A Relational Model for Predicting Farm-Level Crop Yield Distributions in the Absence of Farm-Level Data

Lysa Porth
Assistant Professor and Guy Carpenter Professor
Warren Centre for Actuarial Studies and Research
Department of Agribusiness and Agricultural Economics, University of Manitoba
Department of Statistics and Actuarial Science, University of Waterloo
Email: lysa.porth@ad.umanitoba.ca

Ken Seng Tan
University Research Chair Professor
Department of Statistics and Actuarial Science, University of Waterloo
200 University Avenue West, Waterloo, N2L 3G1, ON, Canada
Email: kstan@uwaterloo.ca

Wenjun Zhu[1]
Assistant Professor
School of Finance, Nankai University, Tianjin, China
Email: zhuwenjun@nankai.edu.cn

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2016 AAEA Annual Meeting, Boston, MA, July 31-August 2, 2016.

[1]**Corresponding author.**

A Relational Model for Predicting Farm-Level Crop Yield Distributions in the Absence of Farm-Level Data

May 26, 2016

Abstract

Individual farm-level expected yields serve as the foundation for crop insurance design and rating. Therefore, constructing a reasonable, accurate, and robust model for the farm-level loss distribution is essential. Unfortunately, farm-level yield data is often insufficient or unavailable in many regions to conduct sound statistical inference, especially in developing countries. This paper develops a new two-step relational model to predict farm-level crop yield distributions in the absence of farm yield losses, through "borrowing" information from a neighbouring country, where detailed farm-level yield experience is available. The first step of the relational model defines a similarity measure based on a Euclidean metric to select an optimal county, considering weather information, average farm size, county size and county-level yield volatility. The second step links the selected county with the county to be predicted through modeling the dependence structures between the farm-level and county-level yield losses. Detailed farm-level and county-level corn yield data in the U.S. and Canada are used to empirically examine the performance of the proposed relational model. The results show that the approach developed in this paper may be useful in improving yield forecasts and pricing in the case 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 tend to result in underestimating the predicted risk relative to the true risk.

Keywords: Relational Model; Aggregation Bias; Shortness of Data; Euclidean Distance; Crop Insurance; Yield Forecasting; Ratemaking.

1 Introduction

The U.S. Federal Crop Insurance Program (FCIP) has provided Multiple Peril Crop Insurance policies since its establishment in 1938. Until the 1980s, FCIP provided protection and ratemaking for all individuals in certain area, which caused adverse selection over time (Skees and Reed, 1986; Woodard et al., 2012). In 1980, the Federal Crop Insurance Act was established, and MPCI contracts were developed based on individual expected yields (i.e., the Actual Production History - APH program). Therefore, individual farm-level expected yields serve as the foundation for crop insurance design and rating to avoid problems such as adverse selection(Ker and Coble, 2003). Unfortunately, farm-level yield data is often insufficient or unavailable in many regions to conduct sound statistical inference due to different reasons (Gerlt et al., 2014). For example, there is only one growing season, and hence one yield observation per year. Moreover, due to crop rotation and other market forces, farmers do not grow the same crop each year. Therefore, farm-level yield observations are usually quite limited. In addition, when new MPCI programs are created, or coverage for new crop types are added, there are possibly no historical records for individual farm-level yields. The shortness of farm-level data may be particularly exagerated in developing countries.

Shortness and scarcity of crop yield data in one of the major challenges that hinder scientific statistical inference of the crop yield distribution, and hence the implementation of efficient MPCI programs(Borman et al., 2013; Porth et al., 2014). In the absence of sufficient qualified farm-level data, many studies have proposed methodologies to link county-level data and farm-level data (Coble and Barnett, 2008; Deng et al., 2007; Miranda, 1991), in order to

improve the prediction of individual farm-level loss distribution. Miranda (1991) uses a farm-specific "beta" to represent systemic risk in farm-level yield and adjust the county-level yield experience. Extending (Miranda, 1991), Coble and Barnett (2008) investigate the systemic risk and idiosyncratic risk by allowing the beta to vary over an assumed normal distribution. By assuming that farm yield is multiplicative conditioned on county yield, Deng et al. (2007) simulate a large number of pseudo farm-level yield distribution for pricing.

To date, literature has focused only on the problem of predicting farm-level yields using county-level yields within the same country, where at least some farm-level historical crop yield data is available to serve as the underlying basis of prediction. These approaches, however, do not address the challenge of statistical inference and actuarial pricing where farm-level yield data is completely absent. This paper for the first time investigates this problem and develops a new two-step relation model to predict individual farm-level yield distributions for country A, in absence of farm-level data, by borrowing information from the neighbouring country B, where detailed farm-level yield experience is available. The data set utilized in this analysis includes both farm-level and county-level corn production experience from the U.S. and Canada. This allows the evaluation of the farm-level yield prediction methodology developed in this study, through using the prediction errors of the estimated farm-level yield distributions relative to the actual farm-level yield distributions. In this study, Canada is the country to be predicted (i.e., country A) and the U.S. data is used as the basis for the prediction (i.e., country B).

The relational model developed in this paper for predicting farm-level crop yield distributions in the case where no individual farm yield losses are available involves two stages. First, we search for the most "similar" county in the U.S. relative to the Canadian county to be predicted. The second stage links the selected U.S. county with the county to be predicted in Canada. The first stage defines a similarity measure based on a Euclidean metric to select the optimal county in the U.S. This measure considers the "Euclidean distance" between the

two data objects, where a shorter distance implies more similar data objects. The Euclidean distance measure incorporates weather information over the target growing season, average farm size, county size and county-level yield volatility. The second stage links the two data sets, through modeling the dependence structures between the farm-level and county-level yield losses in the U.S. and applying the dependence structures to the Canadian data.

Given that this is the first paper to address the issue of farm-level yield prediction in the absence of farm-level data within the country, there is no benchmark model for comparison. However, the performance of the prediction methodology developed in this study is tested by comparing the pseudo farm-level yield data to the actual Canadian farm-level yield data (which is available, but, not utilized). In addition, we generalize the methods in Deng et al. (2007) to provide additional benchmarks for the relational model.

It has been shown that when county-level data is used as a substitute for farm-level data, aggregation bias may result (Gerlt et al., 2014) and this may lead to underestimating the predicted risk relative to the true risk. Therefore, the relational model proposed in this paper may help to overcome aggregation bias. The empirical results show that the relational model developed in this paper is able to predict the farm-level yield accurately. On the other hand, the existence of aggregation bias may lead to misleading and erroneous pricing results.

The remainder of this paper proceeds as follows. Section 2 introduces the data and notation used in this paper. Section 3 discusses the selection of the optimal U.S. proxy counties using the Euclidean distance measure. Next, the relational framework is introduced in Section 4 and robustness checks are discussed in Section 5. Section 6 shows the advantages of the relational model proposed in this paper with a MPCI pricing example. Section 7 concludes the paper.

2 Data and Notations

2.1 Data

A proprietary data set that includes actual farm-level and county-level corn production experience from the U.S. and Canada is utilized, including information regarding crop yield and farm size. The data sets cover the period from 1996 to 2011, and in Canada includes 1440 farms, and in the U.S. includes 5817 farms. In addition, corresponding weather data is used, including temperature and precipitation. These two countries are selected for the focus of the study because they both have relatively long and credible data sets available at both the farm and county levels. This allows the evaluation of relational model for predicting farm-level yields, using the prediction errors of the estimated farm-level yield distributions relative to the actual farm-level yield distributions. In this study, Canada is the country to be predicted (i.e., country A) and the U.S. data is used as the basis for the prediction (i.e., country B). Each of the data sets are discribed in more detail next.

U.S. Data Set:

The U.S. data set contains both farm-level and county-level yields per acre for different types of crops such as corn, sugarcane, soybeans, etc. across the country. For example, the corn data from Minnesota includes historical records from more than 120,000 farms in 78 counties. The time range of the data is from 1975 to 2013. In addition, U.S. weather data is considered including daily average temperature and rainfall from 1975 to 2013.

Canadian Data Set

The Canadian data set contains farm-level and municiple-level (which is equivalent to county-level in the U.S.) yields from the province of Manitoba covering 216 crop types from 19238 farms. The time range of the data is from 1996 to 2011. In addition, corresponding weather data is considered including daily temperature (maximum, minimum, and average) from 24

weather stations, as well as daily precipitation from 30 weather stations in Manitoba over the period of 1975 to 2011.

It is noteworthy that although there is a large county-level data set, the farm-level data is limited, with typically no more than 10 years of observations. This is to be expected, however, and is a motivation of this paper.

2.2 Notations

- For $c = 1, 2, \ldots, C$, let c be the index of U.S. counties and C denote the number of counties in U.S.
- For $d = 1, 2, \ldots, D$, let d be the index of Canadian counties and D denote the number of counties in Canada.
- For each U.S. county c,
  - Let $y_{c,t}^{\text{US}}$ denote the county-level yields for county c in year t, where $t = t_0, \ldots, T$ denotes the index year of county-level observations.
  - Let $A_{c,t}^{\text{US}}$ be the total number of acres for county c in year t, which represents the county size.
  - Denote $W_{c,t,p}^{\text{US}}, p = 1, \ldots, P$ as weather variables for county c, and there are P weather variables in total.
  - For each farm $f^{(c)}$, $f^{(c)} = 1, 2, \ldots, F^{(c)}$, where $F^{(c)}$ is the total farm number:
    - Let $s_f^{(c)}$ be the index year of farm-level observations, where $s_f^{(c)} = s_{f,0}^{(c)}, \ldots, S_f^{(c)}$.
    - Let $y_{f^{(c)},s_f^{(c)}}^{\text{US}}$ be farm-level yields for farm $f^{(c)}$, in year $s_f^{(c)}$.
- For each Canadian county d,
  - Let $y_{d,t}^{\text{CA}}$ denote the county-level yields for county d in year t, where $t = t_0, \ldots, T$

denotes the index year of county-level observations.

- Let $A_{d,t}^{\text{CA}}$ be the total number of acres for county d in year t, representing the county size.
- Denote $W_{d,t,p}^{\text{CA}}, p = 1, \ldots, P$ as weather variables for county d and there are P weather variables in total.
- For each farm $g^{(d)}$, $g^{(d)} = 1, 2, \ldots, G^{(d)})$, where $G^{(d)}$ is the number of total farms:
  - Let $s_g^{(d)}$ be the index year of farm-level observations, where $s_g^{(d)} = s_{g,0}^{(d)}, \ldots, S_d^{(d)}$.
  - Let $y_{g^{(d)},s_g^{(d)}}^{\text{CA}}$ be farm-level yields for farm $g^{(d)}$, in year $s_d^{(g)}$.

3 Searching for the Optimal U.S. Proxy County

3.1 Euclidean Distance Measure

Suppose that we are interested in predicting the farm-level yield in county d_0 in Canada using supplemental data in the U.S. The first step is to search for the most "similar" county in the U.S. In this paper, the similarity between two variables is described based on a similarity metric. There are different types of similarity metrics , and one of the most commonly used ones is the (normalized)"Euclidean distance", where a shorter distance implies more similar data objects. [1] Generally, given two points $\boldsymbol{p} = \{p_i : i = 1, \ldots, N\}$ and $\boldsymbol{q} = \{q_i : i = 1, \ldots, n\}$, the Euclidean distance $d_{\boldsymbol{p},\boldsymbol{q}}$ between the two points can be written as

$$d_{\boldsymbol{p},\boldsymbol{q}} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}. \tag{1}$$

[1] A detailed introduction of similarity measures is available in Goshtasby (2012).

In order to eliminate the intensities of bias and scale, the data is normalized by the means and standard deviations, i.e., normalized Euclidean distance $d^N_{\boldsymbol{p},\boldsymbol{q}}$ is used,

$$d^N_{\boldsymbol{p},\boldsymbol{q}} = \sqrt{\sum_{i=1}^{n}\left(\frac{p_i-\mu_{\boldsymbol{p}}}{\sigma_{\boldsymbol{p}}}-\frac{q_i-\mu_{\boldsymbol{q}}}{\sigma_{\boldsymbol{q}}}\right)^2}. \quad (2)$$

3.2 Selecting the Optimal U.S. Proxy County

In this study the Euclidean distance measure incorporates the following information:

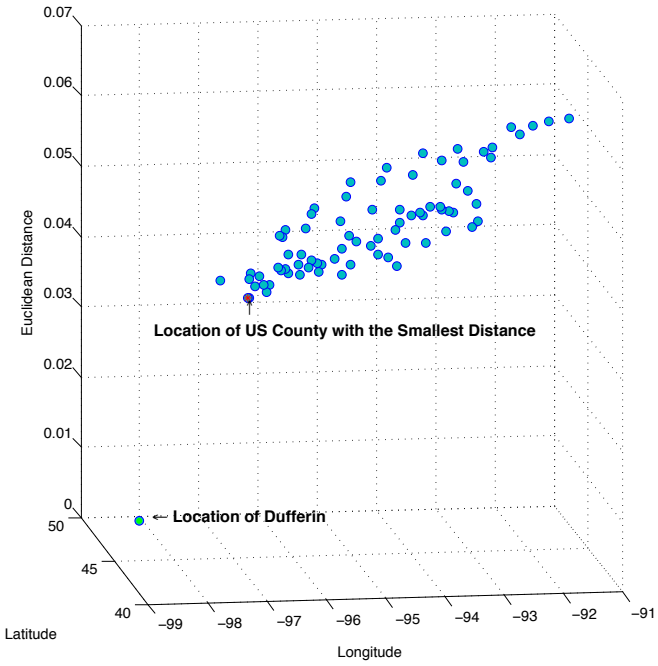
- Growing season weather information,
- Average farm size,
- County size,
- Yield volatility on county level, $CV^{\mathrm{US}}, CV^{\mathrm{CA}}$, where CV is the coefficient of variation (CV), which is defined as the ratio of standard deviation over mean.

Growing season for corn starts in April and the harvest ends in November. Therefore, the weather variables considered in this analysis correspond to the growing season, and include monthly mean temperature and cumulative precipitation from April to November. More specifically, they are TempApr, TempMay, TempJun, TempJul, TempAug, TempSep, TempOct, TempNov, PreApr, PreMay, PreJun, PreJul, PreAug, PreSep, PreOct, PreNov.

As an example, we are interested in predicting yield in county named "Dufferin" (county d_0) in Manitoba, Canada. Based on the weather variables, as well as the county-level yield experience in the U.S. and Canada, we search for the counties in both countries that are most "similar". This similarity is based on the Euclidean distance measure defined according to Equation (2), where the U.S. county that has the smallest distance is selected. For the example "Dufferin", a mapping for the US/Canada is displayed in Figure 1, which shows

how d^N_{c,d_0} changes spatially. The x-axis is for longitude, y-axis is for latitude and z-axis is for the distance, d^N_{c,d_0}. We can see that the selected county has the smallest d^N_{c,d_0}, and also locates very closely to the county "Dufferin".

Figure 1: Scatter Plot of Locations of Counties.



4 Relational Prediction Model

4.1 Stochastic Specification

Without losing generality, we assume that both the county-level and farm-level yields in the U.S and Canada satisfy the following processes:

$$\log\left(y^{\text{US}}_{c,t}\right) = \mu_{c,t} + \varepsilon_{c,t}, \quad t = t_0, \ldots, T, \tag{3}$$

$$\log\left(y^{\text{CA}}_{d,t}\right) = \mu_{d,t} + \varepsilon_{d,t}, \quad t = t_0, \ldots, T, \tag{4}$$

$$\log\left(y^{\text{US}}_{f^{(c)},s^{(c)}_f}\right) = \mu_{f^{(c)},s^{(c)}_f} + \varepsilon_{f^{(c)},s^{(c)}_f}, \quad s^{(c)}_f = s^{(c)}_{f,0}, \ldots, S^{(c)}_f, \tag{5}$$

$$\log\left(y^{\text{CA}}_{g^{(d)},s^{(d)}_g}\right) = \mu_{g^{(d)},s^{(d)}_g} + \varepsilon_{g^{(d)},s^{(d)}_g}, \quad s^{(c)}_f = s^{(c)}_{f,0}, \ldots, S^{(c)}_f, \tag{6}$$

where $\varepsilon_{c,t} \sim WN(0, \sigma_c^2)$, $\varepsilon_{d,t} \sim WN(0, \sigma_d^2)$, $\varepsilon_{f^{(c)}, s_f^{(c)}} \sim WN(0, \sigma_{f^{(c)}}^2)$, $\varepsilon_{g^{(d)}, s_g^{(d)}} \sim WN(0, \sigma_{g^{(d)}}^2)$, and "$WN$" represents a white noise process.

Following Deng et al. (2007), it is further assumed that the drift terms of the yield processes are linear. More specifically,

$$\mu_{c,t} = a_c + b_c(t - t_0), \quad t = t_0, \ldots, T, \tag{7}$$

$$\mu_{d,t} = a_d + b_d(t - t_0), \quad t = t_0, \ldots, T, \tag{8}$$

$$\mu_{f^{(c)}, s_f^{(c)}} = a_{f^{(c)}} + b_{f^{(c)}}(s_f^{(c)} - s_{f,0}^{(c)}), \quad s_f^{(c)} = s_{f,0}^{(c)}, \ldots, S_f^{(c)}, \tag{9}$$

$$\mu_{g^{(d)}, s_g^{(d)}} = a_{g^{(d)}} + b_{g^{(d)}}(s_g^{(d)} - s_{g,0}^{(d)}), \quad s_g^{(d)} = s_{g,0}^{(d)}, \ldots, S_g^{(d)}, \tag{10}$$

4.2 Relational Model

The question of interest is to predict the farm-level yields in county d_* in Canada, with information from county c_* in the U.S. county borrowed based on the Euclidean measure. The relational model proposed in this paper first models the dependence between the U.S. county-level and farm-level yields, and then applies the resulting dependence for the Canadian county-level and farm-level yields. Additionally, it is assumed that $\mathsf{Cov}(\varepsilon_{c,s}, \varepsilon_{f^{(c)},s}) = \sigma_{cf^{(c)}} = \rho_{cf^{(c)}}\sigma_{f^{(c)}}\sigma_c$, $\mathsf{Cov}(\varepsilon_{d,s}, \varepsilon_{g^{(d)},s}) = \rho_{dg^{(d)}}\sigma_{g^{(d)}}\sigma_d$, where $\rho_{cf^{(c)}} = \rho(\varepsilon_{c,s}, \varepsilon_{f^{(c)},s})$ and $\rho_{dg^{(d)}} = \rho(\varepsilon_{d,s}, \varepsilon_{g^{(d)},s})$ are correlation coefficients. Therefore, the variance-covariance matrix of $\varepsilon_{c,s}$ and $\varepsilon_{f^{(c)},s}$ as

$$\Sigma^{\mathrm{US}} = \begin{pmatrix} \sigma_c^2 & \sigma_{cf^{(c)}} \\ \sigma_{cf^{(c)}} & \sigma_{f^{(c)}}^2 \end{pmatrix}, \tag{11}$$

and the variance-covariance matrix of $\varepsilon_{d,t}$ and $\varepsilon_{g^{(d)},n_g^{(d)}}$, as

$$\Sigma^{\mathrm{CA}} = \begin{pmatrix} \sigma_d^2 & \sigma_{dg^{(d)}} \\ \sigma_{dg^{(d)}} & \sigma_{g^{(d)}}^2 \end{pmatrix}. \tag{12}$$

Since county c_* is the U.S. county selected based on the Euclidean measure and used to predict the farm-level yields in county d_* in Canada, parameters a_{c_*}, b_{c_*}, a_{d_*}, b_{d_*}, $a_{f^{(c_*)}}$, Σ^{US}, and $\sigma_{d_*}^2$ are estimated based on the available data. In order to predict the unknown farm-level distribution parameters of Canada, $a_{g^{(d_*)}}$, $b_{g^{(d_*)}}$, $\sigma_{d_* g^{(d_*)}}$, and $\sigma_{g^{(d_*)}}^2$, the following additional assumptions are made.[2].

In the relational model proposed in this paper, we assume that the

$$\mu_{g,s} = \alpha \times \mu_{f,s}, \tag{13}$$

$$\sigma_g = \beta \times \sigma_f \tag{14}$$

where $s \in \{s_{f,0}, \ldots, S_f\} \cup \{s_{g,0}, \ldots, S_g\}$. In addition, it is assumed that they have the same dependence structure, which indicates that in the special bivariate normal case when $(\varepsilon_{c_*}, \varepsilon_f) \sim MN(\mathbf{0}, \Sigma^{\mathrm{US}})$, we have $\rho_{cf^{(c)}} = \rho_{dg^{(d)}}$. Algorithm 4.1 summarizes the step-by-step farm-level yield forecasting procedure for the relational model.

Algorithm 4.1 (**Relational Farm-Level Prediction Model**).

Step 1: *Detrend the U.S. county-level data according to Equation* (3), (7), *detrend U.S. farm-level data according to Equation* (5), (9), *and obtain* $\varepsilon_{c_*,t}$, *and* $\varepsilon_{f,s}$, *respectively, where* $f = 1, \ldots, F,\ t = 1, \ldots, T,\ s \in \{s_{f,0}, \ldots, S_f\}$.

Step 2: *Estimate the marginal distributions of* $\varepsilon_{c_*,t}$ *and* $\varepsilon_{f,s}$, *denoted as* $f_{c_*}(x; \boldsymbol{\theta})$ *and* $f_f(y; \boldsymbol{\theta})$, *where* $\boldsymbol{\theta}$ *is the parameter vector of the distributions.*

[2]Without confusion, we eliminate the superscripts of (c_*) and (d_*) in the rest of the paper.

Step 3: *Estimate the dependence structure of* $\varepsilon_{c_*,t}$ *and* $\varepsilon_{f,s}$*, denoted as* $l_{c_*,f}(x, y; \boldsymbol{\eta})$*, where* $f = 1, \ldots, F$ *and* $\boldsymbol{\eta}$ *is parameter vector of the dependence structure. In the bivariate normal case, it is equivalent to estimate* Σ^{US}*.*[3]

Step 4: *Detrend Canadian county-level data according to Equation* (4), (8)*, and obtain* $\varepsilon_{d_*,t}$*.*

Step 5: *By applying the relational assumptions in Equation* (13) *and Equation* (14)*, determine the margins and the dependence structure of* $\varepsilon_{d_*,t}$ *and* $\varepsilon_{g_*,s}$*, denoted as* $f_{d_*}(x; \boldsymbol{\theta})$*,* $f_{g_*}(y; \boldsymbol{\theta})$*, and* $l_{d_*,g_*}(x, y; \boldsymbol{\eta})$*, where* g_* *is the Canadian farm in county* d_* *we are interested in predicting.*

Step 6: *Simulate* N *bivariate distribution according to* $f_{d_*}(x; \boldsymbol{\theta})$*,* $f_{g_*}(y; \boldsymbol{\theta})$*, and* $l_{d_*,g_*}(x, y; \boldsymbol{\eta})$*, denoted as* $(\hat{\varepsilon}_{f,i}, \hat{\varepsilon}_{g_*,i})$*, where* $i = 1, \ldots, N$*. In the special case of the bivariate normal distribution assumption, simulate* N *pairs of bivariate normal samples with variance covariance matrix* $\Sigma^{CA} = \Sigma^{US}$*.*

Step 7: *For each simulated sample, recover predicted yields in year* s*,* $\hat{y}^{CA}_{g_*,s}$*, according to Equation* (6)*, namely,*

$$\hat{y}^{CA}_{g_*,s,i} \quad = \quad e^{(\alpha\mu_{f,s} + \hat{\varepsilon}_{g,i*})}, \ i = 1, \ldots, N. \tag{15}$$

4.3 Prediction Results

One advantage of Algorithm 4.1 in Section 4.2 is that it is able to predict the whole distribution of farm-level yields for Canadian county d_*. In order to assess the performance of the relational model, we compare the predicted pseudo farm-level yield distribution to the actual Canadian farm-level yield data (which is available, but, not utilized). Due to the limited farm-level data, which contains less than 10 years observation for each farm, the mean and

[3]In multivariate normal setting, the dependence structure is fully determined by the covariance. More generally, different dependence structure models, such as copula models, can be applied.

standard deviation of the predicted pseudo farm-level distribution are compared with actual farm-level data. To be more specific, for each county d_* being predicted, the relative mean squared error (RMSE) for the statistic, Q, is calculated as,

$$RMSE_Q^{(d_*)} = \sqrt{\frac{1}{F^{(d_*)}} \sum_{f=1}^{F^{(d_*)}} \left(\frac{\tilde{Q}_f - \hat{Q}_f}{\tilde{Q}_f} \right)^2}, \tag{16}$$

where Q represents mean or standard deviation; $\tilde{Q}$ is the estimator of Q calculated with actual Canadian farm-level data; $\hat{Q}$ is the estimator of Q calculated from the predicted pseudo farm-yield data. In addition, the calculation of the RMSE, considers the farms in each county that contain a minimum of 6 years of observations, and hence $F^{(d_*)}$ in Equation (16) is the total number of farms in county d_* that satisfy this criterion.

As mentioned previously, this is the first paper to develop an approach to forecast farm-level yield data in the absence of farm-level data within the same county. Therefore, in order to evaluate the performance of the relational framework developed in this paper, the methods of Deng et al. (2007) are generalized to serve as a benchmark model. Deng et al. (2007) model farm-level yield distributions, assuming that farm yield is multiplicative conditioned on county yield. More specifically, for each farm i in county j, crop yield observation in year s, $y_{i,s}$ and $y_{j,s}$ have a multiplicative relationship:

$$y_{i,s} \quad = \quad y_{j,s} \times \eta_{i,s}, \tag{17}$$

where η_{is} are defined as the idiosyncratic shocks in farm i relative to county j. The algorithm of the benchmark model is summarized as follows:

Algorithm 4.2 (**Benchmark Prediction Model**).

Step 1: *Calculate the detrended county-level yields as*

$$y_{c_*,t}^{Det\text{-}US} = \frac{y_{c_*,t}^{US}}{e^{\mu_{c_*,t}}} e^{\mu_{c_*,T}}, \tag{18}$$

$$y_{d_*,t}^{Det\text{-}CA} = \frac{y_{d_*,t}^{CA}}{e^{\mu_{d_*,t}}} e^{\mu_{d_*,T}}, \tag{19}$$

where $t = 1, \ldots, T$,; $\mu_{c_*,t}, \mu_{d_*,t}$ *are calculated according to Equation* (7) *and Equation* (8)*, respectively.*

Step 2: *Calculate the U.S. farm-level idiosyncratic shock for farm* f *year* s_f, η_{f,s_f}^{US},

$$\eta_{f,s_f}^{US} = \frac{y_{f,s_f}^{US}}{y_{c,s_f}^{US}}, \tag{20}$$

where $f = 1, \ldots, F; s_f = s_{f,0}, \ldots, S_f$*;* F *is the total number of farms in county* c_**; and* S_f *is the total number of observations in farm* f*.*

Step 3: *By assuming that Canada and U.S. share the same idiosyncratic shocks, simulate pseudo farm-level yields as*

$$\boldsymbol{y}^{Pseudo\text{-}CA} = \boldsymbol{y}_{d_*}^{Det\text{-}CA} \otimes \boldsymbol{\eta}_f^{US}, \tag{21}$$

where $\boldsymbol{y}_{d_*}^{Det\text{-}CA} = (y_{d_*,s_{f,0}}^{Det\text{-}CA}, \ldots, y_{d_*,S_f}^{Det\text{-}CA})'$ *is a* $((S_f - s_{f,0} + 1) \times 1)$ *vector;* $\boldsymbol{\eta}_f^{US} = (\eta_{f,1}^{US}, \ldots, \eta_{f,F}^{US})$ *is a* $(1 \times F)$ *vector;* $\otimes$ *represents the Kronecker product* [4]*.*

The resulting $((S_f - s_{f,0} + 1) \times F)$ matrix of $\boldsymbol{y}^{\text{Pseudo-CA}}$ contains the pseudo farm-level Canadian yields predicted from the benchmark model in Algorithm 4.2. The prediction results for both Algorithm 4.1 and Algorithm 4.2 are summarized in Table 1 [5]. The first

[4]Kronecker product of $(n \times 1)$ vector $\boldsymbol{a}$ and $(1 \times m)$ vector $\boldsymbol{b}$ is defined as $\boldsymbol{a} \otimes \boldsymbol{b} = \begin{pmatrix} a_1 b_1 & \ldots & a_1 b_m \\ & \ldots & \\ a_n b_1 & \ldots & a_n b_m \end{pmatrix}$.

[5]Counties that do not cover the whole period from 1996-2011 are excluded from the analysis.

column records county names. Columns 2 and 3 (Column 4 and 5) summarize results for the relational model proposed in this paper (benchmark model). We start the investigation of the relational model by simply assuming that $\alpha = \beta = 1$ in Equations (13) and Equation (14), which implies that the farm-level and county-level yields in Canada and U.S. share the same drifts and variance. This assumption is generalized in the next section in an optimization framework.

Table 1: Summary Results of the Relational Model. The first column records county names. Columns 2 and 3 (Columns 4 and 5) summarize results for the relational model proposed in this paper (benchmark model). "Avg" represents mean estimation and "Std" represents standard deviation estimation.

County Name	**Relational Model**		**Benchmark Model**	
	Avg	**Std**	**Avg**	**Std**
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%
North Norfolk	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%
South Norfolk	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%

Prediction results show that the relational model proposed in this paper performs better, and more accurate predicts the mean of the farm-level yield distribution compared to the

benchmark model, with only two exceptions, including county Dufferin (38.5% worse) and county Labroquerie (3.02% worse). The average RMSE for mean prediction is 28.85% from the relational model, which is 41.62% better than the benchmark model (40.85%). In the best case of the relational model prediction, the RMSE_{Avg} is 9.57% in county Brokenhead. In terms of the standard deviation prediction, the best result from the relational model is 24.42% in county Desalaberry, which is 72.71% better than the benchmark. Although the RMSE_{Std} in some counties are large, the prediction based on relational model is still better than the benchmark model.

It is noteworthy that the Deng et al. (2007) method is to predict farm-level distribution within the same county, while the objective of this paper is, in the absence to actual farm-level yield data, to predict farm-level distribution by borrowing information from another county. Hence, compared to (Deng et al., 2007) method, the benchmark model in Algorithm 4.2 has an additional assumption that Canada and U.S. share the same idiosyncratic shocks, $\boldsymbol{\eta}_f^{\text{US}}$, in order to provide an idea of the performance of the relational model. Therefore, the benchmark model are not the same as Deng et al. (2007) and the results of the benchmark model in Table 1 does not indicate that the methodology in Deng et al. (2007) is not good because they are essentially different models to address issues.

5 Robustness Checks

5.1 An Optimization Framework

Recall that the general relational model proposed in this paper assumes α and β are preset to be 1 in Equation (13) and Equation (14) in the prediction results in Subsection 4.3. In this section, the appropriateness of this assumption is investigated. More specifically, the optimal α and β in a general relational model is explored, based on an optimization framework in this subsection.

Let us assume that $\tilde{y}_{g,s}$ is the predicted farm-level yield based on the relational model, satisfying

$$\begin{aligned} \log\tilde{y}_{g,s} &= \tilde{\mu}_{g,s} + \tilde{\varepsilon}_{g,s}, && (22) \\ \tilde{\mu}_{g,s} &= \alpha\mu_{f,s} && (23) \\ \tilde{\varepsilon}_{g,s} &= \beta\varepsilon_{x,s} && (24) \end{aligned}$$

where $\varepsilon_{x,s}$ is a random variable such that $\mathsf{Var}(\varepsilon_{x,s}) = \sigma_f^2$. These assumptions guarantee that Equations (13), (14) are satisfied. The objective is to minimize the RMSE, denoted as

$$\mathcal{L}(\alpha, \beta) = \mathsf{E}\left(\frac{y_{g,s} - \tilde{y}_{g,s}}{y_{g,s}}\right)^2 \quad (25)$$

Proposition 5.1. *For the predicted yield $\tilde{y}_{g,s}$ satisfying Equation* (22) *to Equation* (24)*, the optimal α and β that minimize the RMSE as Equation* (25) *are expressed as*

$$\begin{aligned} \alpha &= \left(\frac{3}{2}\sigma_g^2\rho_{gx}^2 - 3\sigma_g^2\rho_{gx}^2 + \mu_g + \frac{3}{2}\sigma_g^2\right)/\mu_{f,s}, && (26) \\ \beta &= \frac{\sigma_g}{\sigma_f}\rho_{gx}. && (27) \end{aligned}$$

where $\rho_{xg} = \rho(\varepsilon_{x,s}, \varepsilon_{g,s})$.

Proof. Proof of Proposition 5.1 is collected in Appendix A. □

Table 2 summarizes the optimal α and β for each county and the corresponding prediction results [6]. The optimal α and β are close to 1, indicating that the assumptions in Equation (13) and Equation (14) are appropriate for the data utilized in this paper.

[6]Since the hypothetical random variable $\tilde{\varepsilon}_{g,s}$ is latent in reality, the prediction results are calculated numerically.

Table 2: Summary Results of the General Relational Model in an Optimization Framework. The first column records county names. Columns 2 and 3 display the optimal α and β. Columns 4 and 5 summarize the RMSE results for the general relational model with the corresponding α and β. "Avg" represents mean estimation and "Std" represents standard deviation estimation.

County Name	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	**Relational Model**	
			Avg	**Std**
Brokenhead	0.9928	0.9522	8.74%	60.44%
Desalaberry	1.0004	1.0008	19.51%	24.46%
Dufferin	0.8815	0.3821	20.85%	60.45%
Grey	1.0225	1.0164	24.76%	58.69%
Hanover	0.9915	1.0099	37.00%	37.29%
Labroquerie	0.9539	0.9539	22.78%	43.99%
Montcalm	1.1083	1.1260	8.27%	254.92%
North Norfolk	1.0088	1.0024	30.48%	73.01%
Pembina	1.0502	0.9605	14.52%	57.01%
Portagelaprairie	1.0395	1.0337	16.30%	83.94%
Rhineland	0.8267	0.8305	30.84%	48.13%
Roland	0.8031	1.2826	34.71%	39.38%
Steanne	1.0721	0.9174	8.52%	25.55%
South Norfolk	1.0330	1.0347	15.55%	52.20%
Stanley	0.8837	1.2820	23.47%	94.93%
Tache	0.9998	1.0001	17.88%	63.27%
Thompson	1.0004	1.0008	26.06%	143.02%
Whitemouth	0.9957	0.9920	13.22%	63.58%

5.2 Robustness Check for Euclidean Distance Measure

This paper develops a relational model to predict farm-level yield in the absence of farm-level losses, through borrowing information from another country based on Euclidean distance. More specifically, for each Canadian county we are interested in predicting, the U.S. county with the smallest Euclidean distance is selected, defined according to Equation (2). The robustness check of this subsection shows the usefulness of the Euclidean distance in selecting the optimal county for the relational analysis.

The robustness check proceeds as follows. The Canadian farm-level yield is predicted based on the relational model, but, without selecting an "optimal U.S. county" using the Euclidean

distance. Instead, in the prediction for each county, we randomly select a county from all U.S. counties and then perform the relational analysis. This exercise is repeated 100 times for each Canadian county and the average RMSE is recorded. The results are displayed in Table 3.

The prediction results based on selecting the optimal county using the Euclidean distance (hereafter we call "Euclidean method") performs much better than the randomly selected counties (hereafter we call "Random method"). On average, the Euclidean method is 155.44 % ($\frac{73.68\% - 28.85\%}{28.85\%}$) better in mean prediction and 70.11% ($\frac{120.85\% - 70.97\%}{70.97\%}$) better in standard deviation prediction. In particular, the mean estimation for the county "Brokenhead" with the Euclidean method is 1002.97% better than the Random method, while the standard deviation estimation for the county "Desalaberry" with the Euclidean method is 227.94% better than the Random method.

6 Pricing Example

The relational model developed in the preceding sections provides a framework to achieve accurate farm-level crop yield forecasting, which is the foundation for crop insurance ratemaking, such as Multi-Peril Crop Insurance (MPCI) contracts. In this section, a pricing example is used to illustrate the importance of estimating unknown farm-level yields to establish robust premium rates. A MPCI contract is priced, with coverage levels varying from 50% to 100%. In calculating premium rates, both the expectation premium principle and the standard deviation premium principle are considered. The expectation premium principle and standard deviation premium principle for a certain loss random variable X are defined as

$$\textbf{Expectation:}\ \text{Premium Rate} = \mathsf{E}(X)(1+\theta); \tag{28}$$

$$\textbf{StdDev:}\ \text{Premium Rate} = \mathsf{E}(X) + \theta\sqrt{\mathsf{Var}(X)}, \tag{29}$$

Table 3: Summary Results of the Relational Model. The first column records county names. To compare, Columns 2 and 3 summarize results for the relational model with a U.S. county selected by minimizing the Euclidean distances. Columns 4 and 5 display the predicting results without selecting an optimal county. "Avg" represents mean estimation and "Std" represents standard deviation estimation.

County Name	**Minimum Euclidean Distance County**		**Randomly Selected County**	
	Avg	**Std**	**Avg**	**Std**
Brokenhead	9.57%	62.94%	105.50%	128.48%
Desalaberry	21.49%	24.42%	64.56%	80.08%
Dufferin	66.64%	176.81%	73.30%	179.10%
Grey	25.20%	55.80%	93.86%	124.45%
Hanover	37.12%	37.40%	104.57%	80.46%
Labroquerie	23.65%	40.21%	123.37%	101.44%
Montcalm	38.40%	117.34%	39.86%	163.48%
North Norfolk	31.17%	55.44%	114.22%	137.44%
Pembina	24.08%	37.08%	48.22%	83.10%
Portagelaprairie	22.19%	56.03%	80.54%	125.87%
Rhineland	38.51%	143.89%	48.97%	158.06%
Roland	41.06%	55.63%	42.76%	123.65%
Steanne	25.53%	36.18%	49.56%	70.60%
South Norfolk	19.80%	33.29%	82.59%	109.66%
Stanley	32.22%	86.39%	47.64%	127.35%
Tache	18.21%	66.01%	76.02%	108.63%
Thompson	30.93%	130.05%	59.49%	218.26%
Whitemouth	13.44%	62.47%	71.26%	95.26%
Avg	28.85%	70.97%	73.68%	120.85%
Min	9.57%	24.42%	39.86%	70.60%
Max	66.64%	176.81%	123.37%	218.26%

where θ is the risk loading parameter. Different θ is considered ranging from 0.1 to 0.4. The range is selected based on supporting literature (Porth et al., 2013; 2014) and market convention.

Premium rates are calculated using the loss-cost ratio (LCR), defined as the ratio of indemnity over liability. More specifically, the liability of a contract with coverage level c is defined as $L = c \cdot \mathsf{E}(y_t)$, and the indemnity of the contract is defined according to $I = \max(L - y_t, 0)$.

Hence the LCR is defined as:

$$LCR = \frac{I}{L} = \frac{\max(c \cdot \mathsf{E}(y_t) - y_t, 0)}{c \cdot \mathsf{E}(y_t)}. \tag{30}$$

To illustrate the existence of aggregation bias and highlight the importance of the relational model proposed in this paper, pricing results based on different underlying data are considered. More specifically, we compare premium rates calculated from county-level yields, predicted farm-level yields (using the relational model), as well as the actual farm-level yields. Estimated densities of the three types of underlying data are pictured in Figure 2. The pricing results for the county with the smallest prediction error in Table 4 [7]. The prices based on the predicted farm-level data are very close to results from the actual farm-level data, demonstrating the usefulness of the relational model proposed in this paper. For both premium principles, the premium rates increase with the coverage level c and the risk loading parameter θ. Generally, pricing with the predicted farm-level data is slightly more conservative compared to the actual farm-level data.

Comparing the pricing results based on county-level data and farm-level data, the discrepancies indicate aggregation bias, which demonstrates the importance of improving the prediction accuracy of farm-level data using approaches such as the relational model proposed in this paper. In the case where actual farm-level data is not available, the aggregation bias can be formally measured by the difference between the variances of county-level yields and predicted farm-level yields. The aggregation bias can be considered from both absolute and relative views as follows,

$$\text{Absolute Aggregation Bias} = \sigma_f^2 - \sigma_c^2, \tag{31}$$

$$\text{Relative Aggregation Bias} = \frac{\sigma_f^2 - \sigma_c^2}{\sigma_f^2}. \tag{32}$$

[7] Refer the first row in Table 1. Due to length restriction, pricing results for other counties are available upon request.

Table 4: Summary of Pricing Results. The first panel shows the results for expectation premium principle ("Expectation") and the second panel for standard deviation premium premium principle ("StdDev"). The underlying data used for pricing are displayed in the column named "Underlying" including county-level data, predicted farm-level data with the relational model, and real farm-level data.

	θ	Underlying	$c = 0.5$	$c = 0.65$	$c = 0.75$	$c = 0.85$	$c = 0.95$
Expectation	0.1	**County**	0.0122	0.0521	0.0772	0.1176	0.1599
		Farm-Predict	0.0184	0.0558	0.1001	0.1380	0.1899
		Farm-Real	0.0184	0.0521	0.0874	0.1373	0.1799
	0.2	**County**	0.0146	0.0560	0.0900	0.1330	0.1826
		Farm-Predict	0.0264	0.0640	0.1050	0.1547	0.1972
		Farm-Real	0.0218	0.0630	0.1029	0.1464	0.1956
	0.3	**County**	0.0160	0.0585	0.1019	0.1416	0.2097
		Farm-Predict	0.0264	0.0711	0.1112	0.1561	0.2072
		Farm-Real	0.0243	0.0664	0.1074	0.1543	0.2042
	0.4	**County**	0.0150	0.0651	0.0967	0.1409	0.2040
		Farm-Predict	0.0219	0.0634	0.1128	0.1537	0.2191
		Farm-Real	0.0238	0.0775	0.1041	0.1644	0.2331
StdDev	0.1	**County**	0.0161	0.0587	0.0837	0.1225	0.1634
		Farm-Predict	0.0236	0.0627	0.1071	0.1440	0.1934
		Farm-Real	0.0232	0.0585	0.0938	0.1430	0.1839
	0.2	**County**	0.0228	0.0695	0.1031	0.1435	0.1895
		Farm-Predict	0.0374	0.0776	0.1184	0.1657	0.2046
		Farm-Real	0.0316	0.0770	0.1161	0.1569	0.2024
	0.3	**County**	0.0247	0.0738	0.1147	0.1577	0.2071
		Farm-Predict	0.0424	0.0914	0.1370	0.1810	0.2162
		Farm-Real	0.0360	0.0811	0.1327	0.1809	0.2118
	0.4	**County**	0.0313	0.0913	0.1233	0.1633	0.2212
		Farm-Predict	0.0395	0.0886	0.1387	0.1781	0.2359
		Farm-Real	0.0447	0.1036	0.1304	0.1868	0.2458

Figure 2: Estimated Densities of the Three Types of Underlying Data.

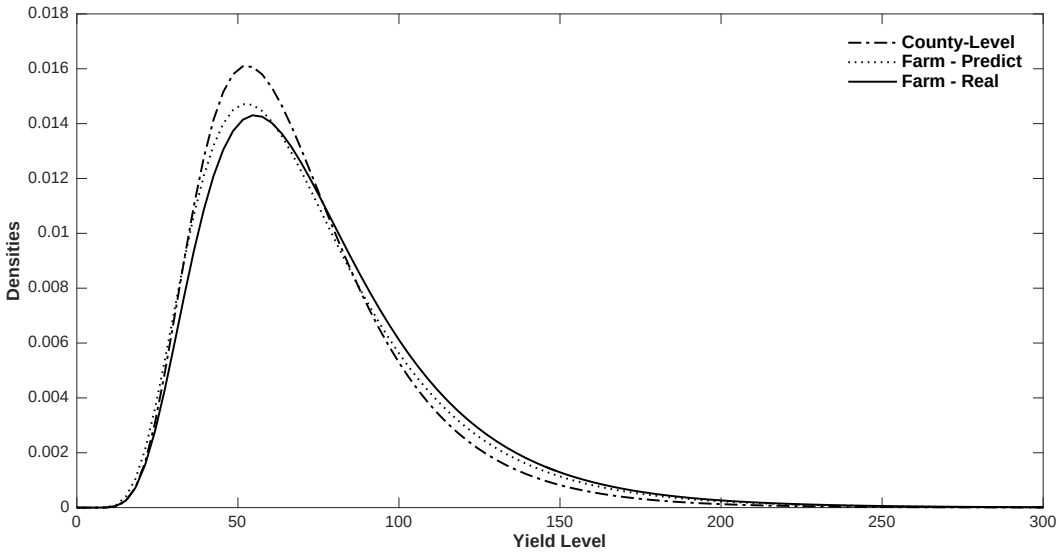


Table 5 summarizes both absolute and relative aggregation biases for each county. Intuitively, aggregation cancels idiosyncratic risks, leading to smaller total risk in aggregated data. Therefore, we expect higher variance in county-level yields compared to farm-level yields. Based on the results in Table 5 confirm this hypothesis, for 16 out of 18 counties, Aggregation biases are positive, indicating that variances of the predicted farm-level yields are greater than the county-level yields. However, this is not conclusive as there are still counties with negative aggregation biases.

7 Conclusion

This is the first paper to develop a relational model to predict farm-level crop yield distributions in the absence of farm-level losses, through borrowing information from a neighbouring country, where detailed farm-level yield experience is available. Detailed corn yield data sets at the farm-level and county-level from the U.S. and Canada are utilized, as well as corresponding weather data including temperature and precipitation, in the empirical analysis to facilitate the examination the efficiency of the proposed relational model. The empirical

Table 5: Summary of Aggregation Biases.

County	Absolute Bias	Relative Bias
Brokenhead	4.28	0.14
Desalaberry	3.44	0.10
Dufferin	27.74	1.01
Grey	8.80	0.32
Hanover	2.93	0.09
Labroquerie	5.85	0.19
Montcalm	4.74	0.15
Northnorfolk	3.95	0.13
Pembina	2.79	0.08
Portagelaprairie	7.59	0.29
Rhineland	-24.33	-0.68
Roland	2.60	0.09
Steanne	-1.54	-0.04
Southnorfolk	10.58	0.42
Stanley	7.21	0.27
Tache	1.13	0.03
Thompson	8.87	0.32
Whitemouth	4.96	0.16

results show that the relational model developed in this paper is able to predict the farm-level yield accurately. The approach developed in this paper may be useful in improving yield forecasts and pricing in the case 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 tend to result in underestimating the predicted risk relative to the true risk.

Although the results of this paper are mainly based on the normal distribution, the relational model presented in this paper is very flexible. In the future research, other marginal distributions and other dependence structures containing heavy tails can be applied to the relational model. With additional data and information, more assumptions can be made to generalize the relational model idea and improve the prediction ability.

A Proof of Proposition 5.1

Note that $\frac{y_{g,s}}{\tilde{y}_{g,s}} = e^{(\mu_{g,s}-\alpha\mu_{f,s})+\varepsilon_{g,s}-\beta\varepsilon_{x,s}}$, and the first order conditions are

$$\begin{aligned}
\frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\alpha} &= 2(-\mu_{f,s})\mathsf{E}\left(1-e^{(\mu_{g,s}-\alpha\mu_{f,s})+\varepsilon_{g,s}-\beta\varepsilon_{x,s}}\right)\cdot e^{(\mu_{g,s}-\alpha\mu_{f,s})+\varepsilon_{g,s}-\beta\varepsilon_{x,s}} = 0, \quad (33)\\
\frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\beta} &= 2(-\varepsilon_{x,s})\mathsf{E}\left(1-e^{(\mu_{g,s}-\alpha\mu_{f,s})+\varepsilon_{g,s}-\beta\varepsilon_{x,s}}\right)\cdot e^{(\mu_{g,s}-\alpha\mu_{f,s})+\varepsilon_{g,s}-\beta\varepsilon_{x,s}} = 0. \quad (34)
\end{aligned}$$

Equation (33) is equivalent to

$$\mathsf{E}\left(e^{\varepsilon_{g,s}-\beta\varepsilon_{x,s}} - e^{\mu_g-\alpha\mu_{f,s}+2\varepsilon_{g,s}-2\beta\varepsilon_{x,s}}\right) = 0. \quad (35)$$

Note that $\varepsilon_{x,s}|\varepsilon_{g,s} \sim N(\frac{\sigma_f}{\sigma_g}\rho_{gx}\varepsilon_{g,s}, (1-\rho_{gx}^2)\sigma_f^2)$, hence,

$$\begin{aligned}
\mathsf{E}\left(e^{\varepsilon_{g,s}-\beta\varepsilon_{x,s}}\right) &= \mathsf{E}\left(e^{\varepsilon_{g,s}}\mathsf{E}\left(e^{-\beta\varepsilon_{x,s}}|\varepsilon_{g,s}\right)\right)\\
&= \exp\left(\frac{1}{2}\beta^2\sigma_f^2 - \beta\sigma_f\sigma_g\rho_{gx} + \frac{1}{2}\sigma_g^2\right), \quad (36)
\end{aligned}$$

In addition,

$$\mathsf{E}\left(e^{\mu_g-\alpha\mu_{f,s}+2\varepsilon_{g,s}-2\beta\varepsilon_{x,s}}\right) = \exp\left(2\beta^2\sigma_x^2 - 4\beta\sigma_f\sigma_g\rho_{gx} + 2\sigma_g^2\right). \quad (37)$$

Therefore, according to Equation (35),

$$\alpha\mu_{f,s} = \frac{3}{2}\beta^2\sigma_x^2 - 3\beta\sigma_x\sigma_g\rho_{gx} + \mu_g + \frac{3}{2}\sigma_g^2. \quad (38)$$

Similarly, according to Equation (34), we have

$$\begin{aligned}
&\exp\left(\frac{1}{2}\sigma_x^2\beta^2 - \beta\sigma_g\sigma_f\rho_{gx} + \frac{1}{2}\sigma_g^2\right)(\sigma_x\sigma_g\rho_{gx} - \beta\sigma_x^2) \quad (39)\\
=\ &2\exp\left(2\sigma_x^2\beta^2 - 4\beta\sigma_g\sigma_f\rho_{gx} + 2\sigma_g^2\right)(\sigma_x\sigma_g\rho_{gx} - \beta\sigma_x^2), \quad (40)
\end{aligned}$$

indicating

$$\sigma_x \sigma_g \rho_{gx} - \beta \sigma_x^2 = 0. \tag{41}$$

Combining Equation (38) and Equation (41), we have

$$\alpha = \left(\frac{3}{2} \sigma_g^2 \rho_{gx}^2 - 3\sigma_g^2 \rho_{gx}^2 + \mu_g + \frac{3}{2} \sigma_g^2 \right) / \mu_{f,s}, \tag{42}$$

$$\beta = \frac{\sigma_g}{\sigma_f} \rho_{gx}. \tag{43}$$

References

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

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

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

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

Goshtasby, A. A., 2012. Image Registration. Springer, Ch. 2: Similarity and Dissimilarity Measures.

Ker, A., Coble, K., 2003. Modeling conditional yield densities. American Journal of Agriculture Economics 85 (2), 291–304.

Miranda, M. J., 1991. Area yield crop insurance reconsidered. American Journal of Agricultural Economics 73, 233–242.

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

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

Skees, J. R., Reed, M. R., 1986. Rate making for farm-level crop insurance: Implications for adverse selection. American Journal of Agriculture Economics 68 (3), 653–659.

Woodard, J. D., Pavlista, A. D., Schnitkey, G. D., Burgener, P. A., Ward, K. A., 2012. Governement insurance program design, incentive effects, and technology adoption: The case of skip-row crop insurance. American Journal of Agriculture Economics 94 (4), 823–837.