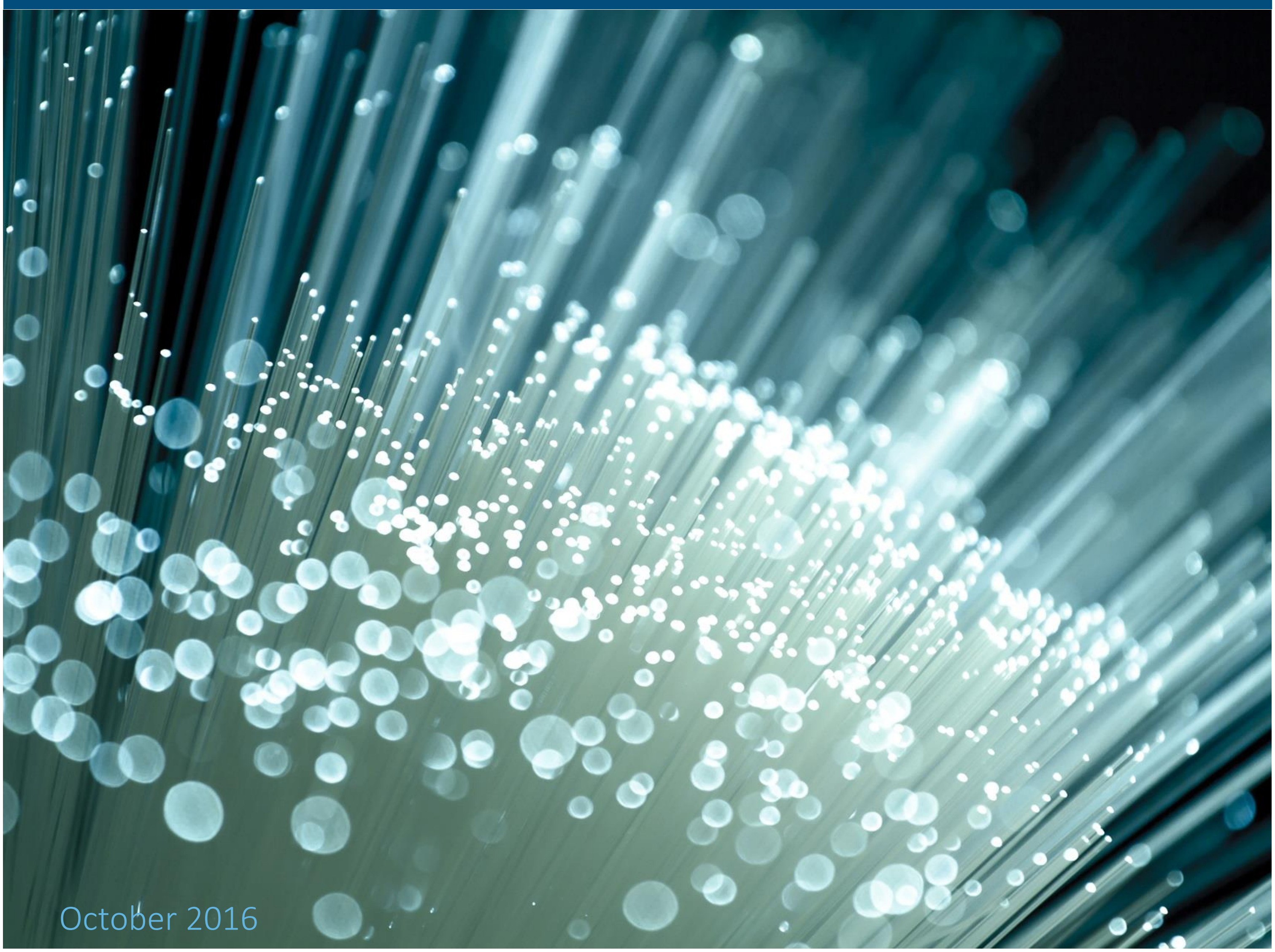
SOCIETY OF ACTUARIES

Farm-Level Crop Yield Forecasting in the Absence of Farm-Level Data



October 2016

Farm-Level Crop Yield Forecasting in the Absence of Farm-Level Data

SPONSOR	Society of Actuaries General Insurance Research Committee	**AUTHORS**	Lysa Porth, PhD, MBA Ken Seng Tan, PhD, ASA, CERA Wenjun Zhu, PhD

TABLE OF CONTENTS

Farm-Level Crop Yield Forecasting in the Absence of Farm-Level Data

Section 1: Acknowledgements

The authors gratefully acknowledge funding for this research by the Society of Actuaries (SOA) General Insurance Research Committee. In addition, the authors are thankful for the various datasets provided by Manitoba Agricultural Services Corporation (MASC), Environment and Climate Change Canada, National Agricultural Statistics Service (NASS), Ag-Analysis (Woodard 2016a; b; c), as well as one proprietary data set. The authors are also grateful to the SOA Project Oversight Committee for their valuable guidance and feedback throughout the project.

Section 2: Background and Scope

Central to designing and delivering an individual crop insurance program are the historical individual farm-level yields, which serve as the foundation for setting coverage levels and rates. However, yield experience at the farm level is often insufficient or unavailable, and there may also be concerns regarding data credibility, to be able to directly calculate an estimate of individual expected losses. This often leads to the use of aggregate county-level data to establish a base premium rate, and this practice has been heavily criticized as contributing to adverse selection and substantial program losses (Coble et al., 2013; Goodwin and Ker, 1998; Skees and Reed, 1986). Therefore, the objective of this research is to address this issue through developing a new relational model to predict farm-level crop yield distributions in the absence of farm-level yield data, in order to improve the accuracy of computing crop insurance premium rates. This may be useful in the scenario where we are interested in the farm-level yield distribution in a country (which we refer to as the base country) in which only county-level data is available, however, there exists more detailed farm-level yield experience from which we can borrow information (which we refer to as the reference country). This is the first paper to develop a relational model in order to relate farm-level yield experience between two separate countries in order to enhance the farm-level dataset to more accurately compute premiums.

Individual Yield Insurance Versus Area-Yield Insurance

Adverse selection and moral hazard are commonly cited for crop insurance market difficulties, and are part of the argument in favor of government subsidization of crop insurance programs (Miranda and Glauber, 1997). Between 1985 and 1993 Hoffman, Campbell and Cook (1994) reported that the loss ratio of the crop insurance program was 2.04, and this was partly attributed to adverse selection and moral hazard problems. These losses motivated critics of the program to recommend alternative program designs (Goodwin and Ker, 1998).

The two main types of crop insurance programs considered in this paper include individual farm-level insurance products and area-based insurance products. The performance of these two types of product designs are an important issue for crop insurance programs and has received a significant amount of research attention. Individual farm-level yield insurance is the most common type of insurance product, and pays an indemnity to a producer when the farmer's actual on-farm yield falls below the average historical expected yield. The payment is calculated as the shortfall in yield multiplied by a pre-determined price guarantee. In Canada, this type of insurance product is referred to as Multi-Peril Crop Insurance (MPCI), and in the U.S. it is called Actual Production History (APH) insurance. An alternative to farm-level yield insurance is area-yield insurance, where indemnities are based on shortfalls in the county average yield, rather than the individual farmer yield. The payment is calculated as the difference between the average historical county yield and the actual average county yield multiplied by a pre-determined price guarantee. In Canada there are currently only a few area-yield products offered, and in the U.S. there is a product known as the Group Risk Program (GRP).

Area-yield insurance was first offered in the U.S. in 1993, and has been encouraged by some over individual farm-level insurance as it is believed that this type of program can assess losses and determine premiums more accurately. This is because county-level yield data serves as the foundation for the program, and county-level data is much more plentiful compared to farm-level data. As a result, area-yield products may avoid some adverse selection and, therefore, reduce program losses. Area-yield products are also less susceptible to moral hazard since an individual farmer is not large enough to influence the county yield average through moral hazard actions. Other cited benefits of area-yield products are comparably lower transaction costs because individual farm-level yields do not need to be established and verified, and there is no need for on-farm loss adjustment. As a result, some literature supports area-yield insurance products as they are thought to provide farmers with a valuable, and less expensive, alternative to farm-level insurance products (Awondo, 2012; Skees, Black, and Barnett, 1997; Baquet and Skees, 1994, Barnett et al., 2005). The main issue with area-yield insurance, however, is basis risk. This refers to the imperfect correlation between the county-level yields and farm-level yields, and some researchers believe that this basis risk makes this type of insurance unattractive to many farmers (Deng et al., 2007). In 2005, area-based products accounted for only 9% of the liability of the Federal Crop Insurance Program (FCIP) in the U.S., and in Canada they would similarly account for a very small portion of products.

Data Scarcity

As mentioned above, data scarcity is one of the motivations for developing area-yield insurance products over farm-level insurance products. There are many contributing factors to shortness of data in crop insurance, and they are discussed in this section. First, in most countries, such as Canada the U.S., there is only one growing season per year, and hence only one yield observation per year. Further, due to crop rotation and other market forces farmers do not grow the same crop each year, and this leaves fewer observations or an inconsistent time series at the farm-level. In addition to limited data, there can also be concerns over the credibility of data. This can be attributed to data from the past that may not be representative of the current environment, due to such factors as changes in technology, biotechnology, farming practices, coverage levels, and other program details (Coble et al., 2013; Porth et al., 2014). The result is a need to balance using as much of the time series as possible so that those infrequent, but, extreme weather events are considered, versus the concern that older data is not representative and, therefore, should be discarded. In order to help balance this incongruity, actuaries attempt to “restate” historical yield or loss experience to bring it on level with the current environment.

In countries with developed crop insurance programs, data is often more substantial and of higher quality. In practice, this means that there is often as many as 40-45 years of historical yield or loss data at the aggregate level. However, at the farm-level data is still quite limited in developed regions with approximately 5-10 years of historical data for a particular crop type, and most often there are gaps in the time series. In countries with developing crop insurance programs, the data challenges can be even more pronounced. This means that farm-level data may be limited to only a few years, or in some cases possibly no historical farm-level yield records. As a result, insufficient farm-level yield data can be one of the major challenges to the successful implementation of crop insurance programs (Borman et al., 2013; Ker and Goodman, 2000; Porth et al., 2014). As an alternative to farm-level yield data, aggregate yield experience at the county level may be available. Since county-level data typically represents several producers' yield experience together in a region, there are often less challenges regarding missing data and insufficient time series.

Aggregation Bias

The concern with using aggregated county-level data in place of farm-level data, however, is that aggregating data may lead to the possible cancellation of idiosyncratic risk, and hence there may be smaller total risk in the aggregated data. The result may be an underestimation of risk in the aggregate county-level data relative to the farm-level data (Awondo et al., 2012). In literature, this occurrence is often referred to as "aggregation bias," and this is known to decrease the correlation between county-level and farm-level yield. For example, Claassen and Just (2009) calculate that farm-level yield variation is understated by 50% when using county-level averages, and 61% of systemic variation and 42% of random variation is lost when yields are aggregated to the county-level. As a result, this increases basis risk and has been shown to make area-yield insurance unreliable for farmers.

Literature: Linking County-Level and Farm-level Data

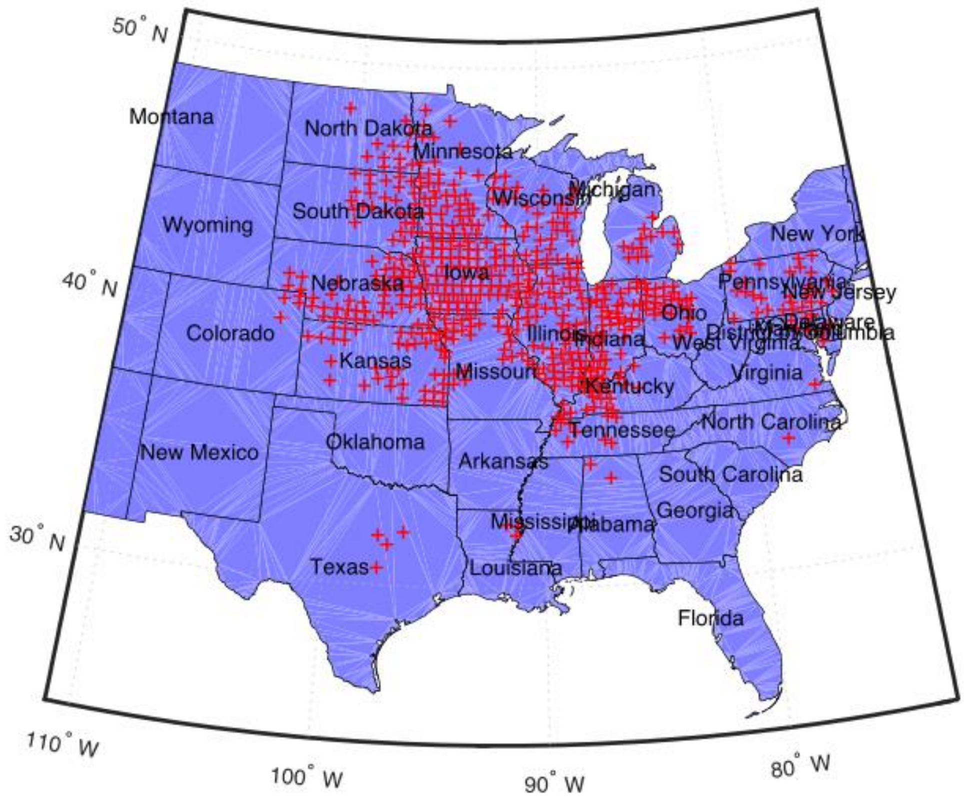
Recognizing that farm-level yield data are likely sparse and may not be representative over time, some research has been conducted to link county-level data and farm-level data in order to enhance the data and provide a proxy for estimating the individual farm-level loss distribution. For instance, Miranda (1991) relates individual farm level yields to the county-level yields via a linear regression model. This framework, which strongly resembles the well-known Capital Asset Pricing Model (CAPM), produces a "beta" coefficient that measures the sensitivity of the farm-level yields relative to the county-level yields. The decomposition of the farm-level yield variations into a systematic component and an idiosyncratic component in turns facilitates the analysis of risk reduction between farm-level yields and county-level yields. Ramaswami and Roe (2004) further establish that this model is valid if systemic and individual risks are additive in individual yields, and if the aggregation is such that the law of large numbers holds.

Extending Miranda (1991), Coble and Barnett (2008) investigate the systemic risk and idiosyncratic risk in farm-level data by allowing the farm-specific beta to vary over an assumed normal distribution. Deng et al. (2007) assume that yields at both farm and county levels have a multiplicative relationship, and provide a simple way of generating additional pseudo farm-level yields from the county-level data. Cooper et al. (2012) generate the farm-level yield distribution that explicitly incorporates the within county yield heterogeneity while accounting for systemic risk and other spatial or intertemporal correlations among farms within the county. They show that individualizing premiums for a farm from its county yield results in substantial mispricing of crop insurance premiums because they do not adequately capture farm level yield variability and yield correlations between farms.

To date, the methods that have been proposed focus on enhancing farm-level yield prediction using aggregate experience at the county-level within the same region and same country. These methods assume that at least some farm-level historical crop yield data exists in the same country to serve as the underlying basis of prediction. Therefore, this paper contributes to the literature by developing a new relational model to predict farm-level crop yield distributions in the absence of farm-level yield data. Our proposed relational model is comprised of two steps. The first step involves determining the optimal county in the reference country, from which the relevant information is borrowed from. The second step is to provide an explicit linkage/relationship between the available data in order to facilitate the prediction of the farm-level yields in the base country. The methodology for the relational model is explained in greater detail in Section 4. Section 4.1 outlines the optimal selection of the reference county via a certain distance metric. Section 4.2 predicts the farm-level yields using moment matching and assuming that both countries have the same dependence structures between the farm-level and county-level yields in order to predict the farm-level yield in the base country.

In this paper a proprietary data set is used, which includes both farm-level and county-level corn production experience from the U.S. and Canada. It is assumed that only the county-level and farm-level data in the U.S. (reference country) and the county-level data in Canada (base country) are used. The geographical locations of the reference counties are displayed in the map of Figure 1. In order to demonstrate the viability of the relational model the farm-level yields in Canada, which are available but not utilized, are used for back-testing and the out-of-sample prediction error is compared. Further, the method outlined by Deng et al. (2007) is generalized to predict the farm-level yields in the base country in order to provide additional benchmark against our proposed relational model. Also, to back-test our proposed relational model against the generalized model of Deng et al. (2007) and empirically observed data, we validate our model by testing the moments assumption as well as its application to pricing MPCI contracts.

Figure 1:
Geographical Locations of Reference Counties in the U.S.



The empirical results demonstrate that the relational model developed in this paper is useful in predicting the farm-level yield distribution in the base country in the absence of farm-level yields, through borrowing information from a reference country that has more detailed information. The relational model achieves lower mean (standard deviation) prediction error of 28.85% (70.97%) compared to the benchmark model of 40.85% (277.65%), and the resulting actuarial premium rates are found to recover the premium rates based on the actual loss experience more closely compared to using only county-level data. The results also confirm the existence of aggregation bias and caution using only county-level yield data to approximate farm-level premium rates, as the degree of underestimation can be as high as 40% in some cases.

The results of this research may be useful for government as well as insurers and reinsurers looking to improve farm-level yield prediction in order to set coverage levels and determine premium rates more accurately. The relational model proposed in this paper may also help to overcome aggregation bias, which may lead to misleading and erroneous pricing results. Further, the relational model may be useful for developing more flexible crop insurance programs that are capable of responding to changing farming conditions/practices more quickly, where lack of data would normally hinder the development of a new insurance product. An example may include a scenario where a certain crop variety is grown in a new region where it has not grown before, due to biotechnology advancements, etc. Another example may include the introduction of a new crop variety for the first time in a country, for which there is no historical yield information. The relational model proposed in this study may be useful in these situations in order to enhance the farm-level yield information through borrowing more detailed information from a reference country. The result may be that insurance coverage in these situations may be available to producers much sooner compared to current practice. This is because it is often necessary to collect sufficient yield or loss experience for as many as five to ten years in order to develop new insurance products and compute the actuarially fair premium rate.

Section 3: An Overview of the U.S. and Canadian Crop Insurance Programs

Crop Insurance Framework in the U.S.

The U.S. Federal Crop Insurance Program (FCIP) was established in 1938, and continues to be the biggest in the world. In 2012 the Federal Crop Insurance Corporation (FCIC) reported total premiums of US$11.7 billion with an insured value of US$117 billion (Shields, 2013). The FCIP is the single largest agricultural program delivered by the Farm Bill, with costs averaging approximately US$9 billion per year (Shields, 2015). There are a relatively large number of insurance products available to producers to cover various risks, however, in general, revenue-based products are the most popular and account for 77% of policies sold, while yield-based products account for 23%. Participation rates are high with approximately 83% of all U.S. crop acreage insured, and corn, cotton, soybeans and wheat are the four largest crops and account for more than 70% of total enrolled acres.

The Standard Reinsurance Agreement (SRA) governs the relationship between the FCIC and the authorized private insurance companies that sell and service the crop insurance policies. In addition, independent insurance agents are paid sales commissions by the insurance companies. The SRA also establishes the terms and conditions for which the FCIC provides subsidy and risk sharing agreements with the approved insurance companies. On average about 62% of the total premium is subsidized by government, and the full cost of selling and servicing the policies is also covered by government. All rates are set by the Risk Management Agency (RMA) of the United States Department of Agriculture (USDA), and unlike other types of insurance where the insurer can deny coverage to an individual, the approved insurance companies selling crop insurance must insure all eligible farmers. Therefore, to help encourage the involvement of the private sector insurance companies, insurance losses are reinsured by the USDA. The risk transferred to the USDA is determined by a fund allocation process, allowing insurers to allocate policies to either the “assigned risk pool” or the “commercial risk pool” based on the level of risk they want to accept. The assigned risk pool reflects riskier policies, and the government takes more responsibility for these policies. Conversely, the insurer would seek to allocate policies they deem less risky to the commercial risk pool. In addition to risk-sharing with the federal government, insurance companies may also choose to purchase reinsurance from the private market. This agricultural insurance framework is intended to bring efficiencies of a private sector delivery system and regulatory and financial strength from the Federal government.

Crop Insurance Framework in Canada

The Canadian agricultural insurance program, AgriInsurance, also has a long history and dates back to its establishment in 1959 under the Crop Insurance Act, and legislated under the Farm Income Protection Act (FIPA). The program is the third largest in the world according to premium volume, following the U.S. and China. In 2012, 254,000 policies were sold covering 70 million acres. In this same year, premium was CD$1.7 billion with corresponding liability of CD$17.3 billion, and indemnities were CD$1.3 billion (AAFC, 2016). As an example of scale of loss of the Canadian program, in 2001 and 2002 droughts over the Prairies and parts of Ontario and Quebec resulted in production losses exceeding $4 billion (AAFC, 2012). The majority of insurance products sold in Canada are yield-based, providing protection against yield shortfall and quality loss. Participation rates vary across provinces, with approximately 72% of crop acres insured on average, and in some provinces this is as high as 90% of crop acres.

The crop insurance program in Canada has a unique multi-layered risk-sharing framework that differs substantially from the U.S. AgriInsurance plans are developed and delivered by government insurance corporations in each of the ten agricultural producing provinces in Canada. The role of the federal government is to provide program oversight, ensuring that the obligations under the Farm Income Protection Act, the Canada Production Insurance Regulations and the Federal-Provincial Territorial Framework Agreement are followed (AAFC, 2016). Each of the ten provincial crop insurance corporations set their own rates and underwrite all corresponding liability for their program, and expected yields and rates are required to be certified by an actuary every 5 years. The federal government does not share in the direct liability of AgriInsurance, however, premium and administration cost subsidy is provided. Provincial and federal governments combined cover 60% of the premium, with 24% paid for by the provincial government and 36% paid for by the federal government. The remaining 40% of premium is paid for by the producer. In addition, crop insurance delivery costs are 100% subsidized by provincial and federal governments proportionately. The Canadian crop insurance framework has proven to be one of the most cost efficient in the world, with average administration costs of approximately 8%.

In addition, to help manage crop insurance losses each province can elect to participate in a deficit financing program, called the Federal-Provincial Reinsurance Fund. The national reinsurance fund is backstopped by the federal government and provincial reinsurance funds are back-stopped by participating provincial governments. In the event that the balance in the provincial insurance fund is insufficient, the provincial government insurance corporation may access the reserves in the “federal reinsurance fund” and in their “provincial reinsurance fund.” Reinsurance funds are reimbursed through the redirection of a portion of regular program premiums depending on the health of the surplus. In the scenario where a “reinsurance fund” is also exhausted, the federal and the provincial governments agree to loan their reinsurance funds respectively the amount necessary to meet their obligations. This loan is provided at an interest free rate, however, it must be repaid in future years through negotiated arrangements. The Federal-Provincial Reinsurance program is then cost neutral to the AgriInsurance program. Currently five of the ten agricultural provinces participate in this program. In addition to the Federal-Provincial Reinsurance Fund, individual provinces may also purchase reinsurance from the private market.

Section 4: Relational Model

In this section, the relational model developed in this paper is described. The first step of the proposed relational model provides an optimal way of selecting a county in the reference country, from which more detailed information is borrowed. The second step establishes the linkage between the county-level data in the base country (i.e. Canada) and the farm and county-level data in the reference country (i.e. U.S.). This allows the prediction of the farm-level yield distribution in the base country. The two steps are discussed in the following two sub-sections.

4.1 Step 1: Selecting the Optimal County in the Reference Country

The effectiveness of the relational model highly depends on the choice of the county that is used to borrow more detailed information from. Intuitively it is desirable to choose a county from the reference country that is as similar as possible to the county in the base country where the farm is located. By similarity we refer to agricultural production that is alike in both countries. Since agricultural production of a farm/county depends on a number of variables, such as the weather conditions, farm and county size, etc., one plausible way of quantifying the similarity is via these variables. Formally we use a similarity measure based on an Euclidean metric to select the optimal county in the reference country. While there are likely many factors that impact agricultural production, in this study the following 20 variables are considered, with each given equal weight:

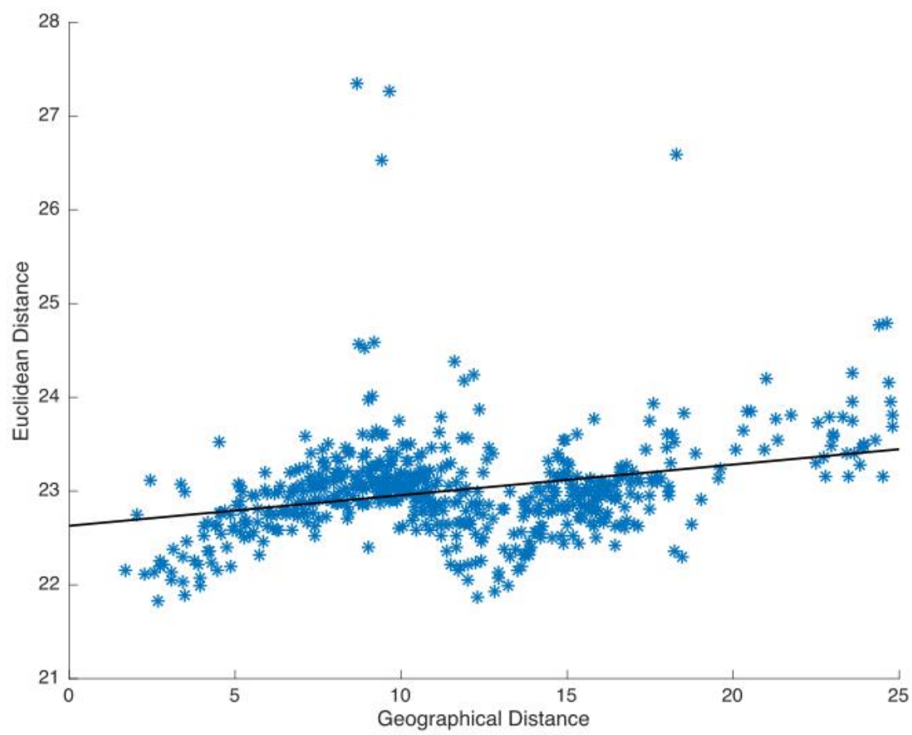
- County size (*variable 1*).
- County-level yield history (*variable 2*)
- Average farm size (*variable 3*).
- The county's coefficient of variation, which is defined as the ratio of standard deviation over mean (*variable 4*).
- Growing season (i.e. April – November) monthly average temperature (*variables 5-12*).
- Growing season (i.e. April – November) monthly cumulative precipitation (*variables 13-20*).

Empirical evidence shows that temperature and precipitation during the growing season are important for determining corn yields. Given that the growing season in Canada and the U.S. roughly starts in April and harvest is in November, the monthly mean temperature and cumulative precipitation from April to November are considered as variables. Together with the average farm size, county size, and the county coefficient of variation, there are 20 variables in total that are used to determine the “closeness” of the two counties via the normalized Euclidean distance. The optimal county is the one that has the lowest Euclidean distance.

As an illustration, suppose we are interested in predicting the farm-level yield in the county called Dufferin, which is located in the province of Manitoba in the base country Canada. We are interested in finding the optimal county in the U.S., from which we can borrow more detailed information, including farm and county-level data. This entails calculating the Euclidean distance over all counties in the U.S. based on the selected variables. The Euclidean distance approach adopted in this study does not explicitly consider the geographical location of the county in the reference country relative to the base country in order to allow for the possibility that the most similar information may belong to a county that is located a considerable geographical distance away from the farm being predicted. However, it is interesting to examine the correlation between the geographical distance and Euclidean distance. Figure 2 plots the results of this calculation with the x- and y-axis, denoting, respectively the geographical distance and Euclidean distance. The correlation between the two variables is 0.3 (p-value: 0.0000), meaning that the Euclidean distance calculated from the selected variables are statistically significant with the geographical locations of the reference county and base county. However, it is important to note that this is not always the case, as there are situations where the selected county is a considerable geographical distance from the base county. This is because the proposed Euclidean distance considers several variables to optimally select the reference county, including county and farm information, as well as weather information.

Figure 2:

Geographical Distance & Euclidean Distance



4.2 Step 2: Linking the Base Country Data and the Reference Country Data

Once the optimal county in the reference country is identified, the remaining task is to explicitly relate the available data between the two countries so that the farm-level yields in the base country, in turn, can be predicted. To do this, we need to impose additional assumptions. Motivated by the desire to have a simple and straight-forward approach, the first assumption is that the joint distribution of the county-level yield and farm-level yield for the base country has a bivariate lognormal distribution. The lognormal distribution is a common distribution for modelling crop yield (Deng et al. 2007), however, other bivariate distributions could be integrated into the relational model, provided the parameters of the distribution can be calibrated. The advantage of the lognormal assumption is that once the mean vector and the variance-covariance matrix are determined, the bivariate lognormal distribution is completely specified so that the calibrated distribution can be used to simulate the respective farm-level yields. The mean and variance of the log-yields at the county-level are readily estimated from the available base country data. The remaining parameters are calibrated by making the following assumptions:

1. The means of the farm-level log-yields in the base country are identical within the county and equal to the average of the means of all the farm-level log-yields within the optimal county selected in Step 1 in the reference country.
2. The variances of the farm-level log-yields in the base country are identical within the county and equal to the average of the variances of all the farm-level log-yields within the optimal county selected in Step 1 in the reference country.
3. The correlation between the county-level log-yields and the farm-level log-yields for both the reference and base countries are identical.

With these assumptions, the farm-level yields can be readily simulated in the base country. It should be noted, however, that the predicted farm-level yields from the base county are the same regardless of which farm we are attempting to predict, namely we have one farm-level prediction for each county. In the future, research could focus on generalizing these assumptions by imposing some functional relationships between farm-level yields and county-level yields.

Section 5: Model Validation and Robustness

In this section, additional sensitivity analysis is conducted on the relational model in order to better understand the accuracy of the farm-level yield prediction, as well as any limitations of the proposed approach. The relational model imposes a few assumptions to facilitate the prediction of the farm-level yield distribution, therefore, it is useful to back-test the model. The countries of Canada and the U.S. are chosen to empirically test the relational model because we are fortunate to have both farm and county-level yields from both countries. As discussed in the preceding sections, the farm and county-level data in the reference country (i.e. U.S.), and the county-level data in the base country (i.e. Canada) are used to develop the relational model. The farm-level yields in Canada are available, but, not utilized in implementing the proposed model. As a result, the availability of the farm-level yields in the base country, even though limited and incomplete, serve as a basis for evaluating the relational model in terms of the actual prediction errors. We compare the prediction errors for each of the 18 counties considered in the province of Manitoba in the base country Canada. The actual farm-level yields cover the period from 1996 to 2011. As there are many farms within a county, the prediction errors for each county are quantified based on the root-mean-square-error (RMSE) metric.

Table 1 displays the RMSEs for estimating the mean and standard deviation (S.D.) of the farm-level yields. In addition to computing the RMSEs for the proposed relational model, we also tabulate the results using a "benchmark model." The benchmark model is a generalization of Deng et al. (2007) so that it can be used to predict farm-level yields in the base country, even though the farm-level data within the county of the base country does not exist. For this reason, the results of this method are referred to as the benchmark farm yield prediction model.

The RMSEs of the benchmark model over 18 counties range from 22.93% to 94.38% for mean estimation, and 17.67% to 1,785% for standard deviation estimation. The corresponding RMSEs (lowest, highest) for the proposed relational model are (9.54%, 66.64%) and (24.42%, 176.81%). In terms of the effectiveness of the relational model and the benchmark farm yield prediction model for estimating mean and standard deviation, both methods have difficulty in accurately estimating the standard deviation. This is particularly true for the benchmark model, which could lead to an unacceptably high RMSE of 1,785% (for county “Thompson”). The performance of the two models for mean estimation appear more encouraging. In particular, the relational model is capable of achieving RMSE of less than 10%, which compares to the lowest RMSE of the benchmark model of close to 23%. Further, comparing the benchmark and relational model on a county-by-county basis, the relational model produces better prediction results in all but two counties (Dufferin and Labroquerie). Regardless of whether the models are evaluated based on mean or standard deviation, on average the relational model also outperforms the benchmark model. For the relational model, the average RMSEs across all 18 counties are 28.85% and 70.97% for mean and standard deviation estimations, respectively. In contrast, the respective averages for the benchmark model are 40.85% and 277.65%.

Table 1:

County Name	Relational Model		Benchmark Model	
	Mean	S.D.	Mean	S.D.
BROKENHEAD	9.57%	62.94%	31.20%	75.59%
DESALABERRY	21.49%	24.42%	37.11%	43.87%
DUFFERIN	66.64%	176.81%	40.98%	91.65%
GREY	25.20%	55.80%	37.57%	59.99%
HANOVER	37.12%	37.40%	38.92%	39.30%
LABROQUERIE	23.65%	40.21%	22.93%	46.51%
MONTCALM	38.40%	117.34%	42.91%	180.57%
NORTHNORFOLK	31.17%	55.44%	94.38%	1237.00%
PEMBINA	24.08%	37.08%	29.88%	93.18%
PORTAGELAPRAIRIE	22.19%	56.03%	28.35%	82.18%
RHINELAND	38.51%	143.89%	47.68%	97.18%
ROLAND	41.06%	55.63%	41.25%	95.21%
STEANNE	25.53%	36.18%	49.62%	17.67%
SOUTHNORFOLK	19.80%	33.29%	36.34%	821.44%
STANLEY	32.22%	86.39%	41.44%	111.48%
TACHE	18.21%	66.01%	34.24%	54.93%
THOMPSON	30.93%	130.05%	44.31%	1785.00%
WHITEMOUTH	13.44%	62.47%	36.20%	64.98%
Avg	28.85%	70.97%	40.85%	277.65%
Min	9.57%	24.42%	22.93%	17.67%
Max	66.64%	176.81%	94.38%	1785.00%

Section 5.1 focuses on the moments equality assumptions, and what effect there would be if the assumptions do not hold, but, instead differ by a proportional constant. Section 5.2 examines the importance of Step 1 in the relational model regarding the selection of the optimal county in the reference country.

5.1: Robustness analysis on moments assumptions

The relational model that we have implemented so far assumes that the means and the variances across all farm-level log-yields within the base country are of the same magnitude and are equal, respectively, to the average of the means and variances of all farms within the optimal county in the reference country. The objective of this subsection is to provide additional sensitivity analysis of the proposed relational model regarding the moments equality assumptions if they do not hold, but instead differ by a proportional constant. By denoting α and β as the proportional constant for mean and standard deviation, respectively, it is possible to determine the optimal values of α and β that minimize the expected RMSE for the selected county.

Table 2 compares the prediction results based on the relational model developed in the proceeding section, referred to as the "standard relational method." In addition, results based on using the optimal values of α and β are considered, referred to as the "optimal relational method." When $\alpha = 1$ and $\beta = 1$, the underlying method reduces to the standard relational method. In addition to reporting the RMSEs of these two methods, Table 2 also reports the improvement factor (in percentage). A positive relative improvement factor indicates a gain in efficiency (in terms of smaller RMSE) for predicting the farm-level yields in the base country using the optimal relational model, as opposed to the standard relational model.

From Table 2, it is of interest to note that the optimal α and β are close to 1. Furthermore, the averages of the optimal α and β across all 18 counties in the base country are 0.9813 and 0.9877, respectively. This finding suggests that the moments equality assumptions are reasonable for the data utilized in this study. In terms of their performances, the optimal relational model has lower RMSEs for predicting the means, with the relative gain in efficiency ranging from 0.32% to 78.46%, and on average the gain is 22.73%. The prediction results for the standard deviation, on the other hand, are mixed. The best improvement factor is as high as 66.55%, but, at the other extreme results can be worse than the standard relational model by a factor of 117.25%. Consistent with our results from the standard relational models and the benchmark models, these results confirm that it is, in general, much more challenging to estimate the standard deviation, as compared to the mean.

Table 2:
The impact of α and β on the relational models: standard v.s. optimal implementations. RMSEs (in percentage) for mean and standard deviation (S.D.) estimations and the resulting percentage of relative improvement across 18 counties in the base country are reported.

County Name	Standard Relational Model		Optimal Relational Model				Relative Improvement	
	Mean (% RMSE)	S.D. (% RMSE)	Optimal α	Optimal β	Mean (% RMSE)	S.D. (% RMSE)	Mean (%)	S.D. (%)
Brokenhead	9.57	62.94	0.9928	0.9522	8.74	60.44	8.67	3.97
Desalaberry	21.49	24.42	1.0004	1.0008	19.51	24.46	9.21	-0.16
Dufferin	66.64	176.81	0.8815	0.3821	20.85	60.45	68.71	65.81
Grey	25.20	55.80	1.0225	1.0164	24.76	58.69	1.75	-5.18
Hanover	37.12	37.40	0.9915	1.0099	37.00	37.29	0.32	0.29
Labroquerie	23.65	40.21	0.9539	0.9539	22.78	43.99	3.68	-9.40
Montcalm	38.40	117.34	1.1083	1.1260	8.27	254.92	78.46	-117.25
North Norfolk	31.17	55.44	1.0088	1.0024	30.48	73.01	2.21	-31.69
Pembina	24.08	37.08	1.0502	0.9605	14.52	57.01	39.70	-53.75
Portagelaprairie	22.19	56.03	1.0395	1.0337	16.30	83.94	26.54	-49.81
Rhineland	38.51	143.89	0.8267	0.8305	30.84	48.13	19.92	66.55
Roland	41.06	55.63	0.8031	1.2826	34.71	39.38	15.47	29.21
Steanne	25.53	36.18	1.0721	0.9174	8.52	25.55	66.63	29.38
South Norfolk	19.80	33.29	1.0330	1.0347	15.55	52.20	21.46	-56.80
Stanley	32.22	86.39	0.8837	1.2820	23.47	94.93	27.16	-9.89
Tache	18.21	66.01	0.9998	1.0001	17.88	63.27	1.81	4.15
Thompson	30.93	130.05	1.0004	1.0008	26.06	143.02	15.75	-9.97
Whitemouth	13.44	62.47	0.9957	0.9920	13.22	63.58	1.64	-1.78
Average	28.85	70.97	0.9813	0.9877	20.75	71.35	22.73	-8.13
Minimum	9.57	24.42	0.8031	0.3821	8.27	24.46	0.32	-117.25
Maximum	66.64	176.81	1.1083	1.2826	37.00	254.92	78.46	66.55

5.2 Sensitivity analysis in term of county selection

This sub-section examines the importance of the optimal selection of the county in the reference country on the proposed relational model. Recall that in Step 1 the optimal county selected in the reference country is the one that gives the lowest Euclidean distance. However, it is of interest to better understand how important the choice of the optimal county in the reference country is on the overall effectiveness of the proposed relational model. In particular, rather than using the optimal county in the reference country, this sub-section examines the results in the scenario where a county in the reference country is chosen at random instead. Table 3 provides some insights to this issue.

In Table 3, two sets of results are produced based on the standard implementation of the relational model using the optimal county in the reference country selected by minimizing the Euclidean distance, as well as the results displayed in the previous two sub-sections. The second set of results is obtained by randomly selecting a county from reference country for predicting the farm-level yield distribution in the base country. For each county in the province of Manitoba in the base country we repeat this procedure 100 times and the average RMSEs are reported in columns 4-5 of Table 3. For ease of comparison, the final two columns of the same table quantify the magnitude of deterioration in using the county selected at random in the reference country. All the values in the final two columns are positive, and these results highlight the critical role of selecting an optimal county in the reference country, rather than a random county, on the performance of the relational model. This provides strong evidence in favor of implementing the relational model with Euclidean distance optimality. In fact, randomly selecting a county in the reference country could penalize the model by more than 1,000%, as shown in the mean estimation for the county Brokenhead. On average, the gains in efficiency by selecting the optimal county in the reference country, as opposed to randomly selecting the county, are 225.14% and 102.71% for mean and standard deviation predictions, respectively.

Table 3:
The impact of county selection in the reference country on the relational models: County selected via Euclidean distance minimization, or randomly. RMSEs (in percentage) for mean and standard deviation estimations, and the resulting percentage of relative deterioration across 18 counties are reported. For the random approach, the reported RMSE for each county is the average of 100 randomly selected counties in the reference country.

	Relational Method with County Selected via					
	Euclidean Distance		Random Approach		Relative Deterioration	
County Name	Mean	S.D.	Mean	S.D.	Mean	S.D.
	(% RMSE)		(% RMSE)		(%)	
Brokenhead	9.57	62.94	105.50	128.48	1,002.40	104.13
Desalaberry	21.49	24.42	64.56	80.08	200.42	227.93
Dufferin	66.64	176.81	73.30	179.10	9.99	1.30
Grey	25.20	55.80	93.86	124.45	272.46	123.03
Hanover	37.12	37.40	104.57	80.46	181.71	115.13
Labroquerie	23.65	40.21	123.37	101.44	421.65	152.28
Montcalm	38.40	117.34	39.86	163.48	3.80	39.32
North Norfolk	31.17	55.44	114.22	137.44	266.44	147.91
Pembina	24.08	37.08	48.22	83.10	100.25	124.11
Portage la prairie	22.19	56.03	80.54	125.87	262.96	124.65
Rhineland	38.51	143.89	48.97	158.06	27.16	9.85
Roland	41.06	55.63	42.76	123.65	4.14	122.27
Steanne	25.53	36.18	49.56	70.60	94.12	95.14
South Norfolk	19.80	33.29	82.59	109.66	317.12	229.41
Stanley	32.22	86.39	47.64	127.35	47.86	47.41
Tache	18.21	66.01	76.02	108.63	317.46	64.57
Thompson	30.93	130.05	59.49	218.26	92.34	67.83
Whitemouth	13.44	62.47	71.26	95.26	430.21	52.49
Average	28.85	70.97	73.68	120.85	225.14	102.71
Minimum	9.57	24.42	39.86	70.60	3.80	1.30
Maximum	66.64	176.81	123.37	218.26	1002.40	229.41

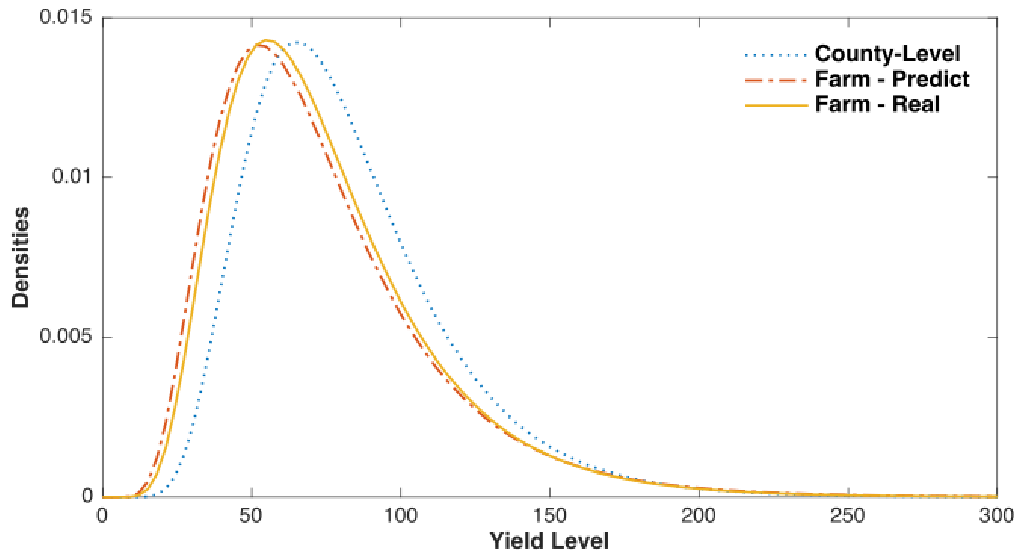
Section 6: Applications to Agricultural Insurance Pricing

The ultimate objective of imputing the farm-level crop yield distribution is to facilitate pricing agricultural insurance contracts, including the individual farm-level yield product (referred to as the MPCI contract). The premium rate of an agricultural insurance contract is often calculated based on the loss-cost ratio (LCR) random variable, which is defined as the ratio of indemnity over liability. The liability, which is known at the inception of the contract, measures the maximum exposure, or the maximum payout, of the agricultural insurance contract based on an expected farm-level yield and an expected commodity price, also considering the coverage (as a percentage) selected by the producer. If the actual farm-level yield for the year falls short of the pre-determined expected yield (i.e. average historical expected yield), then the agricultural insurance contract is triggered. This yield shortfall multiplied by the guaranteed commodity price and the selected coverage ratio determines the indemnity paid to the farmer.

The premium of the above agricultural insurance product depends on the premium principle adopted by the insurance company. The expectation and the standard deviation premium principles are two of the most common premium principles utilized in the insurance industry. Therefore, for the purpose of our pricing illustration both the expectation and standard deviation premium principles are considered. In our numerical examples, the following parameter values are considered: the loading parameter of the premium principle, and the coverage ratio of the insurance policy. These values are consistent with market practices; see for example, Porth et al. (2013, 2014). For each premium principle, the premium rates are calculated based on the following three methods: the historical county-level yields, the predicted farm-level yields (using the standard relational model), and the historical farm-level yields. The first method is referred to as the "county model," the second and third methods are referred to as the "predicted farm model" and "real farm model," respectively. The yield densities estimated using these three approaches based on the data from the county Brokenhand in the province of Manitoba in the base country are depicted in Figure 3.

Visually it can be concluded that the farm-level yield distributions, based on both "farm models" are very close to each other. On the other hand, the distribution estimated from the "county model" deviates substantially from the other two farm-level yield distributions in some important ways. Notably, the county model is located on the right side of the farm models. This implies that using the estimated county yield distribution to approximate the farm yield distribution could underestimate the fair insurance premium rates. It is worthwhile noting that based on Table 1, the prediction of standard deviation is not very satisfying. It was shown in 5.1 that the proposed relational model has some limited ability in predicting higher order moments, with the prediction errors of the standard deviation quite large compared to the mean prediction errors. However, based on the distribution simulated in Figure 3 the prediction seems to work well, and this is because here we are showing the results of the county with the best performance.

Figure 3:
The yield densities estimated based on the historical county data, farm-level yield prediction model (i.e. relational model), and the historical farm data.



The computed premium rates are reported in Table 4. For both premium principles, the premium rates increase with the coverage level and the risk load, which is to be expected. It is reassuring that the premium rates estimated from the predicted farm-level yield predicted using the relational model (labelled as "Farm-Predict") are close to that based on the direct historical farm-level yield data (labelled as "Farm-Actual"), depending on the coverage level and loading factor considered. It should be emphasized that the relational method does not rely on any of the available farm-level data in the county in the base country (though it uses farm-level data from the reference county).

It is important to note that the percentage values in the parenthesis refer to the percentage difference between the pricing results based on county-level yields and farm-level yields. More specifically, the percentages in the "Farm-predict" row can be interpreted as the difference in pricing results between the county-level contract based on county-level yields and the farm-level contracts based on farm-level data predicted from the relational model. Similarly, the percentages in the row "Farm-data" row can be interpreted as the difference in pricing results between the county-level contracts and farm-level contracts based on actual farm-level data. Therefore, it is not necessarily that lower percentage levels represent better performance of the model. Based on the results, there are no significant relationships between the performance of the relational model with coverage level or theta. For example, for the expectation premium principle, if we compare the percentage differences in pricing results between "Farm-predict" and "Farm-actual", we can see that the smallest error occurs when theta = 0.1 and coverage = 65%.

In most cases the premium rates from the relational model are higher compared to those from the actual farm-level yields, indicating that pricing based on the relational method can be slightly more conservative. This can be interpreted as a positive attribute in view of the parameter and model uncertainties. Finally, and most importantly, the results suggest that caution should be used when only the county-level yield data is used to approximate the farm-level premium rates, even though this is a common market practice (Gerlt et al., 2014). The discrepancy in the premium rates obtained from the “county model" and the “farm model" not only can be substantial, however, in many cases the county-level model can severely underestimate the actual results at the farm-level. To see this, the relative difference (in percentage) in premium rates between the “county model" and the “farm model" are computed, and these values are shown in parenthesis in Table 4. Most of these values are positive, indicating that the premium rates calculated from the “county model" are smaller than that from the “farm model". The degree of underestimation can be as high as over 40% in a few cases.

In addition to comparing the pricing results from both the expectation and the standard deviation premium principles, we also produce another set of premiums calculated from a new premium principle recently proposed by Zhu, Tan and Porth (2015). This premium principle, which is denoted as the multivariate weighted premium principle (MWPP), has a number of desirable properties. First, it is a general premium principle in that it easily recovers many of the common premium principles, including the Esscher premium principle, weighted premium principle, economic premium principle, etc. Second, the MWPP is also one of the few premium principles that satisfies all of the stylized properties of a premium principle, as described in Young (2004). Third, and most important, the MWPP has the flexibility of incorporating additional external factors into the pricing framework. This feature is particularly important in view of the complexity of the agricultural insurance sector, as compared to other traditional P&C insurance.

The systematic risk attributed to weather, for example, could be integrated into the MWPP. Economic factors, such as GDP, or inflation, could similarly be integrated into the pricing framework based on the MWPP. In Table 4, the MWPP premiums are calculated by exploiting additional information from the monthly average temperature and cumulative precipitation during the growing season as auxiliary factors. The results are consistent with those obtained using the expectation and the standard deviation premium principles. For example, underestimation of the premiums is observed for most of the cases when the premiums are calculated based on the “county model” rather than the “farm model.” We also find that the premiums obtained by both “farm models” are consistent with each other.

Table 4:
Agricultural insurance premium rates calculated using historical county yields, farm-level prediction yields (i.e. relational model), and historical farm-level yields, under the expectation and standard deviation premium principles. The value in parenthesis denotes the relative difference (in percentage) in premium rate between the “county model" and the “farm model".

θ	Method	c = 0.5	c = 0.65	c = 0.75	c = 0.85	c = 0.95
Expectation Premium Principle						
0.1	County	0.0138	0.0463	0.0746	0.1193	0.1697
	Farm-Predict	0.0253 (45.7%)	0.0613 (24.4%)	0.0922 (19.1%)	0.1337 (10.7%)	0.1856 (8.6%)
	Farm-Data	0.0179(23.4%)	0.0611 (24.2%)	0.0845 (11.8%)	0.1281 (6.9%)	0.1770 (4.1%)
0.2	County	0.0135	0.0450	0.0889	0.1272	0.1748
	Farm-Predict	0.0265 (49.1%)	0.0619 (27.2%)	0.1125 (21.0%)	0.1459 (12.8%)	0.2085 (16.2%)
	Farm-Data	0.0176 (23.4%)	0.0566 (20.5%)	0.0917 (3.0%)	0.1439 (11.6%)	0.1953 (10.5%)
0.3	County	0.0159	0.0658	0.0969	0.1435	0.1874
	Farm-Predict	0.0295 (46.1%)	0.0689 (4.5%)	0.1282 (24.4%)	0.1612 (11.0%)	0.2190 (14.4%)
	Farm-Data	0.0184 (13.6%)	0.0644 (-2.2%)	0.1099 (11.9%)	0.1483 (3.2%)	0.2038 (8.0%)
0.4	County	0.0166	0.0639	0.1104	0.1538	0.2039
	Farm-Predict	0.0217 (23.2%)	0.0776 (17.6%)	0.1222 (9.7%)	0.1669 (7.8%)	0.2134 (4.4%)
	Farm-Data	0.0188 (11.5%)	0.0730 (12.5%)	0.1138 (3.0%)	0.1718 (10.5%)	0.2201 (7.4%)
Standard Deviation Premium Principle						
0.1	County	0.0179	0.0526	0.0808	0.1249	0.1731
	Farm-Predict	0.0309 (42.1%)	0.0681 (22.8%)	0.0986 (18.1%)	0.1395 (10.4%)	0.1892 (8.5%)
	Farm-Data	0.0228 (21.5%)	0.0680 (22.6%)	0.0907 (10.9%)	0.1336 (6.5%)	0.1805 (4.1%)
0.2	County	0.0211	0.0573	0.1023	0.1382	0.1823
	Farm-Predict	0.0374 (43.5%)	0.0752 (23.8%)	0.1266 (19.2%)	0.1570 (12.0%)	0.2153 (15.3%)
	Farm-Data	0.0263 (19.7%)	0.0698 (18.0%)	0.1046 (2.2%)	0.1559 (11.4%)	0.2030 (10.2%)
0.3	County	0.0284	0.0862	0.1170	0.1604	0.1993
	Farm-Predict	0.0461 (38.4%)	0.0897(3.8%)	0.1484 (21.1%)	0.1788 (10.3%)	0.2285 (12.8%)
	Farm-Data	0.0319 (11/1%)	0.0845 (-2.1%)	0.1301 (10.0%)	0.1650 (2.8%)	0.2151 (7.4%)
0.4	County	0.0320	0.0921	0.1360	0.1756	0.2198
	Farm-Predict	0.0385 (17.0%)	0.1059 (13.1%)	0.1483 (8.3%)	0.1909 (8.0%)	0.2288 (3.9%)
	Farm-Data	0.0349 (8.2%)	0.1000 (7.9%)	0.1409 (3.5%)	0.1954 (10.1%)	0.2355 (6.6%)
MWPP						
	County	0.0167	0.0582	0.1072	0.1476	0.2076
	Farm-Predict	0.0274 (39.1%)	0.0786 (26.0%)	0.1111 (3.5%)	0.1517 (2.7%)	0.2089 (0.6%)
	Farm-Data	0.0245 (32.0%)	0.0662 (12.1%)	0.1104 (2.9%)	0.1554 (5.0%)	0.2047 (-1.4)

The large discrepancy between the county and farm models is a form of basis risk, and is a consequence of the so-called aggregation bias. Another approach to quantify the aggregation bias is to compare the difference between the variances corresponding to county-level yields and predicted farm-level yields. The difference can be expressed absolutely or relatively, and these are reported in Table 5 for each county. Intuitively, aggregating data leads to the possible cancellation of idiosyncratic risks, and hence there is smaller total risk in the aggregated data. Therefore, we expect higher yield variance at the farm level relative to the county level. This is confirmed by the results in Table 5, which shows that 16 out of 18 counties have positive aggregation bias.

Table 5:

Absolute and relative aggregation biases across the counties.

County	Absolute Bias	Relative Bias
BROKENHEAD	4.28	14%
DESALABERRY	3.44	10%
DUFFERIN	27.74	101%
GREY	8.80	32%
HANOVER	2.93	9%
LABROQUERIE	5.85	19%
MONTCALM	4.74	15%
NORTHNORFOLK	3.95	13%
PEMBINA	2.79	8%
PORTAGELAPRAIRIE	7.59	29%
RHINELAND	-24.33	-68%
ROLAND	2.60	9%
STEANNE	-1.54	-4%
SOUTHNORFOLK	10.58	42%
STANLEY	7.21	27%
TACHE	1.13	3%
THOMPSON	8.87	32%
WHITEMOUTH	4.96	16%

Section 7: Summary

This is the first research to develop a two-stage relational model to predict farm-level crop yield distributions in the absence of farm-level data. The first step of the algorithm uses variables, such as farm and county-level yield data, farm size, county-level coefficient of variation, and weather information, including temperature and precipitation over the growing season to search for the optimal reference county to borrow information from in the neighboring reference country (i.e. the U.S). The second step creates an explicit linkage between the two data sets in order to predict the farm-level yield in the base country (i.e. Canada). A unique data set that includes detailed corn yield data and farm characteristics at the farm-level and county-level, as well as weather data from the U.S. and Canada are utilized to empirically validate the model. The empirical results show that the relational model developed in this research is able to predict the farm-level yield accurately. The relational model achieves lower mean (standard deviation) prediction error of 28.85% (70.97%) compared to the benchmark model of 40.85% (277.65%), and the resulting actuarial premium rates are found to recover the premium rates based on the actual loss experience more closely compared to using only county-level data. The results also highlight the importance of the first step of the relational model algorithm, which selects an optimal reference county for borrowing the information based on a Euclidean distance metric, showing that the yield prediction results are substantially improved compared to randomly selecting a county to borrow the data from. Lastly, the results also confirm the existence of aggregation bias and caution using only county-level yield data to approximate farm-level premium rates, as the degree of underestimation can be as high as 40% in some cases.

The approach developed in this paper may be useful in improving yield forecasts and pricing in the situation where farm-level data is limited or not available. Further, this approach may also help to address the issue of aggregation bias, when county-level data is used as a substitute for farm-level data, which tends to results in underestimating the predictive risk relative to the true risk. In order to implement this approach in practice, government or private insurers must have access to detailed farm and county-level data from other countries. In reality, however, most countries do not currently make agriculture related data public, particularly at the farm-level. Therefore, countries with crop insurance programs should consider their policies on making some data public in order to leverage improved yield forecasting and pricing approaches that require shared data. Future research may consider examining additional selection algorithms beyond the Euclidean distance metric used in this research. In addition, although the results of this paper are mainly based on the normal distribution, future research may consider other marginal distributions or other dependence structures that contain heavy tails. Further, additional variables may be considered in the future in order to generalize the relational model and possibly improve the prediction ability.

References

Awondo, S., Datta, G., Ramirex, O., and Gonsah, E. (2012). Estimation of crop yield distribution and insurance premium using shrinkage estimator: A hierarchical bayes and small area estimation approach. In Agricultural & Applie Economics Association Annual Meeting, Seattle, Washington.

Barnett, B. J., Black, J.R., Hu, Y. and Skees, J.R. (2005). Is Area Yield Insurance Competitive with Farm Yield Insurance? Journal of Agricultural and Resource Economics, 2:285-301.

Baquet, A. E., and Skees, J.R. (1994). Group Risk Plan Insurance: An Alternative Management Tool tor Farmers. Choices (Q1):25-28.

Borman, J. I., Goodwin, B. K., Coble, K. H., Knight, T. O., and Rejesus, R. (2013). Account- ing for short samples and heterogeneous experience in rating crop insurance. Agriculture Finance Review, 73(1):88–101.

Bühlmann, H. (1980). An Economic Premium Principle. ASTIN Bulletin, 11:52–60.

Bühlmann, H., Delbaen, F., Embrechts, P., and Shiryaev, A. N. (1996). No-arbitrage, change of measure and conditional esscher transforms. CWI Quarterly, 9(4):291–317.

Claassen, R., and Just, R.E. (2009). Heterogeneity and Distributional Forms of Farm-Level Yields. American Journal of Agricultural Economics 93(1): 144-160.

Coble, K. and Barnett, B. J. (2008). Implications of integrated commodity programs and crop insurance. Journal of Agricultural and Resource Economics, 40(2):431–442.

Coble, K. H., Miller, M. F., Rejesus, R. M., Boyles, R., Knight, T. O., and Goodwin, B. K. (2011). Methodology analysis for weighting of historical experience. Technical report, USDA Risk Management Agency.

Coble, K.H., Knight, T.O., Miller, M.F., Goodwin, B.J., Rejesus, R.M., and Boyles, R. (2013). Estimating Structural Change in U.S. Crop Insurance Experience. Agricultural Finance Review, 73: 74-87.

Cooper, J., Zulauf, C., Langemeier, M., and Schnitkey, G. (2012) Implications of within County Yield Heterogeneity for Modeling Crop Insurance Premiums. Agricultural Finance Review 72:134–155.

Deng, X., Barnett, B. J., and Vedenov, D. V. (2007). Is there a viable market for area-based crop insurance? American Journal of Agricultural Economics, 89(2):508–519.

Furman, E. and Zitikis, R. (2008). Weighted Premium calculation principles. Insurance: Mathematics and Economics, 42:459–465.

Gerlt, S., Thompson, W., and Miller, D. (2014). Exploiting the relationship between farm-level yields and county-level yields for applied analysis. Journal of Agricultural and Re- source Economics, 39(2):253–270.

Gerber, H. and Shiu, S. (1994). Option pricing by esscher transforms. Transaction of Society of Actuaries, 46:99–140.

Glauber, J. W., and K. J. Collins. (2002). Crop Insurance, Disaster Assistance, and the Role of the Federal Government in Providing Catastrophic Risk Protection. Agricultural Finance Review, 62(2): 80–101.

Goodwin, B.K., and Ker, A.P. (1998). Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance. American Journal of Agricultural Economics, 80:139-153.

Goshtasby, A. A. (2012). Image Registration, chapter 2: Similarity and Dissimilarity Mea- sures. Springer.

Harri, A., Erdem, C., Keith, H. C., and Thomas, O. K. (2009). Crop yield distributions: A reconciliation of previous research and statistical tests for normality. Applied Economic Perspectives and Policy, 31(1):163–182.

Heilmann, W. R. (1989). Decision theoretic foundations of credibility theory. Insurance: Mathematics and Economics, 8(1):77–95.

Hoffman, W., Campbell, C., and Cook, K. (1994). Sowing Disaster: The Implications of Farm Disaster Programs For Taxpayers and the Environment. Washington, DC: Environmental Working Group.

Hubalek, J. and Neilsen, B. (2006). On the Esscher transform and entropy for exponential Lévy models. Quantitative Finance, 6(2):125–145.

Kamps, U. (1998). On a class of premium principles including the Esscher principle. Scandinavian Actuarial Journal, 1998(1):75–80.

Miranda, M. and Glauber, J. (1997). Systematic risk, reinsurance, and the failure of crop insurance markets. American Journal of Agriculture Economics, 79(1):206–215.

Ozaki, V. A., Goodwin, B. K., and Shirota, R. (2008). Parametric and nonparametric statistical modeling of crop yield: implications for pricing crop insurance contracts. Applied Economics, 40(9):1151–1164.

Porth, L., Tan, K. S., and Weng, C. (2013). Optimal reinsurance analysis from a crop insurer's perspective. Agriculture Finance Review, 73(2):310 – 328.

Porth, L., Zhu, W., and Tan, K. S. (2014). A credibility-based Erlang mixture model for pricing crop reinsurance. Agricultural Finance Review, 74(2):162–187.

Ramaswami, B. and Roe, T.L. (2004) Aggregation in area-yield crop insurance: The linear additive model. American Journal of Agriculture Economics, 86(2):420-431

Sherrick, B., Zanini, F., Schnitkey, G., and Irwin, S. (2004). Crop insurance valuation under alternative yield distributions. American Journal of Agricultural Economics, 86(2):406– 419.

Skees, J. R., Black, J.R., and Barnett, B.J. (1997). Designing and Rating an Area Yield Crop Insurance Contract. American Journal of Agricultural Economics, 79:430-438.

Skees, J., and Reed, M. (1986). Rate-Making and Farm-Level Crop Insurance: Implications for Adverse Selection, American Journal of Agricultural Economics, 68:653-959.
Wang, S. (1996). Premium calculation by transforming the layer premium density. ASTIN Bulletin, 26(1):71–92.

Woodard, J., Paulson, N., Vedenov, D., and Power, G. (2011). Efficiency in the modeling of dependence structures: An application of alternative copulas and agricultural insurance rating. Agricultural Economics, 42:101–112.

Woodard, J. D. (2014). A conditional distribution approach to assessing the impact of weather sample heterogeneity on yield risk estimation and crop insurance ratemaking. North American Actuarial Journal, 18(2):279–293.

Woodard, J. D., 2016a. Ag-analytics.org data warehouse. Available online at ag-analytics. org, accessed: 2015-09-30.
URL ag-analytics.org

Woodard, J. D., 2016b. Big data and ag-analytics: an open source, open data platform for agricultural & environmental finance, insurance, and risk. Agricultural Finance Review 76 (1), 15–26.

Woodard, J. D., 2016c. Data science and management for large scale empirical applications in agricultural and applied economics research. Applied Economic Perspectives and Policy.

Young, V. R. (2004). Premium principles. Encyclopedia of Actuarial Science.

Zhu, W., Tan, K. S., and Porth, L. (2015). On a class of premium calculation principle based on multivariate weighted distributions. Working Paper, University of Waterloo.

About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving 24,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement, and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and non-actuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org