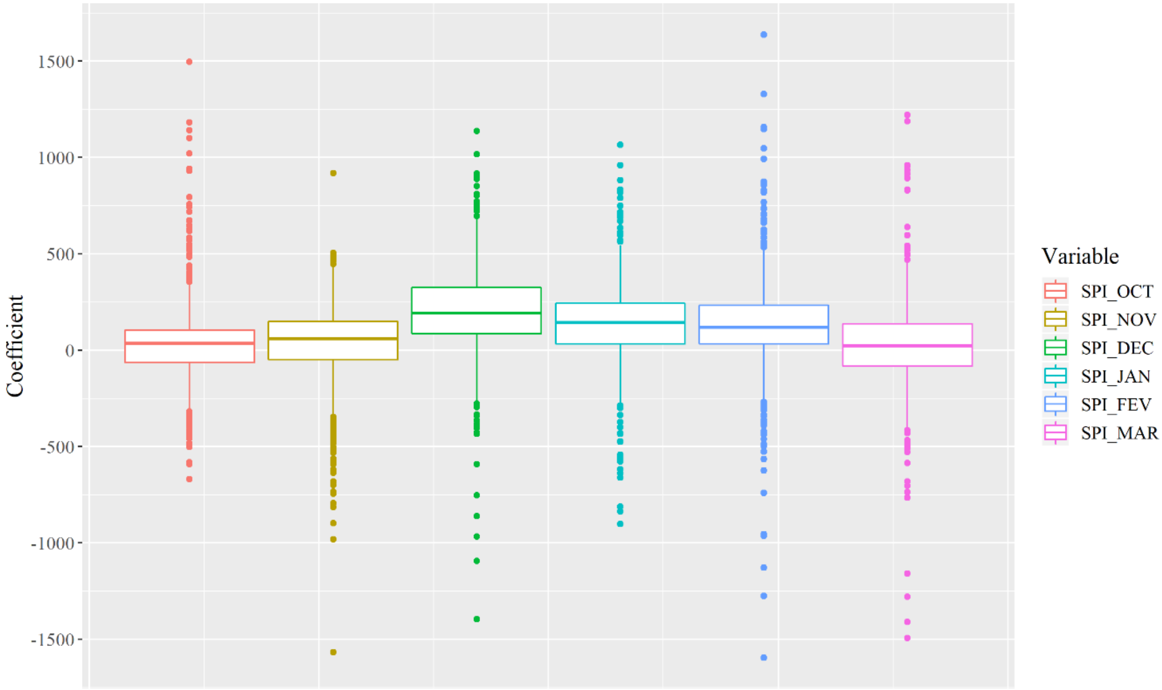

Figure 8. Boxplot of the GWQLASSO coefficients for cluster 1.

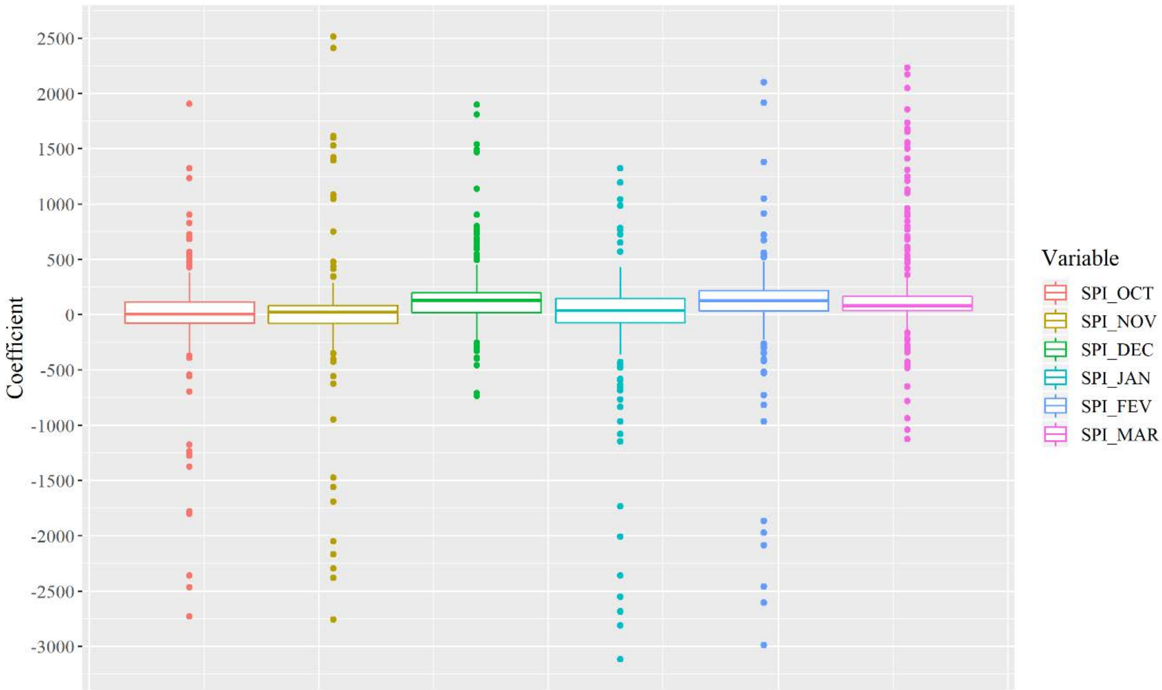

Figure 9. Boxplot of the GWQLASSO coefficients for cluster 2.

The relationship between the December SPI and yields, for cluster 1, varies by more than five times when we compare municipalities on the west of Paraná to the ones in the center region of the state (Figure 10). This goes in line with the characteristics of these regions, soils to the west, mostly in the northwest, are sandy, and the climate is classified as Cfa, with higher temperatures in the summer, both unfavorable to soybean. These characteristics result in diminished yields and higher susceptibility to drought, which are translated in the coefficients for December

SPI. A higher variability in yields, when compared to other regions in the state, coupled with susceptibility to drought is also observed by Franchini et al (2016).

The center and east of the state are classified as Cfb in Koeppen's system. This means that these municipalities have lower temperatures in the summer, what benefits soybean plants. Also, the soil in these regions have more clay, what leads to a higher capacity to contain water and consequentially mitigate the effects of drought, resulting in a less pronounced relation to the index, for both cluster 1 and 2 (Figures 10 and 11).

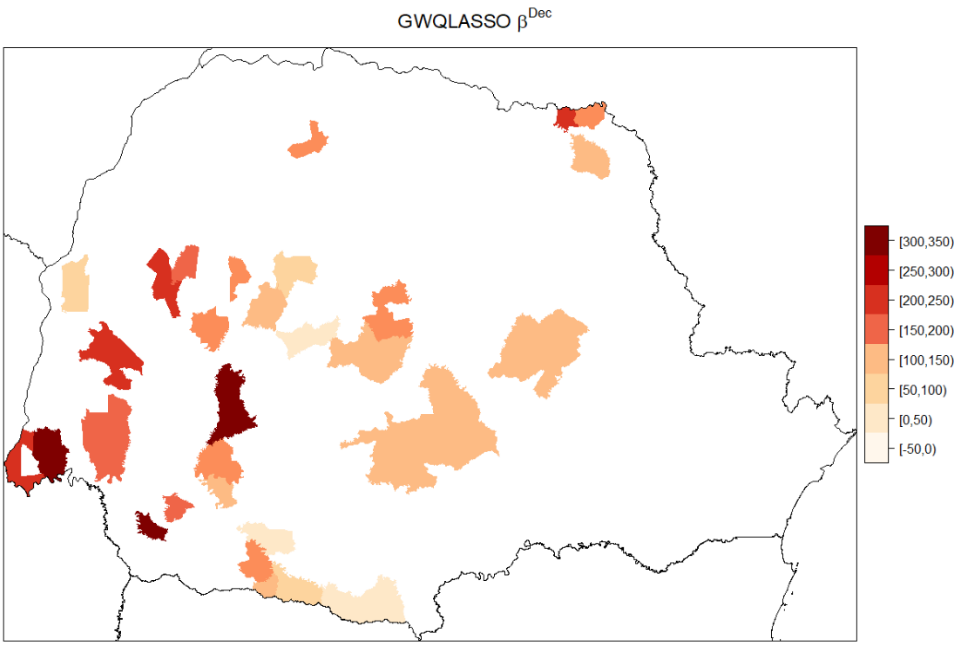

Figure 10. GWQLASSO β^{Dec} coefficients spatial distribution for cluster 1.

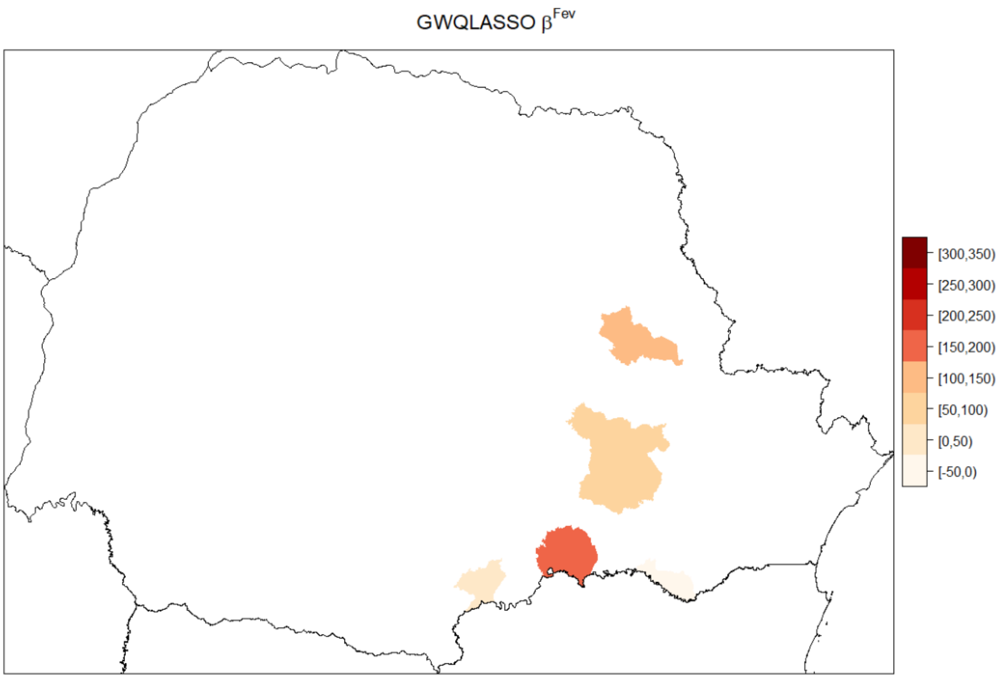

Figure 11. GWQLASSO β^{Mar} coefficients spatial distribution for cluster 2.

3.4.3. Weather Index Insurance premium and performance

WII is generally associated with lower premiums, mainly due to the lack of in situ crop inspection after a claim is filed. Here we compare a traditional yield insurance product with a 65% coverage to our proposed WII product with a 100% coverage. While this comparison may not be fair to our product, as by definition a higher coverage means a higher premium, we found the 65% coverage level to be the most common for soybean in Paraná. Coverage levels above 90% are rare in Brazil and suffer from two problems, the higher premiums and the inferior percentage in subsidization by the government, thus they are not attractive to the farmer.

Our results show that the index insurance may vary from half to three times the price of the common yield insurance. There is a tendency of pricier index insurance, compared to the yield insurance, as we move to the western portion of Paraná (Figure 12), what is expected as this region is more susceptible to drought.

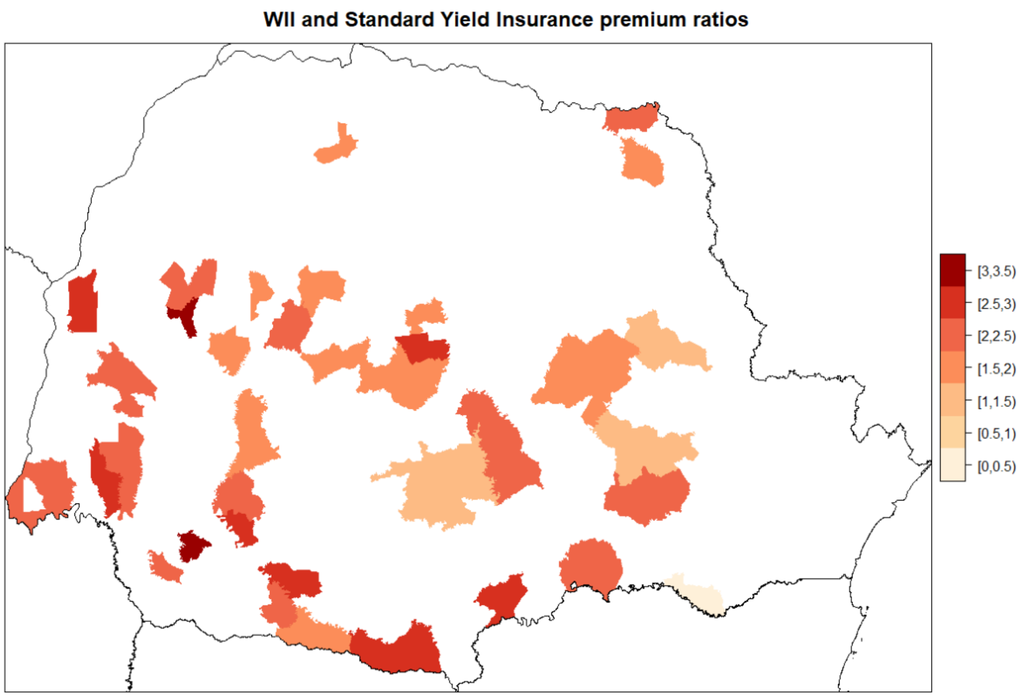

Figure 12. Ratio of GWQLASSO calculated premiums to standard yield insurance premiums.

In our design we do not consider gains from scale and the spatial diversification of risk by the insurer, this would lead to a lesser difference between our product and the commercial product depicted here. Even so, the WII results in a net gain for the producer, as for both clusters and risk measurements it performs better than the yield insurance. The results from both risk measures indicate that both GWQLASSO and quantile regression provide similar risk reduction, with both being more effective than a yield insurance product with a 65% coverage level (Tables 3 through 6). Borrowing from results in Conradt (2015) we can also derive that GWQLASSO is superior to ordinary least squares.

Table 3. Relative risk reduction per municipality in cluster 1 according to the Spectral Risk Measure (SRM)

Municipality	$RR_{GWQLASSO}$			$RR_{GWQLASSO/QR}$			$RR_{GWQLASSO/YI}$		
	2.5%	Median	97.5%	2.5%	Median	97.5%	2.5%	Median	97.5%
Alto Piquiri	-0,1269	-0,0194	0,1459	-0,1127	-0,0056*	0,1117	-0,0691	0,0478*	0,2450
Ampére	-0,1291	-0,0150	0,1711	-0,1113	-0,0012	0,1221	-0,0816	0,0401*	0,2398
Andirá	-0,1317	0,0841	0,4253	-0,1722	0,0191*	0,2521	-0,0758	0,1470*	0,4714
Cambará	-0,1907	0,0076	0,2491	-0,1984	0,0025*	0,2538	-0,1523	0,0732*	0,3356
Campo Mourão	-0,1007	-0,0090	0,0953	-0,0927	0,0041*	0,1055	-0,0546	0,0398*	0,1621
Céu Azul	-0,1610	0,0004	0,2551	-0,1595	-0,001*	0,2099	-0,1101	0,0543*	0,2525
Clevelândia	-0,1156	-0,0077	0,1239	-0,1149	-0,0089*	0,1219	-0,0639	0,0424*	0,1830
Coronel Vivida	-0,2051	-0,0012	0,2519	-0,1915	0,0027	0,2465	-0,1703	0,0452*	0,2845
Formosa do Oeste	-0,1223	0,0513	0,3600	-0,1938	-0,0255*	0,1583	-0,0884	0,0939*	0,3761
Foz do Iguaçu	-0,2061	-0,0061	0,2812	-0,1968	-0,0007	0,2641	-0,1477	0,0468*	0,2916
Guaraniaçu	-0,0838	0,0086	0,1902	-0,1242	-0,0096*	0,1387	-0,0047	0,1035*	0,3146
Guarapuava	-0,0654	-0,0146	0,0306	-0,0536	-0,0015*	0,0517	-0,0159	0,0365*	0,0877
Ivaiporã	-0,0754	0,0101	0,0962	-0,0729	0,006*	0,0817	-0,0289	0,0664*	0,1580
Janiópolis	-0,0819	0,0148	0,1739	-0,1032	-0,0142*	0,0665	-0,0403	0,0678*	0,2515
Mamborê	-0,0609	0,0161	0,1068	-0,0766	-0,0017*	0,0708	-0,0072	0,0737*	0,1718
Manoel Ribas	-0,0960	-0,0051	0,0966	-0,0940	-0,0049*	0,0866	-0,0569	0,0423*	0,1548
Mariluz	-0,1125	0,0021	0,1730	-0,1149	0,0011*	0,1650	-0,0496	0,0722*	0,2705
Mariópolis	-0,1355	-0,0196	0,1762	-0,1117	0,001	0,1146	-0,0928	0,0255*	0,2021
Matelândia	-0,2177	-0,0355	0,2408	-0,1952	-0,0178*	0,2228	-0,1713	0,0107*	0,2166
Nova Esperança	-0,1745	-0,0164	0,1832	-0,1541	0,0047*	0,1982	-0,1191	0,0520*	0,2792
Palmas	-0,0514	0,0006	0,0545	-0,0511	0,0003	0,0556	-0,0098	0,0432*	0,1019
Pato Branco	-0,1411	-0,0264	0,1552	-0,1354	-0,0198*	0,1392	-0,0949	0,0344*	0,2290
Pitanga	-0,1261	-0,0224	0,0809	-0,1127	-0,008*	0,0985	-0,0768	0,0360*	0,1489
Prudentópolis	-0,0671	-0,0100	0,0406	-0,0643	-0,0025	0,0627	-0,0207	0,0386*	0,0914
Quedas do Iguaçu	-0,1423	-0,0341	0,1200	-0,0852	0,0212*	0,1112	-0,0926	0,0204*	0,1876
Roncador	-0,0955	-0,0026	0,1026	-0,0945	-0,0007	0,1072	-0,0466	0,0548*	0,1758
Salto do Lontra	-0,1587	-0,0316	0,1312	-0,1331	-0,0128*	0,1211	-0,1310	0,0035*	0,1887
Santo Antônio da Platina	-0,1100	0,0181	0,1963	-0,1099	0,0089*	0,1605	-0,0549	0,0811*	0,2803
São Jorge d'Oeste	-0,1958	-0,0146	0,2451	-0,1800	-0,0077*	0,2292	-0,1669	0,0148*	0,2518
São Miguel do Iguaçu	-0,1811	0,0151	0,3174	-0,2162	-0,0129*	0,2007	-0,1401	0,0603*	0,3187
Terra Roxa	-0,2323	0,0040	0,3032	-0,2307	-0,0034	0,2747	-0,1869	0,0432*	0,2877
Tibagi	-0,0546	0,0040	0,0777	-0,0510	-0,0017*	0,0481	-0,0094	0,0529*	0,1341
Toledo	-0,2095	-0,0251	0,2577	-0,1832	0,0057*	0,2260	-0,1671	0,0121*	0,2199
Ubiratã	-0,1043	-0,0065	0,0989	-0,1068	-0,0033*	0,1055	-0,0677	0,0425*	0,1584

*Significant at the 5% significance level
Source: Authors

Given that quantile regression and GWQLASSO are similar, other characteristics must be considered when choosing the estimation method for a new product. While quantile regression is effective and uses less computational power than GWQLASSO, it may suffer in locations with less data, being that it does not incorporate information from neighboring sites. Also, it must be estimated for each location individually, along with the screening for significant explanatory variables. This is not a problem for GWQLASSO, as already said, it not only takes surrounding data into account, but also indicates which variables are important for the model.

Table 4. Relative risk reduction per municipality in cluster 1 according to the Mean Semi-deviation

Municipality	$RR_{GWQLASSO}$			$RR_{GWQLASSO/QR}$			$RR_{GWQLASSO/YI}$		
	2.5%	Median	97.5%	2.5%	Median	97.5%	2.5%	Median	97.5%
Alto Piquiri	-0,1264	-0,0063	0,1975	-0,1179	-0,0075*	0,1260	-0,0695	0,0611*	0,2836
Ampére	-0,1329	0,0027	0,2175	-0,1182	-0,0010	0,1383	-0,0803	0,0626*	0,2942
Andirá	-0,1192	0,1149	0,5029	-0,1770	0,0145*	0,2711	-0,0740	0,1663*	0,5406
Cambará	-0,2145	0,0042	0,2565	-0,2207	0,0039*	0,2835	-0,1647	0,0705*	0,3415
Campo Mourão	-0,1131	-0,0083	0,1159	-0,1062	0,0035*	0,1195	-0,0691	0,0454*	0,1772
Céu Azul	-0,1823	0,0135	0,3513	-0,1821	0,0034	0,2479	-0,1391	0,0501*	0,2829
Clevelândia	-0,1209	-0,0042	0,1430	-0,1177	-0,0021	0,1336	-0,0727	0,0500*	0,2040
Coronel Vivida	-0,2210	-0,0029	0,3084	-0,2217	-0,0002	0,2879	-0,1888	0,0433*	0,3327
Formosa do Oeste	-0,1360	0,0647	0,4381	-0,2105	-0,0324*	0,1685	-0,0956	0,1005*	0,4181
Foz do Iguaçu	-0,2108	0,0094	0,3584	-0,2213	-0,0010*	0,3078	-0,1740	0,0479*	0,3105
Guaraniaçu	-0,0795	0,0211	0,2260	-0,1127	-0,0073*	0,1362	0,0077	0,1257*	0,3703
Guarapuava	-0,0719	-0,0177	0,0360	-0,0585	-0,0012*	0,0587	-0,0217	0,0341*	0,0897
Ivaiporã	-0,0740	0,0154	0,1006	-0,0688	0,0062*	0,0830	-0,0230	0,0713*	0,1632
Janiópolis	-0,0900	0,0280	0,2251	-0,1225	-0,0184*	0,0775	-0,0409	0,0883*	0,3023
Mamborê	-0,0645	0,0197	0,1175	-0,0765	-0,0024	0,0782	-0,0064	0,0777*	0,1889
Manoel Ribas	-0,1054	-0,0015	0,1108	-0,1014	-0,0023	0,1090	-0,0597	0,0513*	0,1739
Mariluz	-0,1200	0,0115	0,2177	-0,1216	0,0068*	0,1904	-0,0566	0,0846*	0,3290
Mariópolis	-0,1422	-0,0030	0,2378	-0,1351	-0,0043*	0,1226	-0,1032	0,0394*	0,2540
Matelândia	-0,2383	-0,0270	0,2879	-0,2206	-0,0153*	0,2695	-0,2096	0,0042*	0,2616
Nova Esperança	-0,1800	-0,0154	0,2114	-0,1637	0,0031*	0,2192	-0,1216	0,0619*	0,3204
Palmas	-0,0509	0,0004	0,0562	-0,0529	0,0009	0,0563	-0,0098	0,0432*	0,1041
Pato Branco	-0,1490	-0,0101	0,1972	-0,1407	-0,0062*	0,1673	-0,0984	0,0503*	0,2741
Pitanga	-0,1350	-0,0257	0,0858	-0,1160	-0,0076*	0,1105	-0,0837	0,0336*	0,1576
Prudentópolis	-0,0706	-0,0164	0,0356	-0,0613	0,0001*	0,0714	-0,0274	0,0324*	0,0884
Quedas do Iguaçu	-0,1467	-0,0209	0,1694	-0,1007	0,0139*	0,1168	-0,1034	0,0335*	0,2417
Roncador	-0,1031	0,0001	0,1166	-0,1056	-0,0033	0,1189	-0,0498	0,0580*	0,1901
Salto do Lontra	-0,1636	-0,0222	0,1632	-0,1433	-0,0165*	0,1332	-0,1265	0,0203*	0,2314
Santo Antônio da Platina	-0,1060	0,0345	0,2447	-0,1200	0,0155*	0,1951	-0,0464	0,1032*	0,3303
São Jorge d'Oeste	-0,2115	-0,0019	0,3096	-0,2074	-0,0022*	0,2820	-0,1863	0,0238*	0,3015
São Miguel do Iguaçu	-0,1936	0,0278	0,4372	-0,2257	-0,0153*	0,2169	-0,1591	0,0621*	0,3582
Terra Roxa	-0,2673	0,0036	0,4015	-0,2827	-0,0060	0,3495	-0,2494	0,0257*	0,3068
Tibagi	-0,0538	0,0110	0,0876	-0,0578	-0,0040*	0,0493	-0,0107	0,0594*	0,1430
Toledo	-0,2288	-0,0173	0,3539	-0,2166	-0,0005	0,2440	-0,2068	0,0039	0,2347
Ubiratã	-0,1190	-0,0087	0,1096	-0,1136	-0,0044*	0,1130	-0,0768	0,0434*	0,1789

*Significant at the 5% significance level
Source: Authors

As for the public policy implications, WII has proved to be a superior alternative to basic crop insurance products as a yield insurance with a 65% coverage level. Considering that Oñate (2016) showed that PROAGRO, a risk management tool similar to a crop-credit insurance, has not increased farmers welfare and is not priced according to regional characteristics, we favor the expansion of government operated or funded parametric insurance products. A WII product could be implemented as a microinsurance policy to small farmers or as a macroinsurance directly to the government. The latter would also be further advantageous, as the efficiency of WII grows with scale (Miranda and Farrin, 2012). Several products of this type have been successfully implemented in developing countries such as the “Comité de Ayuda a Desastres y Emergencias Nacionales” (CADENA) program in Mexico (de Janvry et al, 2016) and the “Pradhan Mantri Fasal Bima Yojana” (PMFBY) index insurance scheme in India (Rathore et al, 2017).

Table 5. Relative risk reduction per municipality in cluster 2 according to the Spectral Risk Measure (SRM)

Municipality	$RR_{GWQLASSO}$			$RR_{GWQLASSO/QR}$			$RR_{GWQLASSO/YI}$		
	2.5%	Median	97.5%	2.5%	Median	97.5%	2.5%	Median	97.5%
Palmeira	-0,0863	0,0083	0,0971	-0,0900	0,0039*	0,1028	-0,0470	0,0544*	0,1531
Piraí do Sul	-0,0359	-0,0019	0,0475	-0,0289	-0,0033*	0,0248	0,0055	0,0425*	0,0964
Ponta Grossa	-0,0427	-0,0058	0,0285	-0,0329	0,0032*	0,0444	0,0022	0,0411*	0,0782
Porto Amazonas	-0,1065	-0,0043	0,0976	-0,1010	0,0011*	0,1135	-0,0623	0,0428*	0,1542
Rio Negro	-0,0640	-0,0006	0,0623	-0,0598	0,0002	0,0665	-0,0096	0,0570*	0,1286
São Mateus do Sul	-0,0684	-0,0331	0,0048	-0,0318	-0,0045*	0,0240	-0,0160	0,0209*	0,0609
União da Vitória	-0,0705	-0,0017	0,0723	-0,0579	0,0056*	0,0690	-0,0279	0,0411*	0,1218

*Significant at the 5% significance level
Source: Authors

Table 6. Relative risk reduction per municipality in cluster 2 according to the Mean Semi-deviation

Municipality	$RR_{GWQLASSO}$			$RR_{GWQLASSO/QR}$			$RR_{GWQLASSO/YI}$		
	2.5%	Median	97.5%	2.5%	Median	97.5%	2.5%	Median	97.5%
Palmeira	-0,0993	0,0032	0,0995	-0,0994	0,0014	0,1127	-0,0543	0,0502*	0,1535
Piraí do Sul	-0,0330	0,0061	0,0579	-0,0309	-0,0034*	0,0268	0,0111	0,0513*	0,1083
Ponta Grossa	-0,0435	-0,0056	0,0304	-0,0338	0,0060*	0,0504	-0,0003	0,0422*	0,0815
Porto Amazonas	-0,1155	-0,0058	0,1047	-0,1101	0,0013*	0,1335	-0,0773	0,0422*	0,1638
Rio Negro	-0,0657	-0,0009	0,0679	-0,0607	0,0007	0,0704	-0,0114	0,0576*	0,1330
São Mateus do Sul	-0,0620	-0,0289	0,0096	-0,0310	-0,0045*	0,0229	-0,0112	0,0261*	0,0685
União da Vitória	-0,0757	0,0013	0,0812	-0,0664	-0,0012	0,0696	-0,0323	0,0444*	0,1291

*Significant at the 5% significance level
Source: Authors

3.5. Conclusion

Despite the efforts by the central government, crop insurance is yet to take off in Brazil. Inconsistent budget, information asymmetry and moral hazard are some of the issues that crippled the program and continue to impede its expansion. In this sense, parametric insurance may present an alternative to the local insurance market. Thus, aiming to foster the growth of parametric insurance in Brazil and contribute to the development of this type of insurance throughout the globe we design a WII product by using a novel approach to model the yield-index relationship, the GWQLASSO. This methodology compounds the flexible modelling and robustness of quantile regression with the spatial component of geographically weighted regression and variable selection prowess of the LASSO method.

We test our assumptions using a crop insurance application in Paraná, Brazil. The 36 years long time series of precipitation and soybean yield data are split in design and evaluation sets, with the latter having only six years of data and thus requiring the use of Bayesian bootstrap to improve the reliability of results. To measure the ability of WII to reduce risk, when compared to yield insurance and between yield modelling approaches, we use two different risk measures, the Spectral Risk Measure and the Mean Semi-deviation.

Regarding the performance of WII in Paraná our findings indicate that index insurance is superior to a 65% coverage yield insurance in 41 municipalities of the state, despite being up to three times more expensive than this product. However, the GWQLASSO approach proved as effective as the regular quantile regression as a hedging tool. The latter may seem as a discouragement to the use of a more complex model, nevertheless, some of the characteristics of GWQLASSO (less data intensive and simpler conjoint variable selection) argue in its favor.

Future studies are needed to confirm the viability of WII in other regions and crops throughout the country. Also, regardless of our efforts to mitigate the effect of the level of aggregation in the crop yield data and lack of precise planting dates, these may lead to a loss of accuracy in product design that is unacceptable in a commercial environment. Therefore, tighter cooperation between risk bearers and insurance researchers and product developers is needed.

References

Acerbi, Carlo. "Spectral measures of risk: A coherent representation of subjective risk aversion." *Journal of Banking & Finance* 26, no. 7 (2002): 1505-1518.

BANCO CENTRAL DO BRASIL - BACEN. Departamento De Regulação, Supervisão E Controle Das Operações Do Crédito Rural E Do Proagro – DEROP. *Programa De Garantia Da Atividade Agropecuária PROAGRO Relatório Circunstanciado 2015 a 2018*. 2018.

Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton. "Geographically weighted regression." *Journal of the Royal Statistical Society: Series D (The Statistician)* 47, no. 3 (1998): 431-443.

Cai, Jing. "The impact of insurance provision on household production and financial decisions." *American Economic Journal: Economic Policy* 8, no. 2 (2016): 44-88.

Carmello, Vinicius, and João Lima SANT'ANNA NETO. "Rainfall Variability and Soybean Yield in Paraná State, Southern Brazil." *International Journal of Environmental & Agriculture Research* 2, no. 1 (2016): 86-97.

Carter, Michael, Alain de Janvry, Elisabeth Sadoulet, and Alexander Sarris. "Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling." *Development Policies working paper* 111 (2014).

Cleveland, William S. "Robust locally weighted regression and smoothing scatterplots." *Journal of the American statistical association* 74, no. 368 (1979): 829-836.

Collier, B., J. Skees, e B. Barnett, 2009 Weather index insurance and climate change: opportunities and challenges in lower income countries. The Geneva Papers on Risk and Insurance Issues and Practice **34**: 401–424.

Collier, B., B. Barnett, e J. Skees, 2010 State of knowledge report–data requirements for the design of weather index insurance.

Chakravarti, J. S. "Agricultural insurance: a practical scheme suited to Indian Conditions." (1920).

Charrad, Malika, et al. "Package 'NbClust'." *Journal of Statistical Software* 61 (2014): 1-36.

Conradt, Sarah, Robert Finger, and Raushan Bokusheva. "Tailored to the extremes: Quantile regression for index-based insurance contract design." *Agricultural economics* 46, no. 4 (2015): 537-547.

Cotter, John, and Kevin Dowd. "Estimating financial risk measures for futures positions: A nonparametric approach." *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 30, no. 7 (2010): 689-703.

de Carvalho, José Ruy Porto, et al. "Model for Multiple Imputation to Estimate Daily Rainfall Data and Filling of Faults." *Revista Brasileira de Meteorologia* 32.4 (2017): 575-583.

De Janvry, Alain, Elizabeth Ramirez Ritchie, and Elisabeth Sadoulet. *Weather index insurance and shock coping: evidence from Mexico's CADENA Program*. The World Bank, 2016.

Dowd, Kevin, John Cotter, and Ghulam Sorwar. "Spectral risk measures: properties and limitations." *Journal of Financial Services Research* 34, no. 1 (2008): 61-75.

Duarte, Gislaine V., Altemir Braga, Daniel L. Miquelluti, and Vitor A. Ozaki. "Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance." *Journal of Applied Statistics* 45, no. 11 (2018): 1920-1937.

Efron, B. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7, no. 1 (1979): 1-26.

Fan, Jianqing, and Runze Li. "Variable selection via nonconcave penalized likelihood and its oracle properties." *Journal of the American statistical Association* 96, no. 456 (2001): 1348-1360.

Farias, José Renato Bouças, Eduardo Delgado Assad, IR de Almeida, Balbino Antônio EVANGELISTA, C. Lazzarotto, Norman Neumaier, and Alexandre Lima Nepomuceno. "Caracterização de risco de déficit hídrico nas regiões produtoras de soja no Brasil." *Revista Brasileira de Agrometeorologia* 9, no. 3 (2001): 415-421.

Fotheringham, Stewart, Chris Brundson, and Martin Charlton. *Geographically weighted regression & associated techniques*. Wiley, 2002.

Franchini, Julio Cezar, Alvadi Antonio Balbinot, Jr., Pablo Ricardo Nitsche, Henrique Debiasi, and Ivani De Oliveira Negrão Lopes. "Variabilidade Espacial E Temporal Da Produção De Soja No Paraná E Definição De Ambientes De Produção." (2016). Accessed September 22, 2018. https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1052786

Gallagher, Paul. "US soybean yields: Estimation and forecasting with nonsymmetric disturbances." *American Journal of Agricultural Economics* 69.4 (1987): 796-803.

Giné, Xavier, Lev Menand, Robert Townsend, and James Vickery. *Microinsurance: a case study of the Indian rainfall index insurance market*. The World Bank, 2010.

Goodwin, Barry K., Monte L. Vandeveer, and John L. Deal. "An empirical analysis of acreage effects of participation in the federal crop insurance program." *American Journal of Agricultural Economics* 86, no. 4 (2004): 1058-1077.

Grootveld, H., and W. G. Hallerbach. "Upgrading value-at-risk from diagnostic metric to decision variable: a wise thing to do? Pp, 33-50 in G, Szegö (Ed,) Risk Measures for the 21st Century." (2004).

Halcrow, Harold G. "Actuarial structures for crop insurance." *Journal of Farm Economics* 31, no. 3 (1949): 418-443.

Hess, Ulrich, J. R. Skees, Andrea Stoppa, B. J. Barnett, and John Nash. "Managing agricultural production risk: Innovations in developing countries." *Agriculture and Rural Development (ARD) Department Report* 32727-GLB (2005).

Hunter, David R., and Kenneth Lange. "Quantile regression via an MM algorithm." *Journal of Computational and Graphical Statistics* 9, no. 1 (2000): 60-77.

Jensen, Nathaniel, and Christopher Barrett. "Agricultural index insurance for development." *Applied Economic Perspectives and Policy* 39, no. 2 (2017): 199-219.

Khalil, Abedalrazq F., Hyun-Han Kwon, Upmanu Lall, Mario J. Miranda, and Jerry Skees. "El Niño–Southern Oscillation–based index insurance for floods: Statistical risk analyses and application to Peru." *Water Resources Research* 43, no. 10 (2007).

Koenker, Roger. *Quantile Regression*. Econometric Society Monographs. Cambridge: Cambridge University Press, 2005. doi:10.1017/CBO9780511754098.

Leblois, Antoine, Philippe Quirion, Agali Alhassane, and Seydou Traoré. "Weather index drought insurance: an ex ante evaluation for millet growers in Niger." *Environmental and Resource Economics* 57, no. 4 (2014): 527-551.

Maestro, Teresa, Barry J. Barnett, Keith H. Coble, Alberto Garrido, and María Bielza. "Drought Index Insurance for the Central Valley Project in California." *Applied Economic Perspectives and Policy* 38, no. 3 (2016): 521-545.

Markowitz, Harry. "Portfolio Selection, Cowles Foundation Monograph No. 16." *John Wiley, New York. S. Moss (1981). An Economic theory of Business Strategy, Halstead Press, New York. TH Naylor (1966). The theory of the firm: a comparison of marginal analysis and linear programming. Southern Economic Journal (January)* 32 (1959): 263-74.

McKee, Thomas B., Nolan J. Doesken, and John Kleist. "The relationship of drought frequency and duration to time scales." *Proceedings of the 8th Conference on Applied Climatology*. Vol. 17. No. 22. Boston, MA: American Meteorological Society, 1993.

McKee, Thomas B. "Drought monitoring with multiple time scales." *Proceedings of 9th Conference on Applied Climatology, Boston, 1995*. 1995.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. Departamento De Gestão De Riscos. *Relatório das indenizações pagas entre 2006 a 2015 – Programa de Subvenção ao Prêmio do Seguro Rural (PSR)*. 2015.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. Departamento De Gestão De Riscos. *Relatório Geral 2017 – Programa de Subvenção ao Prêmio do Seguro Rural (PSR)*. 2017.

Miranda, Mario J. "Area-yield crop insurance reconsidered." *American Journal of Agricultural Economics* 73, no. 2 (1991): 233-242.

Miranda, Mario J., and Katie Farrin. "Index insurance for developing countries." *Applied Economic Perspectives and Policy* 34, no. 3 (2012): 391-427.

Mishra, Ashok K., and Vijay P. Singh. "A review of drought concepts." *Journal of hydrology* 391, no. 1-2 (2010): 202-216.

Mishra, Pramod K. "Is rainfall insurance a new idea? Pioneering work revisited." *Economic and Political Weekly* (1995): A84-A88.

Ozaki, Vitor Augusto, and Ricardo Shirota. "Um estudo da viabilidade de um programa de seguro agrícola baseado em um índice de produtividade regional em Castro (PR)." *Revista de Economia e Sociologia Rural* 43, no. 3 (2005): 485-503.

Rao, Kolli N. "Weather Index Insurance: Is it the Right Model for Providing Insurance to Crops?" *ASCI Journal of Management* 41, no. 1 (2011): 86-101.

Rathore, Vandana, and M. J. Rao. "The performance of PMFBY and other crop insurance models in India." *Int. J. Adv. Res. Dev* 2 (2017): 2455-4030.

Rubin, Donald B. "The Bayesian bootstrap." *The annals of statistics* (1981): 130-134.

Schnitkey, Gary. "Area Risk Protection Insurance Policy: Comparison to Group Plans." (2014).

Skees, Jerry R., J. Roy Black, and Barry J. Barnett. "Designing and rating an area yield crop insurance contract." *American journal of agricultural economics* 79, no. 2 (1997): 430-438.

Skees, Jerry Robert. *Developing rainfall-based index insurance in Morocco.* Vol. 2577. World Bank Publications, 2001.

Skees, Jerry R., Jason Hartell, and Anne G. Murphy. "Using index-based risk transfer products to facilitate micro lending in Peru and Vietnam." *American Journal of Agricultural Economics* 89, no. 5 (2007): 1255-1261.

Staniswalis, Joan G. "The kernel estimate of a regression function in likelihood-based models." *Journal of the American Statistical Association* 84, no. 405 (1989): 276-283.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.

van Lier, Quirijn de Jong. "Water Availability to Plants." In *Application of Soil Physics in Environmental Analyses*, pp. 435-452. Springer, Cham, 2014.

Wang, Wentao, Shoufang Xu, and Tianshun Yan. "Structure identification and model selection in geographically weighted quantile regression models." *Spatial Statistics* (2018).

Weschenfelder, Adriana Burin, Karine Pickbrenner, and Eber José de Andrade PINTO. "Análise da frequência de ocorrência e a classificação das precipitações diárias máximas anuais na região da Laguna dos Patos (sub-bacia 87)." (2011).

APPENDIX

APPENDIX A. Crop insurance in Brazil

Crop insurance in Brazil is operated both by the government and the private sector. Proagro, a kind of public insurance, is a program linked to the contracting of rural credit administered by the Central Bank, not covered by the legislation and regulations pertaining to the insurance sector, although in practice it has insurance characteristics. This program aims to relieve the farmer from the payment of rural credit operations whose discharge has been rendered unfeasible by the occurrence of natural phenomena that have interfered in agricultural and livestock production. The classic insurance mechanisms are operated by private companies as insurers, reinsurers and brokers, in a competitive competitive market. The regulation and supervision of this sector is done by the CNSP - National Council of Private Insurance and SUSEP - Superintendence of Private Insurance.

Proagro was established through Law No. 5.969, on December 11, 1973, with the primary objective to exempt the farmer from financial obligations related to credit operations that were made impossible to settle by the occurrence of natural phenomena, pests and diseases affecting crops, livestock or herds. The program ensures the discharge of loans contracted via rural credit. The administration pertains to the Central Bank and covers up to 100% of the credit granted by a financial institution.

Later, on October 13, 2009, Proagro's structure was modified again, through Law No. 12.058. This amendment introduced the program named Proagro Mais, aimed for family farming. With this, Proagro serves both large-scale, business and family agriculture with up to four fiscal modules. The main objective of the program was also to exonerate producers from financial obligations related to rural credit, fully or partially, in the face of natural phenomena, pests or diseases that affect herds and plantations, as well as to indemnify the own resources invested producers in the event of claims. The difference in coverage between the two segments of producers is that Proagro covers the agricultural credit, while Proagro Mais covers agricultural costs and investment costs, in addition to the latter guaranteeing a minimum income. Proof of losses is made by the financial institution, based on a technical valuation report issued by a qualified professional.

Adhesion to Proagro is formalized when rural credit is contracted, considering the limits of funding for each of the crops. To participate in the program, among other requirements, farmers must comply with the agro-climatic risk zoning (ZARC) of the Ministry of Agriculture, Livestock and Food Supply and pay a fee called an additional contribution, fixed in percentage terms, on the total nominal value of the credit taken.

The private rural insurance history in Brazil began in 1954, when agricultural insurance was institutionalized, through Law No. 2.168, in order to protect harvests and herds. This law established that the Brazilian Reinsurance Institute (IRB), previously created by Decree-Law No. 1.186 of April 3, 1939, would operate as a reinsurer and retrocedent, organize and direct a consortium of insurers, establishing the scope and conditions of reinsurance operations, for each type of agricultural insurance. The same act exempted federal taxes and fees from insurance operations. The Decree-Law also established the crop insurance stability fund with the purpose of guaranteeing the stability of the operations, covering the additional coverage of the catastrophic risks, allowing the gradual adjustment of the premiums and contributing to the improvement of insurance. Since its inception, the fund has sought to use funds to reimburse the retrocessionaires of IRB-Brazil Re, the administrator of the fund's resources, with the amounts exceeding the maximum allowable losses for agricultural insurance operations.

In order to develop the agricultural insurance market, a joint stock company was created called National Agricultural Insurance Company (Companhia Nacional de Seguro Agrícola). Law No. 2.168 provided for the establishment of another stabilization fund that would be tied to the company, aiming at maintaining the level of premiums on a reasonable basis and attending to catastrophe. The regulation of Law No. 2.168 was made by Decree No. 35.370, dated April 12, 1954, establishing that the operation of agricultural insurance operations would be exercised by the National Agricultural Insurance Company or other private insurance companies governed under the provisions of Decree-Law No. 2.063, dated March 7, 1940, the legal framework for private insurance operations carried out by corporations, mutuals or cooperatives. Law No. 2.168 was succeeded by decrees defining the criteria, general conditions and prizes for specific agricultural activities, namely: cattle raising, wheat, vine, rice, herbaceous cotton and small crop of multiple crops, crop harvest and livestock of equidae. Thus, for thirteen years, until the dissolution of the National Agricultural Insurance Company, through the publication of Decree-Law No. 73, federal crop insurance operated as described.

The current legal framework for private insurance is still based on Decree-Law 739, when the National Private Insurance System was established, consisting of the National Council of Private Insurance (CNSP), Superintendence of Private Insurance (SUSEP), reinsurers, companies authorized to operate in private insurance and authorized brokers. Decree-Law No. 73 defined the roles and limits of action of each of the constituents, as well as of the CNSP itself, whose main competences are to establish the guidelines and norms of the private insurance policy, regulate the constitution, organization, operation and supervision of the parties subordinated to the Decree-Law, as well as the application of the penalties envisaged, to establish indexes and technical conditions on tariffs, investments and equity relations to insurance companies, lay down the general characteristics of insurance contracts, to establish the general norms of accounting and statistics for insurers, delimit the capital of insurers and reinsurers, establish the general guidelines for reinsurance operations, discipline co-insurance operations, to prescribe the criteria for the constitution of insurers, setting the legal and technical limits of insurance operations, to discipline insurance brokerage and the profession of broker, regulate the installation of insurance scholarships and to establish the conditions for the constitution and extinction of self-regulating entities of the brokerage market, as well as to regulate their administration and exercise of disciplinary power.

The CNSP is chaired by the Minister of Finance and has as members the Superintendent of SUSEP, representatives of the Ministries of Justice and Social Security, the Central Bank of Brazil and the Brazilian Securities and Exchange Commission. The deliberations of the CNSP on specific purposes must be taken by listening to the relevant advisory committee. Insurance for the rural sector, therefore, must pass through the Rural Consultative Committee of the National Council of Private Insurance.

SUSEP is an autarchic entity, jurisdiction of the Ministry of Finance, with administrative and financial autonomy. The mission of SUSEP, as executor of the policy emanating from the CNSP, is to oversee the constitution, organization, operation and operations of insurance companies. Specifically, according to article 36 of Decree-Law No. 73, SUSEP is responsible for processing applications for authorization, for constitution, operation, merger, expropriation, reverse split, transfer of share control and reform of the statutes of the insurers, to issue instructions and issue circulars concerning the regulation of insurance operations, to establish conditions of policies, plans of operations and tariffs to be used compulsorily by the insurance market, to approve the limits of operations of insurance companies, in accordance with the criterion established by CNSP, to examine and approve the conditions of special coverages, as well as to fix the applicable taxes, to authorize the movement and release of the assets and amounts insured as a guarantee of technical reserves and tied capital, to supervise the implementation of accounting and statistical standards established for insurers, to supervise the operations of the insurers regarding compliance with the legislation and apply the applicable penalties, to proceed with the liquidation of insurance companies that have canceled the authorization to operate in the country, and to organize their services and execute their budget.

The CNSP has a broader mandate, aiming at establishing the guidelines of the functioning structure of the private insurance market in Brazil. On the other hand, SUSEP deals with the operational actions of control and market surveillance, that is, the execution of policies, regulation and monitoring of the activities of insurance companies, reinsurers and brokers.

Decree No. 60.459, of March 13, 1967, regulated by Decree-Law No. 73, assigned important tasks to SUSEP, regarding the approval of new insurance contracts. According to article 8, insurers interested in marketing a new insurance contract should send it to SUSEP for analysis of the terms of the agreement, as well as the corresponding actuarial technical notes. The superintendency is empowered to request supplementary information, to determine changes, to promote the suspension of all or part of the conditions and technical actuarial notes, as well as to instruct insurers to include mandatory clauses in the general insurance conditions. It should be noted that the actuarial technical notes presented by the insurers should explain the pure premium, the loading, the interest rate and all other parameters related to the measurement of risk and aggregate costs.

The minimum premiums approved by SUSEP must be adopted by insurance companies for the purposes of calculating technical reserves and reinsurance, and the Superintendency may also approve actuarial technical notes for calculating provisions proposed by insurance companies for each specific case.

SUSEP is empowered to publish studies on benchmark premium rates calculated by scientific or insurance market entities in order to establish adequate actuarial bases for existing risk conditions and may require that such rates be used for the calculation of provisions techniques.

Through Decree-Law No. 73, in addition to the dissolution of the CNSA and its incorporation into the structure of the Ministry of Agriculture, the rural insurance stability fund (FESR) was created, incorporating the crop insurance stability fund and the stabilization fund, instituted in 1954, as described. The purpose of the FESR is to ensure the stability of rural insurance operations and to cover supplementary coverage of the risks of catastrophe. The resources are managed by IRB-Brasil Re and the proceeds come from the maximum allowable surplus as income from rural insurance operations and from credits made by the Federal Government to cover operational deficiency.

CNSP Resolution No. 46, of February 12, 2001, defines the FESR administration and control by its manager. The first point of this resolution is that by amendment introduced by CNSP Resolution 50 of September 3, 2001, the FESR will only guarantee the stability of operations of the agricultural modality that guarantees to the producer indemnification for damages caused to the insured crops, covering exclusively the expenses of direct costing of periodic crops and the budget of maintenance expenses of permanent. The FESR can also be used for livestock, aquaculture, forestry and rural lending for private financial institutions.

Insurers interested in operating the FESR must submit to the fund manager an operations plan with regions and cultures which will be included in the reinsurance program. The FESR guarantee is subject to the approval by SUSEP of the contractual conditions and technical actuarial note of the rural insurance modality for each year. Resolution No. 46 establishes that the approval of the actuarial technical note is conditioned to the presentation of the reinsurance coverage.

In order to stimulate private insurance contracting by rural producers, the federal government has created a policy in which percentages ranging from 35% to 45% of the insurance premium are paid by MAPA, depending on the crop or species to be insured. This program was established by Law No. 10.823 of December 19, 2003, and is regulated by Decree No. 5.121, of June 29, 2004.

For the management of the subsidy policy, the Interministerial Rural Insurance Steering Committee (CGSR) was created, which is responsible for approving and disseminating the percentages on the rural insurance premium and the maximum amounts of the economic subsidy, the specific operational conditions, the crops and animal species covered by the program, the regions to be covered by the intended benefit, the technical conditions to be met by the beneficiaries and the proposal of the Triennial Rural Insurance Plan or its annual adjustments, providing for the guidelines and conditions for granting the economic subsidy.

In addition, it is the responsibility of the committee to implement the benefit provided by law, encourage the creation of pilot projects by insurers, including new crops, animal species and types of coverage, establish guidelines and coordinate the development of methodologies and the dissemination of studies and statistical data to assist the development of rural insurance as an agricultural policy instrument and establish financial limits for the subsidy, per beneficiary and area unit.

The CGSR is composed by one member from the Ministry of Agriculture, Livestock and Supply who will preside, one from the Ministry of Finance, one from the Ministry of Planning, Budget and Management, one from the Ministry of Agrarian Development, one from SUSEP, one from the Agricultural Policy Secretariat of MAPA and one from the National Treasury Secretariat of the Ministry of Finance.

The PSR is under the legislation of the CNSP and the participation of the insurers is conditioned to the analysis and approval by SUSEP of its insurance products and the previous acceptance of the norms of the program, by registering with the Executive Secretariat of the CGSR. Decree 5.121 established the need to publish a Triennial Rural Insurance Plan (PTSR).

The PTSR includes general guidelines of the subsidy policy, specifying the regions, crops and animal species that are the object of the grant, the eligible insurance lines, as well as the definition of the risks covered, parameters and contractual provisions required, percentages and amounts of subsidy to the rural insurance premium, the financial limits per beneficiary or area unit, an estimate of the overall contribution of resources and the evolution of the financial flow during the years of validity and the dates of its validity, especially the deadline for settlement of financial obligations with the insurance companies, before the end of the MAPA exercise.

In practice, the operationalization of the PSR consists in MAPA paying a percentage of the premium to insurance companies that have carried out subsidized rural insurance operations, which are obliged to reduce the premium collected from the beneficiaries, for their insurance products, by an amount equal to the value of the grant. The coordination and monitoring of the application of the program resources is carried out by the CGSR.

A relevant obligation created under the program, which contributes to the reduction of accidents, is that the contracting of temporary crop insurance is carried out in accordance with the ZARC and, in its absence, with other agroclimatic zoning of official institutions at the discretion of the CGSR. The loss ratio tends to be reduced because the zoning must obligatorily consider probabilistic criteria in the delimitation of planting dates and crop risks. To be reimbursed, the insurers qualified in the PSR who have carried out operations that can be subsidized, send to the Executive Secretariat of the CGSR the information and documents required to prove the value of the subsidy, within the term and form contained in the program regulation. The financial obligations assumed by the Union, under the program, must be settled with insurers in the same rural insurance contracting exercise.

The modalities supported by the economic subsidy to the premium are agriculture, livestock, aquaculture and forestry. The risks covered are those approved by SUSEP, provided that within the said modalities which are beneficiary of the subsidy.

The rural producer may receive the subsidy up to the limit of R$ 72,000.00, for more than one crop, provided that the sum does not exceed the limit. The producers can receive the subsidy for the other modalities independently, with an extra limit of R$ 24,000.00 for each of the livestock, aquaculture and forest modalities. Cumulatively, the maximum eligible subsidy per producer, individual or legal entity, in compliance with the Union, is R$ 144,000.00.