SFB 649 Discussion Paper 2013-030

Can expert knowledge compensate for data scarcity in crop insurance pricing?

Zhiwei Shen*
Martin Odening*
Ostap Okhrin*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Can expert knowledge compensate for data scarcity in crop insurance pricing?[a]

Zhiwei Shen[b,c], Martin Odening[c], Ostap Okhrin[d]

Abstract: Although there is an increasing interest in index-based insurances in many developing countries, crop data scarcity hinders its implementation by forcing insurers to charge higher premiums. Expert knowledge has been considered a valuable information source to augment limited data in insurance pricing. This article investigates whether the use of expert knowledge can mitigate model risk which arises from insufficient statistical data. We adopt the Bayesian framework that allows for the combination of scarce data and expert knowledge, to estimate the risk parameter and buffer load. In addition, a benchmark for the evaluation of expert information is created by using a richer dataset generated from resampling. We find that expert knowledge reduces the parameter uncertainty and changes the insurance premium in the correct direction, but that the effect of the correction is sensitive to different strike levels of insurance indemnity.

Keywords: expert knowledge, data scarcity, crop insurance pricing, Bayesian estimation

JEL: C14, Q19

---

[a] The financial support from the Deutsche Forschungsgemeinschaft via CRC 649 “Economic Risk”, Humboldt-Universität zu Berlin is gratefully acknowledged.

[b] Zhiwei Shen (Corresponding author) E-mail: zhiwei.shen@agrar.hu-berlin.de

[c] Department of Agricultural Economics, Humboldt Universität zu Berlin, Invalidenstr. 110, 10115 Berlin, Germany.

[d] Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Economics, Humboldt Universität zu Berlin, Spandauer Strasse 1, 10178 Berlin, Germany.

1 Introduction

The ability to quantify risk exposure is a well-known prerequisite of insurability. In an ideal case, a rich set of data is available to precisely estimate the likelihood and size of insurance losses (e.g., Berliner 1982; Schmit 1986). Unfortunately, this prerequisite is frequently violated in actuarial practice. For example, loss distributions for rare (catastrophic) events are per se difficult to quantify because of their small number of occurrences. Moreover, for new insurance products the damage history may be limited. The lack of historic data may lead to an incorrect estimation of loss distributions and introduces a new source of risk into the insurer's decision problem (Courbage and Liedtke 2003). Accordingly, actuaries tend to charge higher premiums than they would have, had the risks been well-specified (Kunreuther et al. 1995). This, in turn, may deter potential customers from buying insurance.

An insurance area, which is frequently plagued by limited data availability, is index-based crop insurance for agricultural producers. Index-based crop insurance has attracted the interest of researchers and practitioners because it can bypass obstacles inherent to traditional crop insurance. Payoffs are based on an easily observable index, such as an average area yield (instead of individual farm yields). Thus, index-based crop insurance is not subject to moral hazard and loss adjustment is straightforward. By construction, however, crop yields can only be observed once in a harvest period, i.e., typically once a year. As a result, the estimation of loss distributions in a particular region is usually based on a times series of 50 observations at most. In developing countries or emerging economies, reliable yield data are even scarcer and this data limitation hinders the implementation of area yield crop insurance despite the huge potential demand for yield risk reduction (Miranda and Farrin 2012).

Several options to mitigate the problem of data scarcity and model risk have been discussed in the literature. First, Odening, Musshoff and Xu (2007) suggest to use a weather index derived from daily observations of temperature and rainfall rather than an average area yield. The use of daily weather data increases the precision of the index estimation, but unfortunately the basis risk of the insurance product increases significantly because individual yield losses are not perfectly correlated with a weather index. In fact, Carter, Barrett and Trivelli (2007) estimate that farmers' willingness to pay for area-yield insurance is twice as high for rainfall insurance. Another incorporation of weather data in crop insurance rating is to use a long set of weather data as additional information to form empirical frequency priors in a Bayesian estimation of a loss-cost ratio density, so that weights can be adjusted and the insurance premium can be derived for individual years (Borman et al. 2013). Second, plant growth models have been employed to simulate the impact of risk factors on crop yields (Deng et al. 2008); however, these models are complex and difficult to calibrate and also contain a lot of parameters which have to be estimated for each region separately. Third, the use of expert knowledge has been proposed as a general response to cope with poor statistical data in bank risk management and insurance pricing (Alderweireld, Garcia, and Léonard 2006; Biener 2013). In the past, expert knowledge for the quantification of insurance risk was mainly used on an ad hoc basis without invoking a formal mathematical framework (Shevchenko and Wüthrich 2006). Lambrigger, Shevchenko and Wüthrich (2007) developed a Bayesian model that allows for a combination of observational data and expert opinions in a more formal and rigorous way. Applications in the area of insurance pricing, however, are rare. Arbenz and Canestraro (2012) take up the modeling approach of Lambrigger, Shevchenko and Wüthrich (2007) and combine loss observations with expert opinions for the estimation of fire insurance claims. They show that the combination of different sources of information can significantly reduce parameter uncertainty. In this article, we pursue a similar approach to assess the value of expert opinions in the context of pricing area yield insurance contracts.

A contribution of our analysis is that, in contrast to previous research, we are able to derive a benchmark for the evaluation of expert information. This is important since subjective judgments on loss probabilities may be erroneous and could add noise to the estimation of loss distributions (e.g., Kynn 2008). Here, we rely on disaggregated yield data that are usually unavailable to insurance companies when they design area yield insurance contracts. The use of this richer data set allows for the assessment of the informational value of external expert knowledge.

The estimation procedure is applied to area yield insurance for rice producers in three provinces of China (Heilongjiang, Jilin, and Liaoning). China is chosen since it is one of the world's largest agricultural producers and its farmers are exposed to pronounced yield risk. The three provinces under consideration are the main production areas for grain and have a vital role for domestic food supply and food security in China. According to Turvey and Kong (2010), there is a high market potential for index-based insurance products in China. To assess the viability of private crop insurance we calculate insurance premia for a hypothetical area yield insurance that takes into account systemic yield risk among three selected provinces via an appropriate risk premium (buffer load). Following Okhrin, Odening and Xu (2012), the spatial dependence of crop yields in the three provinces is modeled by a multivariate copula. We focus on the estimation of the stochastic dependence of yield losses in space since these dependencies cause systemic risk for insurers which is considered to be the most important obstacle for the implementation of index-based crop insurance (Miranda and Glauber 1997; Duncan and Myers 2000).

The remainder of the article is organized as follows. The next section describes our theoretical model: After an introduction of the insurance pricing framework, we briefly review vine copulas as an instrument to capture high dimensional stochastic dependence of area yields. Thereafter, we describe a Bayesian inference model that allows for the incorporation of expert knowledge into the estimation of posteriori loss distribution functions. In subsequent section, these models are applied to Chinese rice yield data. We present risk premia for a hypothetical area yield insurance and analyze the effect of expert knowledge in different scenarios. Moreover, we apply a resampling strategy to calculate an empirical loss distribution to evaluate experts' opinions. The last section provides conclusions on the benefit of expert knowledge in insurance pricing and offers suggestions for further research.

2 Theoretical Framework

Here we take the supply oriented view of an insurance company that wants to fix the price of insurance contracts so that insurance premia cover indemnity payments in each time period at a given confidence level. This objective requires charging risk premia, which take into account covariate risk between the insurance contracts, on top of the actuarial fair price. Following Wang and Zhang (2003) and Okhrin, Odening and Xu (2012), we evaluate the systemic risk of an insurance portfolio by calculating the buffer fund (BF). The BF is the value at risk (VaR) of the total net losses of a portfolio held by the insurer. It indicates the amount of financial reserve needed to prevent ruin from indemnity payments. Formally stated:

(1) $$BF = \inf\left\{ l \in \Re : \mathrm{P}\left(\sum_{i=1}^{d} w_i \cdot (L(X_i) - \pi_i) \geq l \right) = 1 - \alpha \right\},$$

where $L(X_i)$ denotes the indemnity payment for the ith contract. In the context of area yield insurance index: i represents regions; π_i is the fair insurance premium defined by $E[L(X_i)]$; w_i refers to the weight of the ith contract; and $1 - \alpha$ is the ruin probability. Dividing the BF by the number of contracts gives the buffer load, $BL = BF / \sum_{i=1}^{d} w_i$, which is the risk loading above the fair premium. The simplified premium is based on the following assumptions: first, diversification of products of the insurer is not taken into account; second, only a single-period model is considered and equity reserves accumulated in years with premium surpluses are ruled out; third, administrative costs are ignored.

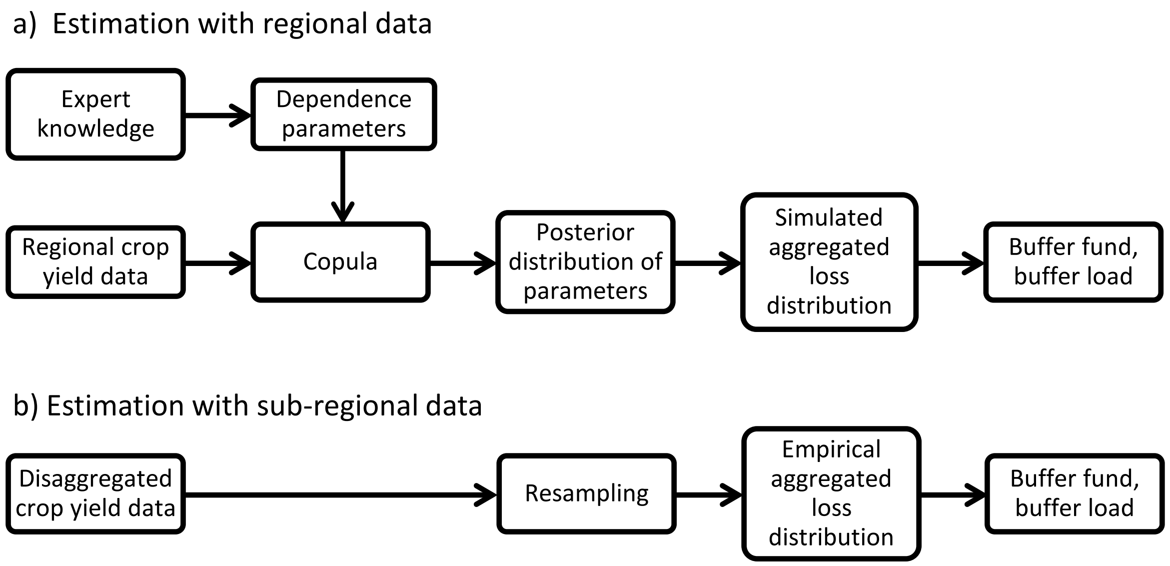
In section 3, we calculate the risk premium for an area yield insurance, which resembles a put option with the following indemnity payment:

(2) $\quad L(X_i) = T \cdot \max(K_i - X_i, 0),$

where X_i denotes the area yield which is the average crop yield within a production region, K_i is the strike level for the *i*th region which can be understood as the coverage of the insurance contract, and *T* is the tick value which converts physical units into monetary terms.

Calculating the BF by (1) requires knowledge of the joint loss distribution of all *n* contracts in the portfolio. Figure 1 provides a diagram of the estimation and validation procedure. Within a Bayesian estimation framework we use time series of area yields to estimate parametric marginal distributions and parametric copulas from which the joint distribution of indemnity payments in all regions can be derived. From the joint loss distribution, we simulate aggregated insurance losses and determine the BF as a quantile (VaR) of the total loss distribution. In an alternative scenario, the same estimation procedure is followed, yet the observational regional yield data is enhanced by expert opinions on the stochastic dependence of losses in the different trading areas (see figure 1a).

Figure 1. Estimation of buffer loads with different data sets



Comparing these two scenarios (with and without expert knowledge) shows the effect of the additional information on the estimated buffer loads. To evaluate the inclusion of expert knowledge, we contrast the results of the parametric estimation with quantiles from an empirical distribution of total (aggregate) losses. Clearly, the empirical loss distribution cannot be reliably estimated from a short time series of area yields. To overcome this problem, we utilize crop yield data which are observed on a disaggregated, sub-regional level (see figure 1b). Assuming that the sub-regions within a region are homogeneous, we calculate average area yields on the regional level by resampling from yield observations on the sub-regional level and then estimate an empirical loss distribution from these resampled data.

Unavailable time series observations of area yields are substituted by cross sectional observations of yields within the insured area. Using this empirical distribution as a benchmark may be criticized: One may ask why disaggregated data are not used for the parametric estimation of insurance losses, since, in practice, insurance companies would also make use of such data in their ratemaking if the data contain more information than aggregated yield data. Here we argue that access to disaggregated data is often more difficult than that to aggregated data. Thus, we assume that disaggregated yield data are not easily available for insurance companies. This assumption it is realistic if insurance suppliers contemplate entering new market segments, such as those in developing countries and transition economies.

In the following subsection, we describe the components of the estimation procedure in greater detail.

2.1 Modeling stochastic dependence of area yields with copulas

The advantage of using copulas arises from the decomposition of a multivariate distribution into margins and a pure dependency component. If *F* is an arbitrary *d*-dimensional continuous distribution function of the random variables $x_1, \dots, x_d$, then it can be decomposed into the associated copula C and its marginal distributions as

(3) $$F(x_1, x_2, \dots, x_d) = C_{\boldsymbol{\theta}}(F_{\psi_1}(x_1), F_{\psi_2}(x_2), \dots, F_{\psi_d}(x_d)),$$

where $\boldsymbol{\theta}$ is the copula parameter and F_{ψ_i} is the univariate continuous marginal distribution with an unknown vector of parameters ψ_i, see Sklar (1959). If *F* belongs to the class of elliptical distributions, then this results in a so-called elliptical copula. Note, however, that in many cases the function of the copula cannot be stated explicitly because the distribution function *F* and the inverse marginal distributions often have only integral representations. There are copula families which overcome this drawback: Archimedean copulas (see Nelsen 2006), a mixture of copula functions, hierarchical Archimedean copula (see Okhrin, Okhrin and Schmid 2013), and pair copula constructions which are more widely known as vines (see Min and Czado 2010; Aas et al. 2009). In this article, we concentrate on vine copulas since they are a natural choice when working in a Bayesian framework. When a multivariate distribution is decomposed into copula and marginal distributions, the multivariate density $f(\cdot)$ can be represented as

$$f(x_1, x_2, \dots, x_d) = c_{\boldsymbol{\theta}}\left(F_{\psi_1}(x_1), F_{\psi_2}(x_2), \dots, F_{\psi_d}(x_d)\right) \cdot f_{\psi_1}(x_1) \cdot f_{\psi_2}(x_2) \cdots f_{\psi_d}(x_d),$$

where $c(\cdot)$ is the copula density defined as $c(u,v) = \partial^2 C(u,v)/\partial u \partial v$ and f_{ψ_i} are the univariate marginal densities.

From the Bayes rule, the joint density can be factorized as

(4) $$f(x_1, x_2, \dots x_d) = f(x_d) \cdot f(x_{d-1}|x_d) \cdot f(x_{d-2}|x_{d-1}, x_d) \cdots f(x_1|x_2, \dots, x_d).$$

The conditional density for two dimensions, using (3) for the first pair, is given by

(5) $$f(x_1|x_2) = c_{12}\left(F_{\psi_1}(x_1), F_{\psi_2}(x_2)\right) \cdot f_{\psi_1}(x_1).$$

Similarly, for three dimensions of variables we have

(6) $$f(x_1|x_2,x_3) = c_{13|2}\left(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right) \cdot f(x_1|x_2)$$
$$= c_{13|2}\left(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right) \cdot c_{12}\big(F_1(x_1), F_2(x_2)\big) \cdot f_1(x_1).$$

Therefore, the joint density (4) can be decomposed into the appropriate pair-copula multiplied by a conditional marginal density using the following general formula (e.g., Aas et al. 2009):

(7) $$f(x|\boldsymbol{v}) = c_{xv_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\} \cdot f(x|\boldsymbol{v}_{-j}),$$

for a d-dimensional vector $\boldsymbol{v}$. Here, v_j is one component of vector $\boldsymbol{v}$ and $\boldsymbol{v}_{-\mathrm{j}}$ refers to the $\boldsymbol{v}$-vector without the v_j component.

Using the conditional marginal density, we can calculate the following conditional distribution of the form $F(x|\boldsymbol{v})$, for every *j*:

(8) $$F(x|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\}}{\partial F(v_j|\boldsymbol{v}_{-j})}.$$

For three dimensions of variables, the vector $\boldsymbol{v}$ is univariate and (8) simplifies to

(9) $$F(x|\boldsymbol{v}) = \frac{\partial C_{xv}\{F(x), F(v)\}}{\partial F(v)}.$$

Thus, the joint density distribution (4) for three dimensions can be expressed as

(10) $$f(x_1,x_2,x_3) = f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot c_{12}\big(F_1(x_1), F_2(x_2)\big)$$
$$\cdot\, c_{23}\big(F_2(x_3), F_2(x_3)\big) \cdot c_{13|2}\left(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right).$$

An extension to arbitrary large dimensions using C, D, or R-vines construction follows the same logic (cf. Kurowicka and Joe 2010). The representation for d-dimensional vine is the product of d marginal densities and $d(d-1)/2$ bivariate copulas.

2.2 Bayesian Copula Estimation with Expert Knowledge

The estimation of the copula parameters and the parameters of the marginal distribution is accomplished within a Bayesian estimation procedure. Bayesian estimation allows for the incorporation of expert knowledge (Zhang and Dukic, 2012) and has been successfully applied to vine copula estimation (Hofmann and Czado 2011; Min and Czado 2010).

According to Bayes rule, the posterior density $p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathcal{O})$ of parameters $(\boldsymbol{\theta},\boldsymbol{\psi})$ for a given observation set $\mathcal{O}$ is

(11) $$p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathcal{O}) \propto p_O(\mathcal{O}|\boldsymbol{\theta},\boldsymbol{\psi}) \cdot p_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{\theta},\boldsymbol{\psi}),$$

where $p_O(\mathcal{O}|\boldsymbol{\theta},\boldsymbol{\psi})$ is the likelihood function and $p_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{\theta},\boldsymbol{\psi})$ is the prior density. The latter may reflect non observational information, such as from regulatory guidelines.

Next we incorporate expert knowledge. Since our analysis focuses on the estimation of systemic risk, experts are only asked to provide estimates of copula parameters while the estimation of marginal parameters ψ_i is solely based on observed data. Let $\mathcal{E}$ denote a set of point estimates $\widehat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ of copula parameters $\boldsymbol{\theta}$, $k = 1, \ldots, K$ provided by *K* experts. In Section 3.2, we describe how point estimates of copula parameters were elicited from expert opinions in detail. We assume that conditionally on $\boldsymbol{\theta},\boldsymbol{\psi}$, the observation set $\mathcal{O}$ is independent of the set of expert point estimates $\mathcal{E}$. The joint conditional density of observations and experts' point estimates can thus be written as

(12) $$p_{\mathcal{O},\mathcal{E}}(\mathcal{O},\mathcal{E}|\boldsymbol{\theta},\boldsymbol{\psi}) = p_{\mathcal{O}}(\mathcal{O}|\boldsymbol{\theta},\boldsymbol{\psi}) \cdot p_{\mathcal{E}}(\mathcal{E}|\boldsymbol{\theta},\boldsymbol{\psi}).$$

The posterior density for parameters $(\boldsymbol{\theta},\boldsymbol{\psi})$ now extends to:

(13) $$\begin{aligned} p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathcal{O},\mathcal{E}) &\propto p_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{\theta},\boldsymbol{\psi}) \cdot p_{\mathcal{O},\mathcal{E}}(\mathcal{O},\mathcal{E}|\boldsymbol{\theta},\boldsymbol{\psi}) \\ &= p_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{\theta},\boldsymbol{\psi}) \cdot p_{\mathcal{O}}(\mathcal{O}|\boldsymbol{\theta},\boldsymbol{\psi}) \cdot p_{\mathcal{E}}(\mathcal{E}|\boldsymbol{\theta},\boldsymbol{\psi}). \end{aligned}$$

Below, we elaborate on the three components that constitute (13).

Recalling the copula model introduced in the previous section, the likelihood function $p_{\mathcal{O}}(\mathcal{O}|\boldsymbol{\theta},\boldsymbol{\psi})$ is conditionally imposed on the copula parameters $\boldsymbol{\theta}$ and the marginal parameters $\boldsymbol{\psi} = (\psi_1, \dots, \psi_d)$ for *d*-dimension $(x_1, \dots x_d)$ of area average yields and is given by:

(14) $$\begin{aligned} p_{\mathcal{O}}(\mathcal{O}|\boldsymbol{\theta},\boldsymbol{\psi}) &= L(x_1, \dots, x_d|\boldsymbol{\theta},\boldsymbol{\psi}) \\ &= \prod_{n=1}^{N} [c\big(F_{\psi_1}(x_{1,n}), \dots, F_{\psi_d}(x_{d,n})\big|\boldsymbol{\theta}\big) \cdot \prod_{i=1}^{d} f_{\psi_i}(x_{i,n}|\psi_i)]. \end{aligned}$$

In general, experts' estimates may be stochastically dependent, particularly if experts have the same informational background. However, we ignore this kind of dependence since it would be difficult to quantify. Moreover, we assume the experts are homogeneous. Under the simplifying assumptions, the likelihood of *K* experts' point estimates can be stated as:

(15) $$p_{\mathcal{E}}(\mathcal{E}|\boldsymbol{\theta},\boldsymbol{\psi}) = \prod_{k=1}^{K} g(\widehat{\boldsymbol{\theta}}_k|\boldsymbol{\theta}),$$

where $g(\cdot\,|\boldsymbol{\theta})$ describe the conditional density of each expert's point estimates. Following Lambrigger, Shevchenko and Wüthrich (2007), using a normal distribution to model expert opinions we use a truncated normal distribution to fit $g(\cdot\,|\boldsymbol{\theta})$ due to the bounded range of copula parameters. Like Arbenz and Canestraro (2012), we assume that experts' point estimates are conditionally unbiased, i.e., $E\big(\hat{\theta}_k\big|\theta\big) = \theta$ with identical variance σ^2, i.e., $(var\big(\hat{\theta}_k\big|\theta\big) = \sigma^2)$. Further assuming independence $p_{\boldsymbol{\theta},\boldsymbol{\psi}}(\boldsymbol{\theta},\boldsymbol{\psi}) = p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \cdot p_{\boldsymbol{\psi}}(\boldsymbol{\psi})$ for the prior density and inserting (14) and (15) into (13) yields:

(16) $$\begin{aligned} p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathcal{O},\mathcal{E}) \propto \prod_{n=1}^{N} & [c(F_{\psi_1}(x_{1,n}), \dots, F_{\psi_d}(x_{d,n})|\boldsymbol{\theta}) \\ & \cdot \prod_{i=1}^{d} f_{\psi_i}\big(x_{i,n}\big|\psi_i\big)] \cdot p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \cdot p_{\boldsymbol{\psi}}(\boldsymbol{\psi}) \cdot \prod_{k=1}^{K} g\big(\widehat{\boldsymbol{\theta}}_k\big|\boldsymbol{\theta}\big). \end{aligned}$$

Estimation of (16) proceeds in three steps using the inference of margins method (Joe 1997). However, the difference from previous literature is that copula parameters are estimated using a Bayesian framework instead of maximum likelihood (also see Bokusheva 2011). First, the margins of each variable are obtained by fitting a parametric distribution to the empirical data. In other words, the marginal parameters are treated as given in a Bayesian framework. Hence, (16) will become

(17) $$p(\boldsymbol{\theta}|\mathcal{O},\mathcal{E}) \propto \prod_{n=1}^{N} c\big(F_{\widehat{\psi_1}}(x_{1,n}), \dots, F_{\widehat{\psi_d}}(x_{d,n})\big|\boldsymbol{\theta}\big) \cdot p(\boldsymbol{\theta}) \cdot \prod_{k=1}^{K} g\big(\widehat{\boldsymbol{\theta}}_k\big|\boldsymbol{\theta}\big).$$

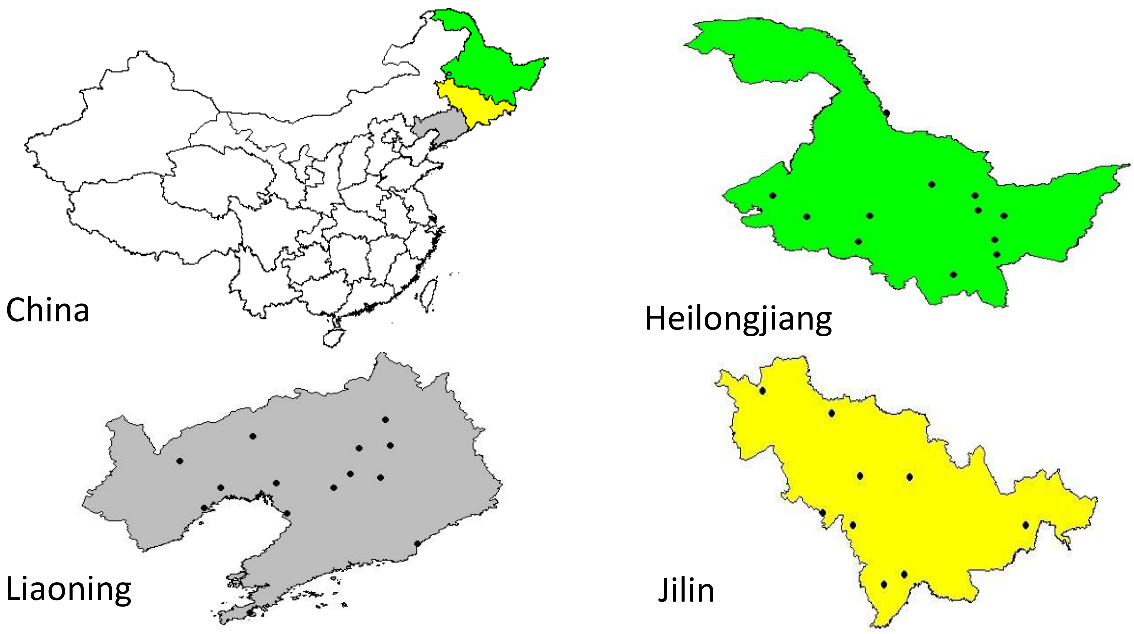
Second, the structure and copula types of each pair in $c(\cdot\,|\boldsymbol{\theta})$ are then determined given the estimated margins $F_{\widehat{\psi_d}}$. Third, once the likelihood function of the vine copula and the prior density are determined, the full posterior distributions of copula parameters are obtained by Markov Chain Monte Carlo (MCMC).

With the posterior distribution of copula parameters at hand, we are able to compute the distribution of aggregated insurance losses through nested Monte Carlo simulations. Based on one copula parameter from the posterior distribution, we generate a sample of 10,000 values to yield a distribution of total net loss. Then, the corresponding buffer fund and risk premium are obtained according to the predetermined ruin probability. This procedure is repeated for any value drawn from the posterior distribution of copula parameters to yield a posterior density for the buffer fund and risk premium.

3 Area Yield Insurance in Northeast China

In this section, the estimation procedure outlined before is used to quantify risk premia of rice yield insurance in three provinces in Northeast China: Heilongjiang, Jilin, and Liaoning. These provinces cover about 787,300 km^2 and have a vital role for domestic food supply and food security since they are the main grain production areas in China. Grain production in the selected areas is seriously affected by weather risk, in particular drought. Between 2004 and 2006, 16,805 km^2 in Heilongjiang and 11,561 km^2 in Jilin were hit by drought (China Meteorological Administration 2008).

Figure 2. Map of the study area



[*] The black dots mark the capitals of the prefectures within a province.

For companies to design insurance in the study area, it is essential to quantify the systemic risk inherent in an area yield insurance designed for each of the three provinces. With regard to the information set available to the insurer, we assume that only provincial crop yield data are available. In our study, however, we have access to sub-regional crop data. This allows us to assess the informational value of expert knowledge. The sub-regional rice yield data refer to the administrative level of a prefecture level and cover the period between 1994 and 2009. These data were collected from Heilongjiang Statistical Yearbook (1995-2010), Jilin Statistical Yearbook (1995-2010), and Liaoning Statistical Yearbook (1995-2010). The locations of the three provinces as well as their prefectures are depicted in figure 2. Provincial rice yields are calculated as weighted averages of prefecture yields. Descriptive statistics of the yield data are presented in the appendix. The expected values and standard variances of prefecture-level yields in Heilongjiang range from 54.55 dt/ha to 79.96 dt/ha and from 7.55 dt/ha to 16.24 dt/ha. Moreover, the size of the cultivated rice area in the prefectures varies considerably. Similar variation can be found in Jilin and Liaoning. After detrending, parametric distributions were fitted to the yield data. Goodness-of-fit tests (i.e., the Kolmogorov-Smirnoff test, χ^2 and Anderson-Darling test) suggest a logistic distribution for all three provinces.

3.1 Elicitation of dependence parameters from experts

Acquisition of expert knowledge and its translation into model parameters is a nontrivial task that requires a sound statistical and psychological approach. Here, we take up the approach of Böcker, Crimmi and Fink (2010) who suggest to turn expert opinions into dependence parameters by means of indirect questions about conditional and joint probabilities of loss events. The procedure consists of successive questions on bivariate relations of random variables and joint probability for all three variables. After experts estimated the marginal and joint probabilities (e.g., $P(X \leq x)$ and $P(X \leq x, Y \leq y)$) in all three provinces and a certain copula model has been determined from the observations, we are able to estimate the bivariate copula parameters for all bivariate variables using (3). These steps are conducted for the conditional marginal probabilities and the conditional joint probabilities of area yields in all three provinces.

Ten experts from insurance and academics with experience in agricultural insurance in China were interviewed in a written survey. As mentioned above, we focus on the elicitation of dependence parameters (marginal probabilities were given to the experts). To elicit the joint loss probabilities, the following question was asked:

"What is your estimate of the joint probability that a shortfall of average rice yield among all the farmers which occurs less than once per decade is observed in both Heilongjiang and Jilin in the same year?"

Similar questions refer to other combinations of the three insurance regions. This formulation avoids specifying the yield in absolute terms. The answers to these questions represent the probability $P(X \leq x, Y \leq y) = C_\theta\big(F_X(x), F_Y(y)\big)$ where $P(X \leq x) = P(Y \leq y) = 0.1$.[1] Table 1 presents results from the expert survey. The averages of estimated joint probabilities are about 3 percent for all three bivariate combinations. The joint loss probability for all three regions is only half as high. We also display the standard deviation to indicate the variation of expert opinions. The coefficient of variation for the joint probability of losses in Heilongjiang and Liaoning amounts to 55 percent and is almost twice as high as that for the two other pairs. Obviously, experts are more ambiguous with regard to the likelihood of joint losses in these two provinces. It may be that the larger distance between Heilongjiang and Liaoning complicates the estimation.

[1] Eliciting more probabilities will likely decrease the willingness of experts to participate in the survey as well as the consistency of the answers. Moreover, when copula structure is predetermined, the joint probability evaluated at one point is sufficient to derive the dependence parameter of copula function.

Table 1. Expert Estimates of Joint Probabilities of Rice Yields in three Provinces

	Heilongjiang and Jilin	Jilin and Liaoning	Heilongjiang and Liaoning	Heilongjiang, Jilin, and Liaoning
Mean	0.027	0.029	0.030	0.014
Standard deviation	0.006	0.009	0.016	0.008

Using the elicited joint probabilities and assuming a particular copula type, the estimate of copula parameter $\hat{\theta}_k$ can be derived according to (3). The mean $E\left(\hat{\theta}_k \middle| \theta\right) = \theta$ and the variance ($var\left(\hat{\theta}_k \middle| \theta\right) = \hat{\sigma}^2$) of the distribution of copula parameter estimates are obtained as sample estimates from the 10 experts, i.e.:

$\hat{\sigma}^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\hat{\theta}_k - \bar{\theta})^2$, $\bar{\theta} = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_k$, respectively.[2]

3.2 Spatial dependence of area yields

Prior to the estimation of the copula parameters, the structures of the vine copula and the copula family for each pair have to be determined. The structure of vine copula usually refers to the decomposition type (C-vines or D-vines) and the order in which the variables enter the vine copula function. Since both structures are identical for three random variables, we do not need to account for different specifications in our application. Following Aas et al. (2009), we determine the order of regions such that their pairwise dependence is maximized on the first level of the tree. The strength of dependency is measured by Kendall's tau (table 2). This leads to the following order: Jilin (1)-Liaoning (2)-Heilongjiang (3). The copula family for each pair is chosen according to the Akaike Information Criterion. First, the copula type is determined for the unconditional bivariate copulas c_{12} and c_{23} in the vine copula (10). We find that the Frank copula has the best fit. Based on this specification, we calculate the conditional margins ($F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)$) and identify the best fitting copula family $c_{13|2}$ for the joint conditional margins which is a Gaussian copula.

Once the structure of vine copula and the copula families for each pair have been specified, we estimate the copula parameters using the Bayesian model (17). An uninformative prior is assumed, that is $p(\boldsymbol{\theta})$ is a uniform density. The posterior distribution of the parameters is obtained by using the Metropolis-Hasting algorithm with Gaussian random walk proposals which are bound in the domain of the copula parameters. For instance, in case of a Gaussian pair copula, the truncated random walk proposal is bound to [-1, 1]. We ran 30,000 iterations in three parallel chains, discarding a burn-in of 20,000 in order to achieve the appropriate convergence (the potential scale reduction factors of Gelman and Rubin (1992) were below 1.1 for all of the parameters). The resulting posterior means and standard deviations for the copula parameters with and without expert knowledge are presented in table 2. We also depict Pearson's linear correlation, which is often used as a standard measure for stochastic dependence in crop insurance despite its potential pitfalls (e.g., Goodwin 2001; Wang and Zhang 2003)

[2] Alternative approaches to calculate estimates of σ^2 are discussed in Arbenz and Canestraro (2012).

Table 2. Estimation of Dependences among three Provinces with Standard Deviation

Dependence measure	Dependence parameters		
	Jilin - Liaoning	Liaoning - Heilongjiang	Jilin - Heilongjiang*
Pearson's rho	0.19 (0.24)	0.50 (0.19)	0.26 (0.23)
Kendall's tau	0.20 (0.19)	0.47 (0.19)	0.07 (0.19)
Copula without experts	1.48 (1.55)	4.77 (1.76)	0.14 (0.10)
Copula with experts	3.63 (0.61)	4.29 (1.14)	0.36 (0.05)

* For the vine copula, the dependence between Jilin and Heilongjiang is conditional on Liaoning.

All dependency measures presented in table 2 indicate that Liaoning and Heilongjiang have the highest correlation among the three provinces, though their geographical distance is the largest one. The dependence between average rice yields in the other two pairs of provinces is positive, but considerably smaller. Note that Pearson's rho and Kendall's tau differ largely for Jilin and Heilongjiang. Comparing the last two rows in table 2 reveals that the inclusion of expert knowledge has a significant impact on the estimation of the copula parameters. Apparently, experts believe that the stochastic dependence between rice yields in Heilongjiang and Jilin as well as in Jilin and Liaoning are higher than reflected by the yield data – for Liaoning and Heilongjiang, the copula parameter estimate is slightly lower. It can be conjectured that experts' estimations were influenced by knowledge about the geographic location of the regions, such that adjacent locations are presumed to show higher correlations. These findings show that the consideration of expert knowledge leads to a reduction of the standard deviation of parameter uncertainty. This finding confirms results from previous studies, e.g., Arbenz and Canestraro (2012).

3.3 Resampling

To derive a loss distribution that can serve as a benchmark for the experts' estimates, we resort to a resampling procedure. Rice yields measured in sub-regions (prefectures) are used to generate further provincial yield data and to calculate the empirical loss distribution and empirical buffer load. Heilongjiang, Jilin, and Liaoning provinces consist of 12, 9, and 14 sub-regions, respectively. Provided that the sub-regions in each province are homogeneous, weighted average rice yield of any combinations of sub-regions can be regarded as a realization of provincial yield (in the same year). To test the assumption of homogeneity of rice yields, we apply a robust Levene's test (Levene 1960), an *F*-test and a Kruskal-Wallis test (Kruskal and Wallis 1952) to the detrended sub-regional rice yield data. The results in table 3 show that the null hypotheses of equal means, equal medians and equal variances within the provinces cannot be rejected, supporting our resampling approach.

Table 3. Results for testing homogeneity

Test for	Variance		Mean		Median	
	Levene's statistic	*p*-value	*F*-statistic	*p*-value	Kruskal-Wallis χ^2	*p*-value
Heilongjiang	0.998	0.450	0.000	0.999	1.118	0.999
Jilin	1.401	0.202	0.000	0.999	3.680	0.885
Liaoning	1.313	0.208	0.000	0.999	0.960	0.999

Resampling is then carried out by taking a weighted average of all combinations C_m^l of observed (detrended) sub-regional yields within a province. The combination $C_m^l = \frac{m!}{l!(m-l)!}$ refers to the number of subsets of l distinct elements of a set m. Here, m denotes the number of sub-regions in a province and l varies from 1 to m. This resampling is done for each province and year in the observation period. For instance, in Jilin, which consists of 9 sub-regions, $(C_9^9 + C_9^8 + \cdots C_9^1)$ = 511 combinations of weighted average yields are generated for each year. Likewise, 4,095 and 16,383 data are resampled for Heilongjiang and Liaoning, respectively.

From the resampled data, we calculate the weighted average net loss $\sum_{i=1}^{3} w_i (L(X_i) - \pi_i) / \sum_{i=1}^{3} w_i$ for area yield insurances spanning all three provinces. Indemnity payments $L(X_i)$ in three provinces are derived for two alternative strike levels K_i, the 50 and 30 percent quantiles of the respective area yields distributions. Without loss of generality, we set $T = 1$ so that losses can be interpreted in yield units, i.e., dt/ha. The weights w_i of the insurance contracts are chosen according to total rice areas (shown in the Appendix).

Figure 3. Cumulative distribution of weighted average net loss with different datasets

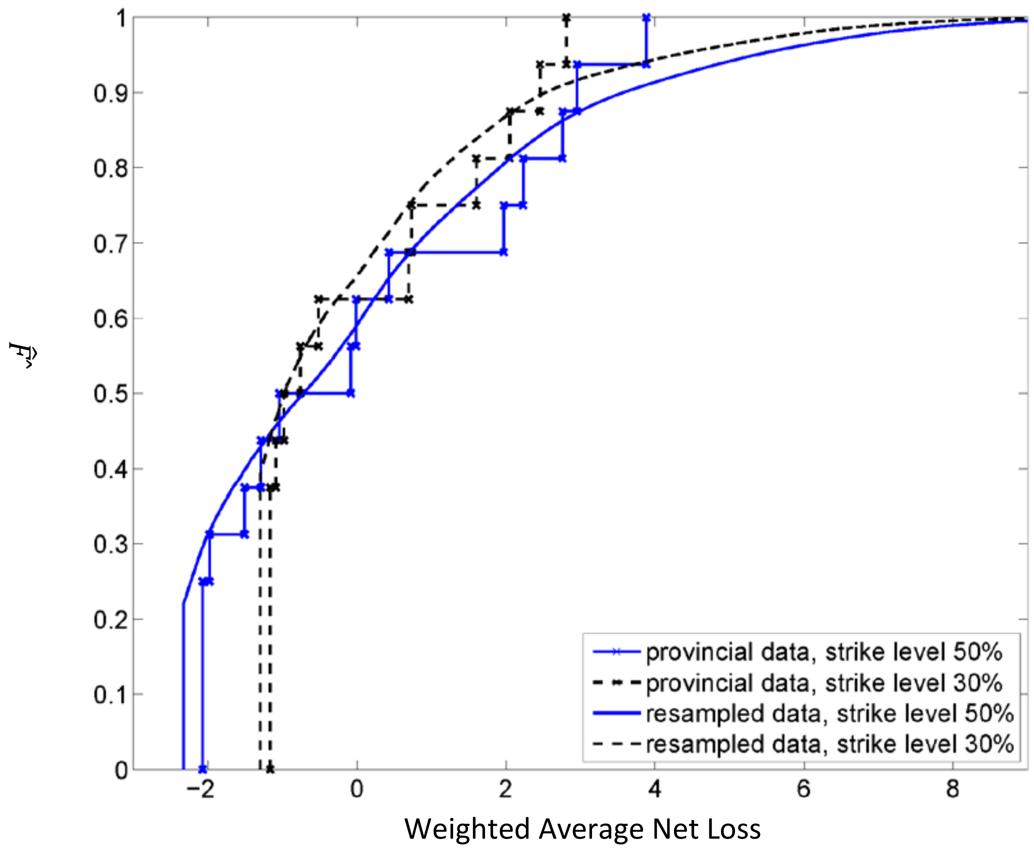


Figure 3 displays the resulting net loss distributions on a per hectare basis. Both distributions are positively skewed. For a strike level of 50 (30) percent, the probability for net losses below -2.33 dt (-1.30 dt) is zero. By construction, these values mark the weighted average fair insurance premium. The distribution is truncated at these values, because no indemnity payments accrue if rice yields exceed the 50 (30) percent quantile of the yield distribution. The 95 percent quantile of weighted net loss is 5.35 dt (4.28 dt) and maximum losses amount to 26 dt (24 dt).

Figure 3 also displays the empirical loss distributions based on only 16 observations of provincial yields, which indicate that these distributions differ from their resampling counterparts. First, the fair prices are smaller (2.07 dt and 1.17 dt, respectively). In addition, the tail risk is underestimated by the empirical distribution. This is due to the fact that the resampled data show more variation of area yields than the historical observations. These differences are translated into the risk premia of the area yield insurance (see table 4).

3.4 Estimation of Risk Premia

Table 4 summarizes estimates of fair prices and buffer loads based on the Bayesian estimation with and without expert knowledge, the empirical distribution, and the resampled distribution. Fair prices are displayed for each province individually, while buffer loads are calculated for an average insurance contract. This is appropriate in our analysis since we are primarily interested in estimating the insurer's size of loss exposure and not interested in how to reallocate risk to the insured.

Table 4. Fair Prices and Buffer Loads

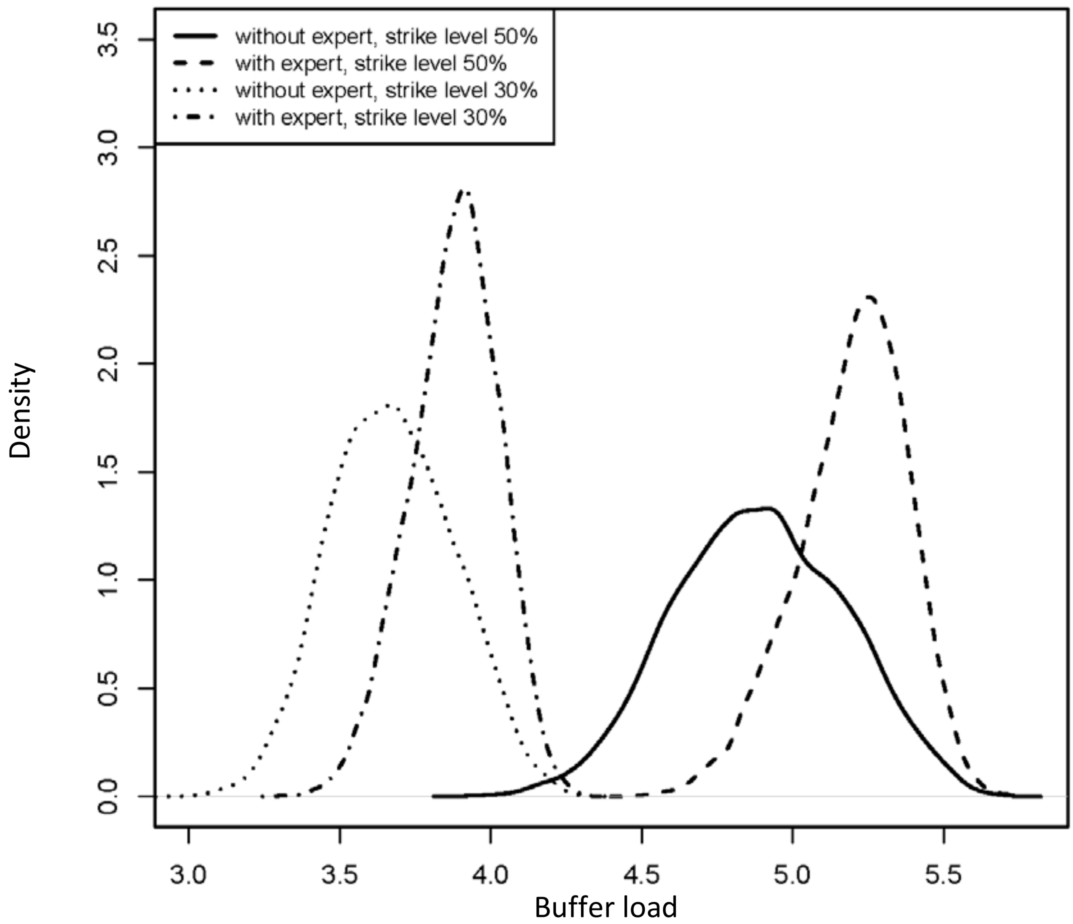
Scenarios	Fair Price				$BL_{0.90}$	$BL_{0.95}$
	Heilongjiang	Jilin	Liaoning	Weighted Average		
Empirical distribution based on provincial data						
Strike Level 50%	2.326	1.538	2.041	2.072	2.931	3.604
Strike Level 30%	1.290	0.793	1.286	1.166	2.415	2.703
Empirical distribution based on resampled sub-regional data						
Strike Level 50%	2.526	2.043	2.157	2.330	3.577	5.348
Strike Level30%	1.385	1.122	1.298	1.302	2.531	4.278
Bayesian copulas estimation based on provincial data						
Strike Level 50%						
without expert	2.103	1.663	2.068	1.987	3.399 (0.216)*	4.890 (0.283)
with expert	2.103	1.664	2.069	1.987	3.607 (0.138)	5.196 (0.182)
Strike Level 30%						
without expert	1.082	0.856	1.065	1.023	2.261 (0.131)	3.674 (0.205)
with expert	1.082	0.856	1.064	1.023	2.374 (0.088)	3.880 (0.146)

* Standard deviations are presented in the parentheses.

According to the copula based estimation without expert knowledge, fair prices vary between 1.66 dt in Jilin and 2.10 dt in Heilongjiang for a coverage of 50 percent, which is about 3 percent of the average rice yield. The corresponding values for a coverage of 30 percent are 0.86 dt and 1.08 dt, respectively. The estimation with expert knowledge results in the same fair prices because fair prices depend on the marginal yield distributions and the latter are not influenced by the expert knowledge in this study. The buffer loads, however, are different, which is not surprising in light of differences in the copula parameter estimates. The buffer load for a 90 percent confidence level increases from 3.40 dt to 3.60 dt if expert knowledge is included in the Bayesian loss estimation. A similar difference occurs for a 95 percent confidence level, for which the buffer loads amount to 4.90 dt and 5.20 dt respectively for with and without expert. An increase of the estimated risk can also be observed for a coverage of 30 percent. Obviously, the buffer load is considerably higher than the fair price of the insurance, which may appear counterintuitive. One should recall, however, the static definition of the buffer load that we use in this study: The buffer load indicates the margin to the fair price that is required to build a desired financial reserve in a single period. In a multi-period setting, it is not necessary to charge the buffer load every year.

Table 4 provides further information about the dispersion of the estimated buffer loads. It can be seen that the variability of the Bayesian estimates becomes smaller if expert knowledge is considered. This finding is also supported by figure 4, which depicts the densities of the buffer loads resulting from the Bayesian copula estimation.

Figure 4. Estimated density of 95% buffer load, estimated from Vine copulas



How do these estimates relate to the benchmark distribution which we derived from the disaggregated yield data? With regard to fair prices, we find that the parametric approach underestimates these values for all three provinces and both strike levels. This is likely due to the fact that the estimation of marginal yield distributions is based on only 16 observations. As mentioned above, the inclusion of expert knowledge has no impact on the estimation of fair prices and thus cannot reduce the underestimation. The effect on the estimation of the buffer loads is somewhat different. Again, the parametric estimates without expert knowledge are smaller than the ones from the resampled distribution; however, taking into account expert knowledge reduces the underestimation and brings the estimated buffer loads closer to the benchmark. In one scenario (coverage 50 percent, confidence level 90 percent) the parametric buffer load is even slightly higher than the benchmark value. This means that in this case study, expert opinions change the insurance premium in the correct direction, but the size of the correction depends on the coverage and confidence level.

4 Conclusions

This study was motivated by the difficulty in assessing systemic yield risks in agricultural crop insurance due to the fact that yields are often observable only once a year and hence available time series data are usually short. The scarcity of yield observations hampers the application of data intensive statistical methods and may result in unreliable estimates of potential insurance losses and thus risk premia. To mitigate this problem, we provide a statistical framework that allows for the incorporation of expert knowledge on joint yield risks into a ratemaking procedure. A Bayesian estimation procedure is employed which allows for the explicit combination of prior information, yield observation, and expert opinions. The stochastic dependence of area yields in different regions is captured by a vine copula, which is able to capture high dimensional dependence structures. The modeling approach is applied to a hypothetical area yield insurance for rice producers in three provinces in Northeast China. We estimate a joint loss distribution for all provinces from which we derive risk premia. The results of the Bayesian estimation are compared with an empirical loss distribution that is generated by resampling from disaggregated yield data. We find that the inclusion of expert knowledge has a significant impact on the estimation results. Insurance experts estimate the probability of joint area yield risk in three provinces to be higher than that which is solely estimated from the yield data. This increase varies between 5 and 6.5 percent, depending on the confidence level and insurance coverage. Additional expert knowledge changes the insurance premium in the correct direction relative to a benchmark derived from sub-regional data, but the size of the correction is sensitive to the specification of the insurance products, e.g., the strike level. Moreover, expected losses, i.e., fair premia, are underestimated. This indicates that the estimated distribution of total insurance losses differs from the benchmark distribution even after taking expert knowledge into account. We conclude that the use of expert knowledge is not a panacea for data scarcity in crop insurance pricing, but that it has potential to mitigate this problem. This finding is relevant for insurers and reinsurers who intend to launch new insurance products, particularly in low income countries where demand for crop insurance is high, but crop yield data are rare. One should note, however, that our evaluation of the data augmentation procedure interferes with several subjective assumptions, such as the specification of marginal distributions, the choice of the copula type and structure, and the distribution of expert parameters. Thus, it is rather difficult to extract the “treatment effect” of including expert knowledge and generalizations of our specific results are not straightforward.

There are several possible extensions of this study. First, one might attempt to receive more information from experts. In our case, expert knowledge was only used to support the estimation of copula parameters, while the copula type and marginal yield distributions were determined by means of yield observations only. Second, the number of experts could be increased, for example, by asking agricultural specialists without an insurance background. Third, to cope with data scarcity, more objective information, in particular historical weather records, should also be considered together with expert knowledge in our Bayesian framework. Even though basis risk exists between weather data and crop yield, the stochasticity of crop yield is mainly determined by weather data, particularly for unfavorable weather events. Fourth, alternative procedures of eliciting probabilities from should be tested and compared. Finally, we suggest conducting further empirical studies to provide a clearer picture of the conditions under which expert knowledge is most helpful.

5 References

Aas, K., C. Czado, A. Frigessi, and H. Bakken. 2009. "Pair-copula Constructions of Multiple Dependence." *Insurance: Mathematics and Economics* 44(2): 182–198.

Alderweireld, T., J. Garcia, and L. Léonard. 2006. "A Practical Operational Risk Scenario Analysis Quantification." *Risk Magazine* 19(2): 93-95

Arbenz, P., and D. Canestraro. 2012. "Estimating Copulas for Insurance from Scarce Observations, Expert Opinion and Prior Information: A Bayesian Approach." *ASTIN Bulletin* 42(1): 271–290.

Berliner, B. 1982. *Limits of Insurability of Risks*. Prentice-Hall.

Biener, C. 2013. "Pricing in Microinsurance Markets." *World Development* 41(C): 132–144.

Böcker, K., A. Crimmi, and H. Fink. 2010. "Bayesian Risk Aggregation: Correlation Uncertaity and Expert Judgement." In *Rethinking Risk Measurement and Reporting Uncertainty*. Risk Books. London.

Bokusheva, R. 2011. "Measuring Dependence in Joint Distributions of Yield and Weather Variables." *Agricultural Finance Review* 71(1): 120–141.

Borman, J.I., B.K. Goodwin, K. Coble, T. Knight, and R. Rejesus. 2013. "Accounting for Short Samples and Heterogeneous Experience in Rating Crop Insurance." *Agricultural Finance Review* 73(1): 88-101.

Carter, M.R., C.B. Barrett, and C. Trivelli. 2007. "Concept Note: Area-Based Yield Insurance Pilot Project for Peruvian Coastal Agriculture." *CRSPs*. http://crsps.net/resource/concept-note-area-based-yield-insurance-pilot-project-for-peruvian-coastal-agriculture/.

China, Heilongjiang Bureau of Statistics. 1995-2010. *Heilongjiang Statistical Yearbook 1994-2009*. China Statistics Press.

China, Jilin Bureau of Statistics. 1995-2010. *Heilongjiang Statistical Yearbook 1994-2009*. China Statistics Press.

China, Liaoning Bureau of Statistics. 1995-2010. *Heilongjiang Statistical Yearbook 1994-2009*. China Statistics Press.

China Meteorological Administration.2008. *Weather Insurance Pre-feasibility Study, A report for Deutsche Gesellschaft für Technische Zusammenarbeit (GZT)*. China, Beijing.

Courbage, C., and P.M. Liedtke. 2003. "On Insurability, Its Limits and Extensions." *Insurance Research and Practice* 18(2): 44–49.

Deng, X., B.J. Barnett, G. Hoogenboom, Y. Yu, and A.G. Garcia. 2008. "Alternative Crop Insurance Indexes." *Journal of Agricultural and Applied Economics* 40(1): 223-237.

Duncan, J., and R.J. Myers. 2000. "Crop Insurance Under Catastrophic Risk." *American Journal of Agricultural Economics* 82(4): 842–855.

Gelman, A., and D.B. Rubin. 1992. "Inference From Iterative Simulation Using Multiple Sequences ", *Statistical Science* 7: 457-511.

Goodwin, B.K. 2001. "Problems with Market Insurance in Agriculture." *American Journal of Agricultural Economics* 83(3): 643–649.

Hofmann, M., and C. Czado. 2011. "Assessing the VaR of a Portfolio Using D-vine Copula Based Multivariate GARCH Models." working paper.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Kruskal, W.H., and W.A. Wallis. 1952. "Use of ranks in one-criterion variance analysis", *Journal of the American Statistical Association* **47**(260): 583–621

Kunreuther, H., J. Meszaros, R.M. Hogarth, and M. Spranca. 1995. "Ambiguity and Underwriter Decision Processes." *Journal of Economic Behavior & Organization* 26(3): 337–352.

Kurowicka, D., and H. Joe. 2010. *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Company.

Kynn, M. 2008. “The ‘Heuristics and Biases’ Bias in Expert Elicitation.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 717(1): 239-264

Lambrigger, D.D., P.V. Shevchenko, and M.V. Wüthrich. 2007. “The Quantification of Operational Risk Using Internal Data, Relevant External Data and Expert Opinions.” *Journal of Operational Risk* 2(3): 3–27.

Levene, H. 1960. “Robust testes for equality of variances.” In *Contributions to Probability and Statistics* (I. Olkin, ed.) 278–292.

Min, A., and C. Czado. 2010. “Bayesian Inference for Multivariate Copulas Using Pair-Copula Constructions.” *Journal of Financial Econometrics* 8(4): 511–546.

Miranda, M.J., and K. Farrin. 2012. “Index Insurance for Developing Countries.” *Applied Economic Perspectives and Policy* 34(3): 391–427.

Miranda, M.J., and J.W. Glauber. 1997. “Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets.” *American Journal of Agricultural Economics* 79(1): 206–215.

Nelsen, R.B. 2006. *An Introduction to Copulas*. Springer.

Odening, M., O. Musshoff, and W. Xu. 2007. “Analysis of Rainfall Derivatives Using Daily Precipitation Models: Opportunities and Pitfalls.” *Agricultural Finance Review* 67(1): 135–156.

Okhrin,O., Y. Okhrin, and W. Schmid. 2013. “Determining the structure and estimation of hierarchical Archimedean copulas.” *Journal of Econometrics* 173(2): 189-204

Okhrin, O., M. Odening, and W. Xu. 2012. “Systemic Weather Risk and Crop Insurance: The Case of China.” *Journal of Risk and Insurance*. doi:10.1111/j.1539-6975.2012.01476.x.

Schmit, J.T. 1986. “A New View of the Requisites of Insurability.” *The Journal of Risk and Insurance* 53(2): 320–329.

Shevchenko, P.V., and M.V. Wüthrich. 2006. “The Structural Modelling of Operational Risk via Bayesian Inference: Combining Loss Data with Expert Opinions.” *Journal of Operational Risk* 1(3):27-50

Sklar, A. 1959. Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* **8**: 229–231

Turvey, C.G., and R. Kong. 2010. “Weather Risk and the Viability of Weather Insurance in China’s Gansu, Shaanxi, and Henan Provinces.” *China Agricultural Economic Review* 2(1): 5–24.

Wang, H.H., and H. Zhang. 2003. “On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach.” *The Journal of Risk and Insurance* 70(1): 111–124.

Zhang, Y., and V. Dukic. 2012. “Predicting Multivariate Insurance Loss Payments Under the Bayesian Copula Framework.” *Journal of Risk and Insurance*. http://onlinelibrary.wiley.com/doi/10.1111/j.1539-6975.2012.01480.x.

6 Appendix

Descriptive Statistics of Rice Yield Data in sub-regions

Prefectures	*Expected yield (dt/ha)*	σ_Y *(dt/ha)*	*Sown area of rice (ha)*
Prefectures in Heilongjiang Province			
1 Harbin	79.96	9.94	473,217
2 Qiqihar	60.26	7.55	179,588
3 Jixi	72.32	9.12	143,286
4 Hegang	54.55	9.70	44,839
5 Shuangyashan	63.28	13.96	43,311
6 Daqing	61.81	16.24	58,896
7 Yichun	67.62	8.71	31,773
8 Jiamusi	67.83	8.67	238,205
9 Qitaihe	66.33	7.97	17,236
10 Mudanjiang	70.18	11.83	44,158
11 Heihe	56.13	9.87	9,490
12 Suihua	74.88	7.72	266,336
Prefectures in Jilin Province			
1 Changchun	87.36	5.03	175,389
2 Jilin City	82.52	9.23	141,613
3 Siping	90.61	10.76	59,238
4 Liaoyuan	78.41	9.76	17,344
5 Tonghua	90.50	9.81	76,024
6 Baishan	62.53	6.69	1,397
7 Songyuan	96.20	12.16	91,627
8 Baicheng	69.62	15.93	99,779
9 Yanbian	50.18	15.76	40,498
Prefectures in Liaoning Province			
1 Shenyang	79.31	9.24	126,800
2 Dalian	59.28	6.93	29,000
3 Anshan	73.81	11.04	38,100
4 Fushun	61.04	10.45	20,200
5 Benxi	59.82	6.90	9,400
6 Dandong	64.18	7.68	52,500
7 Jinzhou	74.79	8.03	29,300
8 Yingkou	94.52	11.57	44,300
9 Fuxin	62.00	10.47	5,800
10 Liaoyang	69.74	11.92	50,400
11Panjin	94.09	6.74	108,500
12 Tieling	75.01	13.02	64,700
13 Chaoyang	58.75	9.74	400
14 Hulvdao	60.77	10.50	9,100

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.
007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.
008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.
009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.
010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.
011 "The Real Consequences of Financial Stress" by Stefan Mittnik and Willi Semmler, February 2013.
012 "Are There Bubbles in the Sterling-dollar Exchange Rate? New Evidence from Sequential ADF Tests" by Timo Bettendorf and Wenjuan Chen, February 2013.
013 "A Transfer Mechanism for a Monetary Union" by Philipp Engler and Simon Voigts, March 2013.
014 "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, March 2013.
015 "Cyclical Variation in Labor Hours and Productivity Using the ATUS" by Michael C. Burda, Daniel S. Hamermesh and Jay Stewart, March 2013.
016 "Quantitative forward guidance and the predictability of monetary policy – A wavelet based jump detection approach –" by Lars Winkelmann, April 2013.
017 "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency" by Markus Bibinger, Nikolaus Hautsch, Peter Malec and Markus Reiss, April 2013.
018 "Fair re-valuation of wine as an investment" by Fabian Y.R.P. Bocart and Christian M. Hafner, April 2013.
019 "The European Debt Crisis: How did we get into this mess? How can we get out of it?" by Michael C. Burda, April 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

020 "Disaster Risk in a New Keynesian Model" by Maren Brede, April 2013.
021 "Econometrics of co-jumps in high-frequency data with noise" by Markus Bibinger and Lars Winkelmann, May 2013.
022 "Decomposing Risk in Dynamic Stochastic General Equilibrium" by Hong Lan and Alexander Meyer-Gohde, May 2013.
023 "Reference Dependent Preferences and the EPK Puzzle" by Maria Grith, Wolfgang Karl Härdle and Volker Krätschmer, May 2013.
024 "Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations" by Hong Lan and Alexander Meyer-Gohde, May 2013.
025 "The 'Celtic Crisis': Guarantees, transparency, and systemic liquidity risk" by Philipp König, Kartik Anand and Frank Heinemann, May 2013.
026 "State Price Densities implied from weather derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Huei-Wen Teng, May 2013.
027 "Bank Lending Relationships and the Use of Performance-Sensitive Debt" by Tim R. Adam and Daniel Streitz, May 2013.
028 "Analysis of Deviance in Generalized Partial Linear Models" by Wolfgang Karl Härdle and Li-Shan Huang, May 2013.
029 "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps" by Markus Bibinger and Mathias Vetter, May 2013.
030 "Can expert knowledge compensate for data scarcity in crop insurance pricing?" by Zhiwei Shen, Martin Odening and Ostap Okhrin, May 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".