European Actuarial Journal (2020) 10:361–375
https://doi.org/10.1007/s13385-020-00234-1

ORIGINAL RESEARCH PAPER

An average model approach to experience based premium rates discounts: an application to Spanish agricultural insurance

José L. Vilar-Zanón[1,3] · **Antonio Heras**[1] · **Estela de Frutos**[2]

Received: 7 June 2019 / Revised: 9 April 2020 / Accepted: 11 May 2020 / Published online: 30 May 2020

Abstract
We address some issues in agricultural insurance, describing drawbacks of the bonus-malus system (BMS) methodology used in Spain and many other EU countries. We develop an alternative experience based premium rate discount system taking into account the adverse years when high losses caused by extreme weather events happen. Our contribution consists of a two-step methodology. Firstly, we use tobit or Tweedie regressions to calculate yearly correction rates. Secondly, we calculate the mean of the correction rates. This average model acts as a buffer against adverse year losses. We compare three alternatives: our two resulting average models and the BMS operating in the Spanish line of business exemplified—table grapes.

Keywords Agricultural insurance · Experience premium rate discounts · Bonus-malus system · Tweedie · Tobit

1 Introduction

The purpose of this article is to discuss the actuarial methodology used in agricultural insurance when it comes to address the experience-rating step. It focuses on the Spanish case. Agricultural (also named crop) insurance is in many senses rather special compared to general insurance lines. From an organizational point of view, the Spanish market works exclusively through a coinsurance pool formed by 21

✉ José L. Vilar-Zanón
jlvilarz@ucm.es

1 Department of Financial and Actuarial Economics & Statistics, Universidad Complutense de Madrid, Madrid, Spain

2 Agroseguro S.A., Madrid, Spain

3 Departamento de Economía Financiera y Actuarial y Estadística, Universidad Complutense de Madrid, Campus de Somosaguas. Facultad de Ciencias Económicas y Empresariales, 28223 Pozuelo de Alarcón, Spain

insurance companies. It relies upon a joint action of public and private entities offering subsidies to farmers for their premiums. It offers coverage in all the agricultural productions against all the natural risks.

From a methodological point of view, the rating process has to overcome some hard problems. In [23] and [19] we can see that premiums must be calculated for coverage levels or deductibles once a farm yield distribution has been fitted. However, non-stationarity of crop yields due to technological change and weather variability calls for regular updates of the fitted distribution, see for instance [18, 22, 24, 26] and [7]. In addition, the yield spatial dependencies due to weather events jeopardizes the basic process of pooling independent risks [2]. In US crop insurance a loss cost ratemaking process is used to calculate the base rate from a mean of loss costs over time and farms [25].

The well-known general insurance problem on how to particularize the premiums by trying to approach them to individual risk behaviours is also very relevant in crop insurance. It reduces to solving the well-known credibility problem consisting of balancing the farm individual information and the collective one. By means of this experience-rating step, people try to measure the differences that could exist among farms or plots, trying to prevent adverse selection and to reward the plots with lower losses. In other words, it tries to assess to what extent the differences between losses are due to what we could call the handling (or management) of the plots, a variable that is not possible to observe directly. The literature on this important topic is rather scarce, and we have only found the reference [21], where the authors tie this topic to the US loss cost ratemaking process commented earlier. The authors define a new factor that represents for each plot an information increase, encapsulating in a single number a plot management compared to its neighbour's. Therefore, they manage to calculate premium rate corrections at the county (following the US terminology) and even at the individual plot levels. Moreover, this is done by changing the distribution-oriented methodology, since they calculate the correction factors applying a Tobit regression matching into the process the zeroes found in the loss data (see for example [1]). We may label this method as an American proposal for solving the experience-rating step.[1] In Spain and many other EU countries [3], the approach has been to adapt the well-known bonus-malus system (BMS). The intention is to introduce into agricultural insurance the benefits that this system has already shown in automobile insurance. Therefore, the goal would be to encourage insureds to follow good practices in their plots management by means of premium discounts and to reduce the incurred risk. Remember also the "bonus hunger", i.e. when an insured person does not declare a claim because the compensation might be lower than the subsequent Bonus-Malus (BM) surcharge, a phenomenon that reduces the administration costs as a collateral effect. The theory of Bonus-Malus systems is explained

[1] In the US there are several practical implementations of the idea of experience-based ratemaking. One reviewer let us know that US rating procedures 'fine-tune' base rates according to historical average yields. Since a higher average yield leads to a rate decrease, farms with higher average yields will have lower relative risk. Also, the Risk Management Agency (RMA) has implemented a biotech discount, lowering the premiums to those farmers that use modern technologies. In addition, discounts are also offered to producers taking coverage at higher levels of aggregation, see [14].

in [6, 15]. The design of such an experience rating system consists of the selection of the number of bonus-malus classes, the transition rules, and the scale of discounts/surcharges. Following this general setting, bonus-malus classes are defined and plots are classified on a yearly basis, depending on the past loss experience, by means of a set of transition rules. Each bonus-malus class has assigned a discount or a surcharge applied to all the plots belonging to it, and we expect this process to individualize the premiums making them closer to each true farm risk, i.e. reducing the remaining heterogeneity. The degree of this achievement is called the efficiency of the BMS. Thus, the higher the efficiency, the better the system. Many measures of efficiency have been proposed, see for instance [4, 11, 12, 17, 20]. The scale of the BM is the set of discounts and surcharges. A keystone in BMS design is the property of financial equilibrium of the BM scale, see [15]. As the premiums are not deterministic anymore (for they now depend on each policy's past loss experience, which is random), this property entails the yearly losses and premiums income to have equal expectations. This prevents the risk of premium insufficiency, i.e. suffering from a premium income chronically lower than the loss expense.

Nevertheless, it seems that the application of BMS theory to agricultural insurance has not been analysed in the literature, as we did not find any reference treating this interesting topic. We contribute to the field of experience rating in agricultural insurance by outlining some flaws observed in the Spanish case and proposing a new methodology free of them. Methodologically, our contribution traces back to the already mentioned US loss cost ratemaking process, because it is strongly related to [21].

The practical experience we bring to this article, proves that the translation of the BMS scheme to crop insurance might not lead to the desired benefits described above, particularly with reference to financial equilibrium. Indeed, the implementation of such a BMS has raised the following issues:

- The BM surcharges often add to the penalizations of the base tariff, resulting in premium increases not commercially affordable.
- The final premiums are largely insufficient, incapable of compensating the losses.
- The first and second points are worsened by the occurrence of unfavourable years caused by adverse weather events (like storms, freeze, floods, droughts etc.), resulting in unsustainable premium increases, along with very high loss increases. These extreme weather events may have a return period, so we can be sure to some extent of their occurrence within that period.

The consequence of these issues is that a BMS translation to agricultural insurance might increase the commercial premiums to the point of not being either acceptable from the commercial point of view or balanced in a technical sense.

The aim of this paper is to propose an experience-based premium discount system without those three drawbacks, while being able to reach the benefits described above. Our main contribution is a new methodology consisting of an average of yearly-based models defined within some adverse events return period. Our model is designed as a buffer acting against the adverse years losses mentioned earlier (issue number three), providing a rating system in financial equilibrium that solves the premium insufficiency (issue number two), with a more stable level of premiums over years (issue number one).

Suppose that we know the return period of the adverse weather events threatening the line of business. Obviously, this will consist of a certain number of years, say n. Then our methodology works in two steps. Firstly, we sequentially choose each of these n years considering it as the year to be explained, using the remaining $(n-1)$ as explanatory years, to calculate as many correction factors as years are in the return period. Note that at this point we have broken the time flow. In each of the n cases, we calculate a set of correction rates using the same new factor as in [21], by applying two alternative regression methodologies (Tweedie and Tobit) able to model the zeroes contained in the yearly loss data. This way we will compare the behaviour of both regression methods. We use Tobit because it was applied in [21]. We also apply Tweedie regressions because the Tweedie distribution has often been used in ratemaking applications. In the second step, we calculate the average correction rates produced by the n regressions. This is what we call the 'average model', designed to play as an adverse year losses buffer inside the return period.[2]

Therefore, we will compare three different experience-based premium rate discount models: one taken from the real world that follows the classical BMS approach, and two others summarized above implementing our average model methodology. We will compare them measuring their efficiencies (a well-known concept in BMS theory) by means of the calculation of their mean squared errors (mse). Finally, we will compare their associated scales of premiums to decide which model is the best one when applied to our data set. As for the discussion of our average model methodology and the use of the mse, we follow the field of Statistical Learning methods, see [9, 10].

In Sect. 2, we introduce the average model methodology together with the Tobit and Tweedie regressions. In Sect. 3, we use a data set from the Spanish "table grapes" line of business to compare the three models: the classical BMS already applied in the line of business, the "mean Tweedie" and the "mean Tobit" models. We end up with a concluding section.

2 Experience rating in agricultural insurance

Let us consider a line of business and a set of n consecutive years. We assume a base tariff as given, and we define P_{jt} and S_{jt} as the premium and loss, respectively, for plot j in year t. We also define $P_{j*} = (P_{jt})_{t=1,\dots,n}$ and $S_{j*} = (S_{jt})_{t=1,\dots,n}$ as the vectors of yearly premiums and losses, respectively, for plot j. Assuming that there are J plots in the geographical area, we also define $P_{*t} = (P_{jt})_{j=1,\dots,J}$ and $S_{*t} = (S_{jt})_{j=1,\dots,J}$ as the vectors of premiums and losses, respectively, for all the plots in a given year t.

[2] See [5] for an alternative Bayesian approach to weighting the experience in order to deal with extreme weather events.

First of all, we need a new *empirical factor F* increasing the information contained in the base tariff. Its purpose is measuring each plot management as compared to other plots located in a neighbouring geographical area (which we will identify with the region or county). We will follow here the idea of [21], where for each plot j, F_j is built as the ratio of its mean loss ratio during the last n years, over the mean of the last n years' loss ratios of the plots i located in the same geographical area:

$$F_j = \frac{\frac{\sum_{t=1}^{n} S_{jt}}{\sum_{t=1}^{n} P_{jt}}}{\frac{\sum_i \sum_{t=1}^{n} S_{it}}{\sum_i \sum_{t=1}^{n} P_{it}}}. \tag{1}$$

In (1) t stands for the years, and i for the plots inside the same geographical area. This empirical factor, defined as a ratio of loss ratios, aims to measure what we may call the 'plot handling'. This idea of plot handling attempts to encapsulate all the circumstances and conditions that could not be included in the base tariff because they are hardly observable. Note that the empirical factor can also be defined based on a more reduced subset of years. If a given year s is not used in the calculation of the empirical factor, we will denote it by $F_j^{(-s)}$. For instance, $F_j^{(-n)}$ will be the empirical factor built based on the first (n-1) years. We can also collect the empirical factors for all the different plots in a vector, denoted as $F^{(-s)} = (F_j^{(-s)})_{j=1,\dots,J}$.

For each plot j, we consider the vector $I_{j*} = (I_{jt})_{t=1,\dots,n}$ of the yearly insured sums or total liabilities (also called 'value of the plots output'). For each j, we also define the vector of the *loss costs* $(S/I)_{j*} = (\frac{S_{jt}}{I_{jt}})_{t=1,\dots,n}$ and the vector of the *premium ratios* $(P/I)_{j*} = (\frac{P_{jt}}{I_{jt}})_{t=1,\dots,n}$. Note that, dividing every component of the loss costs' vector by the corresponding component of the premium ratios' vector, we obtain the vector of the *loss ratios* $(S/P)_{j*} = (\frac{S_{jt}}{P_{jt}})_{t=1,\dots,n}$. Similarly, we can also define for a given year t the vectors $I_{*t} = (I_{jt})_{j=1,\dots,J}$, $(S/I)_{*t} = (\frac{S_{jt}}{I_{jt}})_{j=1,\dots,J}$, $(P/I)_{*t} = (\frac{P_{jt}}{I_{jt}})_{j=1,\dots,J}$ and $(S/P)_{*t} = (\frac{S_{jt}}{P_{jt}})_{j=1,\dots,J}$.

Next, we perform a regression analysis relating the loss ratio as dependent variable and the new empirical factor as the only independent variable. To avoid endogeneity problems, we consider the loss ratio for the last year n, $(S/P)_{*n} = (\frac{S_{jn}}{P_{jn}})_{j=1,\dots,J}$, and the empirical factor defined based on the first (n-1) years, $F^{(-n)} = (F_j^{(-n)})_{j=1,\dots,J}$. Denoting by α and β the parameters to be estimated, for every plot j we can write

$$\frac{S_{jn}}{P_{jn}} \sim \alpha + \beta F_j^{(-n)}. \tag{2}$$

Now (2) has the following difficulty: the insurance loss data have a very large amount of zeroes that would make that regression not meaningful. This happens because most of the policies do not have any claim during the year. To model this characteristic, [21] apply a Tobit (normal censored) regression. In this model, the zeroes are outcomes of a censored variable that could actually take negative values. Once we have checked that $F^{(-n)}$ is significant, for each plot j we could calculate a modified premium rate in year n:

$$P^R_{jn} = \left(\hat{\alpha} + \hat{\beta}F_j^{(-n)}\right)\frac{P_{jn}}{I_{jn}} \tag{3}$$

(where $\hat{\alpha}$ and $\hat{\beta}$ are the estimators of the regression parameters). Once we get the premium rates, it is easy to calculate the modified premiums for every plot j in year n (a vector P^{EXP}, the experience premiums)

$$P^{EXP}_{jn} = I_{jn}P^R_{jn}. \tag{4}$$

However, taking into account the three problems described earlier, this approach has an important flaw. It is well known that there is a great variability of the losses in crop insurance, due to the yearly differences in weather conditions. If the premium rates (3) were calculated on the basis of a factor $F^{(-n)}$ produced in favourable years (with low losses), the resulting premiums (4) could not be sufficient to compensate the losses of future unfavourable years, when extreme weather events might significantly increase the losses.

At first sight, we could remedy the second flaw by considering some mean loss of past years as the dependent variable in (2). However, this regression would be spurious as the dependent and independent variables would be based on the same information (the past losses). Therefore, the dependent variable in (2) can be neither the yearly loss nor a mean of past yearly losses.

Our proposal for solving this problem consists of defining n regressions similar to (2), following the explanation developed in the Introduction. Each regression corresponds to an explained year t and the $(n-1)$ explanatory years $\{1,\dots,t-1,t+1,\dots,n\}$, all of them inside a return period n associated to an extreme weather event threatening the line of business. In each case, the yearly loss ratio is the dependent variable, and we build up the new factor $F^{(-t)}$ using the remaining $(n-1)$ years belonging to the extreme event return period: in other terms, the regression corresponding to year t would be $\frac{S_{jt}}{P_{jt}} \sim \alpha_t + \beta_t F_j^{(-t)}\,(j = 1,\dots,J)$.

Note that in this formulation we abandon the natural time flow interpretation, breaking it. For each year $t \in \{1,\dots,n\}$, the corresponding Tobit regression will produce new premium rates $P^R_{jt,Tobit}$ for every plot j, defined as:

$$P^R_{jt,Tobit} = \left(\hat{\alpha}_t + \hat{\beta}_t F_j^{(-t)}\right)\frac{P_{jt}}{I_{jt}}. \tag{5}$$

For each plot j, the final output of the analysis is the proposed premium rate for the next period $n+1$, calculated as an average of the modified premium rates for the years $1,\dots,n$:

$$\overline{P}^R_{j,Tobit} = \frac{1}{n}\sum_{t=1}^{n} P^R_{jt,Tobit}. \tag{6}$$

These new premium rates have been obtained under the Tobit model assumptions, following the methodology of [21]. As for the occurrence of many zeroes, we can give a try to an alternative methodology, fitting a Tweedie model instead of a Tobit model at the stage of (2).

The Tweedie distribution can be defined as a compound distribution with Poisson primary and Gamma secondary distributions. It has been widely applied in insurance claims modelling because it is a mixed distribution that assigns a positive probability to zero, representing no claims, and also has a positive continuous part, representing the density of the positive claim amounts. The fact of putting some probability at the zero outcome is quite important for our application, because it allows interpreting the sample zeroes in their true sense, as null realizations of the random variable.[3]

If we suppose a Tweedie model for the loss ratios instead of the Tobit model, since the function linking the independent and dependent variables is logarithmic, the regression corresponding to year t will be

$$\log\left(\frac{S_{jt}}{P_{jt}}\right) \sim \alpha_t + \beta_t F_j^{(-t)} \; (j = 1, \ldots, J). \tag{7}$$

Thus, we come to a formula similar to (5), assigning a modified premium rate to each plot j

$$P^R_{jt,Tweedie} = e^{\hat{\alpha}_t + \hat{\beta}_t F_j^{(-t)}} \frac{P_{jt}}{I_{jt}}. \tag{8}$$

The exponential term is the individual factor by which the premium rate of plot j should be corrected according to the information given by the new factor $F_j^{(-t)}$. In the same way as in (6), the premium rates to be applied for the next period will be $\overline{P}^R_{j,Tweedie}$, the average of the n yearly premium rates:

$$\overline{P}^R_{j,Tweedie} = \frac{1}{n} \sum_{t=1}^{n} P^R_{jt,Tweedie}. \tag{9}$$

As we have noted in the Introduction, the main objective of any experience rating system is to increase the efficiency. The efficiency is generally thought of as the

[3] The Tweedie distribution depends on a parameter p, known as the ‘index parameter’, and the different values of this parameter allow to obtain as particular cases some important continuous, discrete and mixed distributions. For example, the Normal, Poisson and Gamma are particular cases of a Tweedie distribution with $p=0$, 1 and 2, respectively. When $1<p<2$, the Tweedie becomes a Poisson-Gamma compound distribution. This is precisely the interesting case for us, because it is a mixed distribution with a discrete mass of probability assigned to zero (interpreted as the probability of no claims) and a positive continuous part (interpreted as the density of the claim amounts). For information about the Tweedie distribution see Chapter 12 of [8], and for its ratemaking applications, see [13]. Regarding the Tobit model, it can also be considered as a mixture of a claim/no claim binary distribution and a truncated normal distribution giving the density of the positive claim amounts, see [16]. Although this decomposition makes sense in insurance problems, it is rarely found in the actuarial literature.

degree to which the premiums approximate the true risk of the policies. In the classical BMS, this efficiency is usually computed by means of the evaluation of some measure of the discrepancies between the premiums and the losses (mean square error, mean absolute error, see [12, 15, 17, 20] and [11]). All those efficiency measures are based on the risk modelling through a random variable Λ, the mean number of claims per time unit, whose density function is called the *structure function*. However, this does not make sense in the problem we study here, because in our data the risks are described by their losses, not by their mean number of claims per time unit.

Therefore, for each plot j we propose to consider its mean loss cost during the n years of the return period

$$\overline{LC}_j = \frac{1}{n}\sum_{t=1}^{n}\frac{S_{jt}}{I_{jt}}. \tag{10}$$

We will define the vector of the mean loss costs for all the plots as $\overline{LC} = (\overline{LC}_j)_{j=1,\dots,J}$.

Then, we will compare these values with the mean premium rates obtained in (6) or (9), by means of the *mean squared error* (*mse*), defined as the average of the squares of the errors. We have chosen the *mse* as the efficiency measure of the experience rating discount system because it can be considered as an adaptation to our problem of the measures of efficiency found in the general theory of BMS mentioned earlier. Moreover, the *mse* is commonly used in Statistical Learning to evaluate the quality of the predictive models, see [9] and [10]. When comparing several alternative experience-rating systems with mean premium rates given by vectors $\overline{P}^R = (\overline{P}_j^R)_{j=1,\dots,J}$, the best one will obviously be the one minimizing (11):

$$mse\left(\overline{LC}, \overline{P}^R\right) = \sum_{j=1}^{J}(\overline{LC}_j - \overline{P}_j^R)^2. \tag{11}$$

We will also measure the variability of the modified premiums by means of their standard deviation. Thus, the larger the standard deviation, the farther will the system be from the situation where all the plots pay the same premium. We want to underline the idea that a system supplying premiums with a lower standard deviation is commercially preferred (assuming equality in mean), because it implies a shorter variability over the policies, thus shorter discounts and penalties to impose to them.

3 Practical application to the Spanish table grapes line of business

We now exemplify our methodology using a data set describing the Spanish table grapes line of business between the years 2012 and 2016.

During those years, the insurance company has been using a classical BMS, built up through a heuristic process, summarized next. Remember that the use of BMS is quite common in many EU countries, as described in [3].

The company BMS takes into account the following variables for each insured:

- Insurance underwriting in the last season (year).
- Claims statement during the last season.
- Number of years with insurance underwriting in the last 10 years.
- Number of years with loss during the last 10 years or seasons.
- Ratio of accumulated indemnities over surcharged risk premiums (excluding the "Consorcio…[4]" Reinsurance fee) for the same 10 years period.

These five variables determine the bonus-malus class for each policy. There are thirteen classes in the BMS: seven bonus classes, five malus classes and one neutral. Policies are classified yearly, and the scale of penalties and bonuses ranges from + 25 to − 40%.[5]

The data set consists of 6,483 policies, each one corresponding to one plot. Each policy is described by fifteen fields, but we are mainly interested in the following: line of business (always "table grapes"), year, ID number, Spanish province where the plot is located, region of the province, premium calculated according to the base tariff, modified premium calculated according to the BMS set up by the company, total insured output value of the plot which stands as the insured sum, and yearly total loss.

We have selected those policies appearing with the same insured plot during the 5 years. This way we obtain 528 policies per year with a one to one relationship between ID numbers and the plots insured during the 5 years. Therefore, we have used 2640 registers to exemplify our methodology.

Table 1 contains the statistical summaries of the loss cost variable during the five years. We can check in Table 2 that the amounts of null loss costs are quite high, confirming the already mentioned technical difficulty for running regressions.

From the statistical summaries given in Tables 1 and 2, we conclude that years 2013 and 2015 were clearly unfavourable. We highlight in Table 3 the mean loss costs for the 5 years: we check that the 2013 and 2015 loss costs are significantly greater than those of the other years.

Our data illustrate the issues outlined in the Introduction. If we use data corresponding to a favourable year to feed the experience rating system, the resulting corrected premiums will be insufficient if an adverse year comes. If we fed it with data coming from a bad year, it would rise the premiums up to an excessively high level, adding in many cases surcharges to a base premiums already high.

[4] The Spanish "Consorcio de Compensación de Seguros" (CCS), is a public business entity attached to the Spanish Ministry of Economy. The CCS relies on its own capital, without any dependence of public funds. From the agricultural insurance perspective, it develops three main lines of activity, playing a key role as a system compulsory reinsurer, monitoring loss adjustments and participating in the coinsurance scheme. The involvement of the CCS in the Spanish system of agricultural insurance is crucial since it decreases the impact of losses making possible risk underwritings.

[5] For more information, see *Sect. 14. Bonificaciones y Recargos* in the document *Seguro de Explotaciones de Uva de Mesa: condiciones especiales* (in Spanish) https://agroseguro.es/fileadmin/propietario/Productos/AGRICOLAS/321%20UVA%20DE%20MESA/PLAN_2018/CES-321-18-1.0.pdf

Table 1 Loss costs statistical summaries

	Min	1st Qu	Median	Mean	3rd Qu	Max
2012	0	0	0	0.05877	0.03197	0.9743
2013	0	0	0	0.1367	0.2124	0.8943
2014	0	0	0	0.05138	0	0.9
2015	0	0	0	0.1742	0.3135	0.99
2016	0	0	0	0.0479	0.006723	0.66

Table 2 Yearly proportions of null loss costs

	2012	2013	2014	2015	2016
Null loss costs	72.34%	57.57%	82%	54.92%	73.86%

Table 3 Mean loss costs for the 5 years period

	2012	2013	2014	2015	2016
Mean Loss Cost	0.05876	0.1367	0.05138	0.17422	0.04789

Therefore, we will apply our methodology, calculating an average model inside a return period relative to the extreme weather events. In the table grapes line of business we will suppose a five years return period, which is in fact our data recollection time length.[6] For each insured ID number (or equivalently, for each plot j between 1 and $J = 528$), and for each year $t \in \{1, 2, 3, 4, 5\}$, we assume as known the premium rates charged by the company after the implementation of the BMS, which we will denote by $P^R_{jt,BM}$.

We then calculate the mean premium rates for each policy j under the Tobit model, $\overset{-R}{P}_{j,Tobit}$, and under the Tweedie model, $\overset{-R}{P}_{j,Tweedie}$, and define the vectors of mean premium rates under both models[7]:

$$\overline{P}^R_{Tobit} = (\overline{P}^R_{j,Tobit})_{j=1,\dots,J};\ \overline{P}^R_{Tweedie} = (\overline{P}^R_{j,Tweedie})_{j=1,\dots,J}. \tag{12}$$

We then compare these results with the premiums rates calculated through the BMS of the company.[8] In Table 4, we show the statistical summary of these premium rates

[6] The selection of $n = 5$ years for the return period has been based on the fact that we can find both favourable and unfavourable years within this period, as we can check in Tables 1, 2 and 3. Rejesus et al. also choose a period of 5 years for the analysis, based on different considerations: "*The choice of five years is arbitrary but reflects a balance between a longer period (that would have greater statistical power) and choosing a shorter period (that would make it possible for more producers to qualify but would have lower statistical power)*" (see [21], pg. 414).

[7] It is important to highlight that the empirical factors $F^{(-t)}, t \in \{1, \dots, 5\}$, were always significant in all the Tobit and Tweedie regressions. For the numerical calculations, we used the R packages 'AER' (https://cran.r-project.org/web/packages/AER/AER.pdf) and 'tweedie' (https://cran.r-project.org/web/packages/tweedie/tweedie.pdf).

[8] The outputs of the statistical procedures are available to the readers upon request.

Table 4 Statistical summaries of the premium rates calculated with the regression models (Tobit and Tweedie), the mean of the company BMS during the 5 years, and the company BMS in 2016

	Min	1st Qu	Median	Mean	3rd Qu	Max	St. Dev
MeanTobit	0.01399	0.07348	0.09102	0.09798	0.1139	0.2727	0.04006
MeanTweedie	0.01474	0.07639	0.09195	0.09753	0.1118	0.2847	0.03657
Mean BMS	0.0126	0.05618	0.06963	0.06809	0.07752	0.1391	0.01882
BMS 2016	0.008035	0.05591	0.07440	0.07337	0.08552	0.1709	0.02548

together with their standard deviations. We also show the premium rates corresponding to the BMS issued by the company in year 2016 and finally the mean BMS premium rates during the 5 years. We see that for the Tobit model, the mean and standard deviation are higher than for Tweedie. This means that the Tweedie model produces a set of premiums better suited to commercial purposes, because their discounts and surcharges are narrower. All the statistics corresponding to these two cases are higher than those corresponding to the company BMS in 2016 (two last rows of Table 4).

We now measure the efficiency of the mean Tobit and mean Tweedie models, calculating the *mse* with respect to the mean loss cost[9]:

$$
\begin{aligned}
mse(\overline{LC}, \overline{P}^R_{Tobit}) &= \sum\nolimits_{(j=1)}^{J} (\overline{LC}_j - \overline{P}^R_{(j,Tobit)})^2 = 0.00963, \\
mse(\overline{LC}, \overline{P}^R_{Tweedie}) &= \sum\nolimits_{(j=1)}^{J} (\overline{LC}_j - \overline{P}^R_{(j,Tweedie)})^2 = 0.01044.
\end{aligned}
\tag{13}
$$

In (13) we see that Tobit's *mse* is slightly lower than Tweedie's. However, we may still prefer the Tweedie model due to its benefits (natural interpretation of the zeroes, lower standard deviation).

We now compare (13) with the efficiency of the premium rates issued by the company. If, for every policy j, we define the mean premium rate by

$$
\overline{P}^R_{j,BM} = \frac{1}{5} \sum_{t=1}^{5} P^R_{jt,BM}
$$

, and the vector of the mean premium rates by $\overline{P}^R_{BM} = (\overline{P}^R_{j,BM})_{j=1,\dots,J}$, the efficiency can be measured by

$$
mse\left(\overline{LC}, \overline{P}^R_{BM}\right) = \sum_{j=1}^{J} (\overline{LC}_j - \overline{P}^R_{j,BM})^2 = 0.01641 \tag{14}
$$

which is greater than the values obtained in (13).

We conclude that the premium rates issued by our two average models (Tweedie and Tobit) are more efficient than those issued by the company BMS. We highlight that this efficiency improvement is achieved thanks to the average model methodology, which we are presenting as our main finding.

[9] The values of the mse have been calculated using the Leave One Out Cross Validation (LOOCV) method.

Table 5 Comparison between premiums and losses. Comparing with "5 years mean losses", we check that the company BMS does not reach the financial equilibrium, while Tweedie and Tobit approximately achieve it (see the column "mean")

	Min	1st Quar	Median	Mean	3rd Quar	Max
Mean Tobit premiums	40	1258	2911	9828	6398	699,000
Mean Tweedie premiums	42.1	1287	2887	9636	6462	639,400
Losses 2016	0	0	0	5832	343.5	1,008,000
5 years mean losses	0	0	1907	9818	7102	445,700
Mean premiums BMS	80.36	1107	2387	5328	4483	142,400
Premiums BMS 2016	7.2	916.9	2245	6561	5074	327,600

Table 6 Premiums in financial equilibrium

	Min	1st Qu	Median	Mean	3rd Qu	Max	St. Dev
Mean Tobit premiums	40	1256	2908	9818	6391	698,261	40,881
Mean Tweedie premiums	43	1311	2942	9818	6584	651,444	39,549

We also have calculated the premiums corresponding to each model using (4), and reported their statistical summaries in Table 5. We find the current BMS premium insufficiency (compare the means column "5 years mean losses" with "Premiums BMS 2016" and also with "Mean premiums BMS"). We see also that the mean Tobit model gives a mean premium (9828) greater than the mean losses of the 5 years (9818), while the mean Tweedie model has a lower mean premium (9636). Thus, the two models reach approximately the financial equilibrium, though not exactly. We can redress this slight discrepancy multiplying the two sets of premium by suitable correction factors to get premiums in financial equilibrium.[10] We report the statistical summaries of the corrected premiums in Table 6. There we find two sets of premiums in financial equilibrium, but the Tweedie case has a lower standard deviation, and thus gives a narrower set of premiums. This may be a good reason to prefer the Tweedie model, additionally to its natural way of modelling the zeroes. In any case, we have found that the average model methodology produces a more suitable set of premiums than the BMS methodology currently in use.

4 Conclusions

In this article, we have reported some important flaws of the classical BMS methodology that is nowadays in use in Spain and in other EU countries: the high variability and insufficiency of the premiums, mostly due to the great losses associated to adverse years with extreme weather events. Motivated by this situation, we have

[10] In order to reach the financial equilibrium, we should multiply the Mean Tobit Premiums by $\frac{9{,}818}{9{,}828}$, and the Mean Tweedie Premiums by $\frac{9{,}818}{9{,}636}$.

developed a new methodology for calculating experience-based premium rate discounts and penalties, which can act as a buffer against the adverse years in agricultural insurance. Moreover, our approach might also be a solution to the chronic premium insufficiency that seems to arise when applying the classical BMS approach.

Our methodology does not rely on the design of a classical BMS. Instead, we incorporate the new experience-based information by means of regression models, closely following the methodology of [21]. Despite the importance of the BMS in agricultural insurance, [21] is, to our knowledge, the only relevant reference in the literature. The authors define a new factor representing additional information for each plot, encapsulating in a single number the quality of the plot management. Then, they calculate the correction factors by applying a Tobit regression. The use of the Tobit regression, instead of less sophisticated regression techniques, is due to the high number of zeroes (null loss costs) found in the loss data.

In our work, we have explored the possibility of applying an alternative technique, the Tweedie regression, instead of the Tobit regression. This seems to us a natural way to deal with the problem of the occurrence of many zeroes in the dependent variable, because the Tweedie distribution is a mixed distribution assigning a discrete mass of probability to zero. Besides, the Tweedie distribution is often used by actuaries in insurance ratemaking problems. We have concluded that the results obtained with Tweedie regressions are very similar to those obtained with Tobit regressions, and even better in some respects (such as less variability of the premium rates).

But our main contribution is the development of an average model methodology, defined over a period of time (the return period) in which we can find not only normal or favourable years, but also unfavourable or adverse years, with extreme weather events. During such adverse years, high losses will occur due to extreme events like storms, freezes, floods, droughts etc. (depending on the line of business). We have found that, no matter the regression model applied (Tobit or Tweedie), the final average model will act as a buffer against the adverse years' losses, supplying premiums that approach much better the mean loss cost over the return period.

Even if the obtained premiums only approximate the financial equilibrium, we can always face this problem by multiplying those premium sets by the adequate correcting factors, finally getting a new sets of financially balanced premiums. This is the final step in order to overcome the three drawbacks mentioned in the Introduction.

We are aware that our methodology is sensitive to the estimation of the return period associated to the natural events affecting the losses in the line of business. Thus, it should be carefully defined, taking into account both the data and the expert's advice. In our application, after hearing the expert advice, we made it equal to five years.

Another key factor is the definition of the empirical factor (1), which should increase the information already contained in the base tariff. We have adopted the factor suggested in [21] that aims to encapsulate the quality of each insured plot handling or management. This is another step where the experts' opinion should be carefully listened. Some new questions arise here, such as: how many periods should be considered, and how to choose the region where we compare the insured's loss cost performance. Another important problem is related to the degree of granularity of the application of the methodology: should it be applied globally (no matter the province or region) or locally (inside each province or region)? In our example, we have applied the new factor globally, given the characteristics of the exemplified line of business (all the provinces and regions are geographically close each other), and following the experts' advice.

Regarding the question of choosing the best approach among the Tobit and Tweedie models, although both give very similar results, in our case we may tip the balance in favour of Tweedie. From a theoretical point of view, Tweedie has a natural interpretation of the great number of zeroes in the dataset. From the applied point of view, in our example it has produced a set of premiums that are better suited for commercial purposes, even if its efficiency, measured by the *mse*, is slightly worse than the obtained in the Tobit model. In any case, we could always apply the two regression methods in order to finally choosing the set of premiums best suited to our purposes (minimum *mse*, lowest standard deviation, financial equilibrium).

Acknowledgements This work was supported by Agroseguro under UCM contract number 437-2016. The authors are very thankful to Pablo Ezquerra and Felix Novoa from AGROSEGURO, for their helpful comments and suggestions.

References

1. Anemiya T (1984) Tobit model: a survey. J Econom 24:3–61
2. Annan F, Tack J, Harri A, Coble K (2013) Spatial pattern of yield distributions: implications for crop insurance. Am J Agr Econ 96(1):253–268
3. Bielza M, Conte C, Gallego F, Stroblmair J, Catenaro R, Dittman C (2009) Risk management and agricultural insurance schemes in Europe.https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/agricultural-insurance-schemes
4. Borgan O, Hoem J, Norberg R (1981) A non-asymptotic criterion for the evaluation of automobile bonus systems. Scand Actuar J 165–178
5. Borman JI, Goodwin BK, Coble KH, Knight TO, Rejesus R (2013) Accounting for short samples and heterogeneous experience in rating crop insurance. Agric Financ Rev 73(1):88–101
6. Denuit M, Maréchal X, Pitrebois S, Walhin J (2007) Actuarial modelling of claim counts. Wiley, Hooken
7. Duarte G, Braga A, Miquelluti L, Ozaki V (2018) Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance. J Appl Stat 45(11):1920–1937
8. Dunn PK, Smyth GK (2018) Generalized linear models with examples in R. Springer, New York
9. Gareth J, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, New York
10. Hasti T, Tibshirani R, Friedman J (2009) The elements os statistical learning data mining, inference and prediction. Springer, New York
11. Heras A, Gil J, García P, Vilar-Zanón J (2004) An application of linear goal programming to bonus malus system design. Astin Bull 34:435–456
12. Heras A, Vilar-Zanón J, Gil J (2002) Assymptotic fairness of bonus malus systems and optimal scales of premiums. Geneva Pap Risk Insur-Theory 27:61–82
13. Jorgensen B, Paes de Souza M (1994) Fitting Tweedie's compound model to insurance claims data. Scand Actuar J 1:69–93
14. Knight T, Coble K, Goodwin B, Rejesus R, Seo S (2010) Developing variable unit-structure premium rate differentials in crop insurance. Am J Agr Econ 92(1):141–151

15. Lemaire J (1995) Bonus malus systems in automobile insurance. Kluwer academic publishers, Dordrecht
16. McDonald J, Moffitt R (1980) The uses of Tobit analysis. Rev Econ Stat 62(2):318–321
17. Norberg R (1976) A credibility theory for automobile bonus system. Scand Actuar J 92–107
18. Ozaki V, Silva R (2009) Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. J Appl Stat 36(4):443–452
19. Ozaki V, Goodwin B, Shirota R (2008) Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts. Appl Econ 40(9):1151–1164
20. Pesonen M (1963) A numerical method of finding a suitable bonus-malus scale. Astin Bull 2:102–118
21. Rejesus R, Coble K, Knight T, Jin Y (2006) Developing experience-based premium rate discounts in crop insurance. Am J Agr Econ 88(2):409–419
22. Shen Z, Odening M, Okhrin O (2018) Adaptive local parametric estimation of crop yields: implications for crop insurance rate making. Eur Rev Agric Econ 45(2):173–203
23. Skees J, Reed M (1986) Rate making for farm-level crop insurance: implications for adverse selection. Am J Agr Econ 68(3):653–659
24. Tolhurst T, Ker A (2014) On technological change in crop yields. Am J Agr Econ 97(1):137–158
25. Woodard J, Sherrick J, Schnitkey G (2011) Actuarial impacts of loss cost ratio ratemaking in U.S. crop insurance programs. J Agric Resour Econ 36(1):211–228
26. Zheng Q, Wang H, Shi Q (2014) Estimating bivariate yield distributions and crop insurance premiums using nonparametric methods 46(18):2108–2118