Spanish Journal of Agricultural Research
15 (4), e0119, 12 pages (2017)
eISSN: 2171-9292
https://doi.org/10.5424/sjar/2017154-10716
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, O.A, M.P. (INIA)

RESEARCH ARTICLE

OPEN ACCESS

Crop insurance demand in wheat production: focusing on yield gaps and asymmetric information

Alba Castañeda-Vera[1,2], Antonio Saa-Requejo[1], Inés Mínguez[1,2] and Alberto Garrido[1]

[1]*Research Centre for the Management of Agricultural and Environmental Risks (CEIGRAM), Technical University of Madrid (UPM), Madrid, Spain*
[2]*Agricultural Systems Group, Dpt. Producción Agraria, E.T.S. Ingenieros Agrónomos, Technical University of Madrid (UPM), Madrid, Spain*

Abstract

Analysis of yield gaps were conducted in the context of crop insurance and used to build an indicator of asymmetric information. The possible influence of asymmetric information in the decision of Spanish wheat producers to contract insurance was additionally evaluated. The analysis includes simulated yield using a validated crop model, CERES-Wheat previously selected among others, whose suitability to estimate actual risk when no historical data are available was assessed. Results suggest that the accuracy in setting the insured yield is decisive in farmers' willingness to contract crop insurance under the wider coverage. Historical insurance data, when available, provide a more robust technical basis to evaluate and calibrate insurance parameters than simulated data, using crop models. Nevertheless, the use of crop models might be useful in designing new insurance packages when no historical data is available or to evaluate scenarios of expected changes. In that case, it is suggested that yield gaps be estimated and considered when using simulated attainable yields.

Additional keywords: risk management; rainfed wheat; crop insurance penetration rate; crop models; Spain.

Abbreviations used: AIC (Akaike information criterion); *AsymB* (asymmetric information indicator for the insurance option Basic); *AsymE* (asymmetric information indicator for the insurance option Extended); *AsymEc* (asymmetric information indicator for the insurance option Extended including the complementary insurance); CyL (Castilla y León); RMSE (root mean square error); SiAR (Sistema de Información Agroclimática para el Regadío); *Ya* (actual yield); *Yexp* (expected rainfed yield); *YinsB* (average insured yield in option Basic); *YinsE* (average insured yield in option Extended before sowing); *YinsEc* (average insured yield in option Extended including the complementary insurance); *YinsZ* (zonal maximum insurable yield); *Ythresh* (average insured yield in insurance option Extended including the complementary insurance after subtracting a 30% deductible).

Authors' contributions: Conception or design; acquisition, analysis, or interpretation of data; drafting of the manuscript: ACV. Statistical analysis: AS. Coordinated the research project: MIM. Critical revision of the manuscript for important intellectual content, supervising the work: AG.

Citation: Castañeda-Vera, A.; Saa-Requejo, A.; Mínguez, I.; Garrido, A. (2017). Crop insurance demand in wheat production: focusing on yield gaps and asymmetric information. Spanish Journal of Agricultural Research, Volume 15, Issue 4, e0119. https://doi.org/10.5424/sjar/2017154-10716

Supplementary material (Tables S1 to S5) accompanies the paper on SJAR's website

Received: 14 Nov 2016. **Accepted:** 22 Nov 2017.

Funding: Spanish Ministerio de Economía y Competitividad (project MULCLIVAR CGL2012-38923-C02-02), ACV received a PhD grant from UPM (Madrid).

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to Alba Castañeda-Vera: albacastanedavera@gmail.com

Introduction

Developing efficient agricultural insurance requires overcoming a number of market imperfections and combating asymmetric information. Very often, these difficulties justify governments' intervention in promoting insurance demand through the implementation of public incentives, such as premium subsidies, and a significant degree of market intervention, guidance and overseeing (Mahul & Stutley, 2010).

Asymmetric information occurs when insured farmers have more information than the insurer about their actual risk and behaviour, and results in two behavioural responses: moral hazard and adverse selection. Moral hazard occurs when farmers' expected indemnity under insurance is larger than under optimal uninsured conditions (Coble *et al.*, 1997). This means that farmers might modify their behaviour after contracting insurance in order to increase the probability of being indemnified (Goodwin, 1994) or reduce the effort to escape risk once they are covered

for it. Adverse selection happens when high-risk farmers contract insurance in greater numbers than low-risk farmers. The insurer, being unable to differentiate them, is forced to average out the risks of both types in calculating the premium. In these situations, the insurer does not have adequate information to calculate the unbiased probability and severity of claims. The result can be unbalanced loss ratios, thereby affecting the actuarial robustness and sustainability of entire insurance systems.

Yield gaps are defined in agronomy as the differences between attainable yields based on climate and soil conditions and actual farmers' yields; they are usually defined with reference to some specified spatial and temporal scale (Lobell *et al.*, 2009). Quantifying such gaps is useful to identify suboptimal crop management and the opportunity for the agronomic improvement of productivity (*e.g.* Lobell *et al.*, 2009; Affholder *et al.*, 2012; van Wart *et al.*, 2013). Yield gap has been proven to be a valuable concept for assessing and understanding the ecological possibilities to meet food demand for an increasing population (van Ittersum *et al.*, 2013). Nevertheless, the use of this concept is not limited to the biophysical aspects of agronomy, but may also be used to support research on socioeconomic aspects of agricultural production. Examples are seen in the limits to technology adoption due to either technical constraints or for economic issues arising from market conditions (Godfray *et al.*, 2010). In the context of crop insurance, a conceptual variation of yield gap might be also defined as the difference between insured and actual yields. Such a gap indicates either a deficiency in the insurer's knowledge or an information advantage of the farmer. Either situation is a case of asymmetric information. The lack of data on actual yields is very often the main cause of this asymmetry (Mahul & Stutley, 2010). Crop models can be used generate data on expected yields for risk analyses and failure studies in the crop insurance market when no historical data are available.

This paper presents analyses of yield gaps in the context of crop insurance, and the possible influence of asymmetric information in the decision of a farmer to contract insurance. A novel method to measure asymmetric information was designed based on yield gaps and applied to wheat insurance in the region of Castilla y León (northern central Spain). The analysis includes simulated yield data using the CERES-Wheat crop model in order to assess the suitability of such models to estimate actual yield risk when no historical data are available. The aims of the study were to (i) quantify the gap between actual wheat yields and the yield data managed by the insurer, (ii) obtain indicators providing evidence of potential asymmetric information in wheat insurance, (iii) explore the impact of asymmetric information on the decision of farmers to contract insurance, and (iv) evaluate the use of a crop model within the context of crop insurance.

Material and methods

Winter cereal insurance

In Spain, multi-peril insurance for cereal farmers has offered since 1982. The database for this study pertains to crop seasons 2010-11 and 2011-12 and includes four insurance modules (P, 1, 2 and S) that were available to cereal farmers. Module S has not been offered since 2013. The modules differ based on the coverage, on whether the crop is grown rainfed or irrigated, on the way indemnities are calculated when a loss occurs (whether indemnities are paid for each individual field or for the whole farm), and on the maximum insurable yield guaranteed (Table S1 [suppl.]). Two types of risks are defined. Type A risks include hail, fire, crop damage caused by wild fauna, flood and excessive rainfall. Type B risks include adversities impeding crop emergence (no-emergence risks) or limiting crop growth (including drought).

For irrigated crops, losses associated to both risk types A and B are indemnified per plot if the final yield is lower than the insured yield, the latter being freely set by the farmer. Under rainfed conditions, however, losses associated to type B risks can be indemnified per plot (Module 2) or per farm (Modules 1 and S). In the latter case the farmer is indemnified only if the average yield of all farm plots is lower than the insured yield, and farmers are obliged to insure all farm plots. Module S limits insured yield to a maximum officially assigned by the Spanish Agency of Agricultural Insurance (ENESA, initials in Spanish). Modules 1 and 2 limit insured yield to a maximum insurable yield calculated for individual farmers based on their insurance history records. Lastly, farmers might select a deductible, but most of them usually opt for the lowest of 30%.

To simplify the analysis, the four modules (P, 1, 2 and S) were combined into two options based on the risks covered. The first option, called *Basic*, comprises module P that guarantees farmers' production with coverage for type A risks. The second option, called *Extended*, includes modules 1, 2 and S, and provides coverage for risks of both type A and type B. Farmers choosing the *Extended* option may increase insured yield at midseason with a complementary insurance payment if yield expectation exceeds that insured before sowing.

Crop insurance premiums are proportional to insured yields and vary depending on the contracted module and deductible. They are subsidized by the national and regional governments. The subsidy varies depending on the option contracted, the farmer's characteristics (age, gender and others), and whether the contract is a renewal or is contracted through a farmer association. The lowest base subsidy is for module P (option *Basic*) increasing for modules S, 2 and 1 (option *Extended*) as 1, 8, 17 and 22% of the premium cost. Additional subsidies based on the farmer's characteristics (young farmers, gender and others), and whether it is contracted collectively through a farmer association are also different for different modules, in such a way that Module P (option *Basic*) has the lower maximum assignable subsidy followed by Module S, Module 2, and lastly Module 1 (option *Extended*) with 16, 33, 52 and 57 % of the premium cost (Table S1 [suppl.]).

Study site

Spain is the fifth largest producer of cereals in Europe, with a harvested production that ranged between 14 and 25.5 million Mg (6.2 Mha of crop) between 2000 and 2013 (Eurostat, 2013). About 47% of the cultivated land dedicated to arable crops (including cereal) is located in Castilla y León region, with about 2.9 Mha (ESYRCE, 2013).

Castilla y León (CyL) is located in northern central Spain. Climate is continental Mediterranean with warm and dry summers and cold and wet winters. It is a high plateau (830 masl average) around the Duero river basin surrounded by mountain ranges. There is a clear gradient of temperature and rainfall from North-East (cooler and wetter) to South-West (warmer and drier).

The most commonly cultivated cereals are barley (*Hordeum vulgare*) and bread wheat (*Triticum aestivum*), with 54% and 30% of total regional production, respectively. Wheat is mainly grown without irrigation (JCyL, 2014) so that yields respond to the high rainfall variability (240 to 700 mm). Average rainfed yields in the study region during 2000–2013 varied between 2.0 and 4.5 Mg/ha (JCyL, 2014). Winter wheat is sown during October–December and spring wheat during February–March. Wheat is harvested between late June and mid-August. The most sensitive phenological stages to meteorological adversities are flowering (April-May, late frosts, drought and heat stress) and grain filling (June-July, heat and drought).

Actual, expected, insured and water limited yields

Actual and expected yields

Actual yields (*Ya*) were obtained from the ESYRCE database (*Encuesta sobre Superficies y Rendimientos de Cultivos*), which originates from a yearly survey started in 1995 (MAGRAMA, 2014). The sample of farms is selected from a 1 km × 1 km grid. Surveys are performed at the farm level and include information on cultivated area, average yield per crop, crop varieties, management practices, farm infrastructure and others. The original database comprised 121,309 yield observations, distributed across the study region (CyL). After aggregation at municipality level, the database for the present analysis contained 7879 yield observations.

Expected yields (*Yexp*) prior to sowing and for each of the counties were estimated assuming a linear trend from the seasonal adjusted *Ya* series for the following year. *Yexp* includes spatial variability of the mean expected yield in 531 municipalities across the study region, in 2011 and 2012.

Insurance-related yields

A database from all municipalities in CyL with at least one insured farmer in 2011 or 2012 was made available by ENESA for 107,709 insurance policies over these two years, contracted in 452 municipalities (belonging to a total of 38 counties). Each record includes information on the number of sold policies, total insured production and area, and insured yield per crop insurance module. These records were aggregated by county, year, and insurance module.

The zonal maximum insurable yield (per municipality) for wheat grown under rainfed conditions for each municipality is published every year in the Spanish Official Gazette (BOE). The ones used for this study were published in BOE (2013). These data were aggregated by county to define the average zonal maximum insurable yield (*YinsZ*).

Three other insurance-related yields were defined for the insurance options *Basic* and *Extended*, the latter including the complementary insurance of the option, as described in the previous section:

- *YinsB* is the average yield insured in the option *Basic*, the yield expected before sowing that is freely set by the farmer.
- *YinsE* is the average yield insured in the option *Extended* (Module 1, Module 2 and Module S), yield insured by the farmers before sowing and is limited to an individual maximum insurable yield.
- At midseason, farmers have the option of contracting a complementary insurance policy, in which the insured yield is adjusted according to farmer's yield expectations at that time of season (*YinsEC*).

YinsZ is defined at county level (as it does not vary from a year to another) and *YinsB, YinsE* and *YinsEC* are defined per year and per county.

YinsZ includes spatial variability, including 531 municipalities in the study region. *YinsB, YinsE* and

YinsEC include spatial and temporal variability (2 years) of insured yield data aggregated at municipality level using the mean insured yield in each of the insurance options in the same 531 municipalities.

Attainable rainfed yields

Attainable rainfed yields (Yw) were simulated using the crop model CERES–Wheat (Godwin *et al.*, 1989). CERES–Wheat as available in the package Decision Support System for Agrotechnology Transfer (DSSAT) vers. 4.5 (Jones *et al.*, 2003; Hoogenboom *et al.*, 2012). CERES–Wheat has been widely applied at the regional scale (*e.g.* Bannayan *et al.*, 2003; Zhao *et al.*, 2011) and preferred for the simulation of winter wheat growth at field scale in semi-arid conditions in Spain after a detailed comparison (Castañeda-Vera *et al.*, 2015).

CERES-Wheat was first calibrated and validated using published data from the trials conducted by the Agricultural Research Program of CyL 2004 and 2010 (Table S2 [suppl.]), following Castañeda-Vera *et al.* (2015). The experimental sites are widely spread throughout the grain production areas and are representative of its climate and soil variability. The objective of these trials was to monitor the attainable rainfed production of different wheat cultivars. Therefore, they are assumed to have been grown in optimal rainfed growing conditions, being soil water availability the solely limiting factor. Crop data belonging to the cultivar Marius (*T. aestivum* cv. Marius) were selected to calibrate the crop model. This cultivar is commonly used by farmers in this region. Being used as a test cultivar, it was sown in each of the trials every year. Crop development data such as emergence, anthesis and physiological maturity dates, sowing dates and plant density, and yield were available (Table S2 [suppl.]).

Soil texture and depth maps were obtained from the ITACyL (Agro-Technological Institute in Castilla y León) soil website (http://suelos.itacyl.es/mapas) to simulated rainfed yields. Three soil textures (sandy, loam and clayey, Fig. 1A) and two depth-types (1.0 m and 0.3 m, Fig. 1B) were defined (Table S3 [suppl.]). Three maps with polygons defining depth, soil texture and climate zones were built using ESRI ArcMap 10.0 GIS and combined into a single map. Different climate zones were selected assuming that accumulated rainfall and evapotranspiration are main driving variables for crop growth under rainfed conditions. Accumulated rainfall and evapotranspiration are aggregated in Thornthwaite's aridity index (Thornthwaite, 1948). The map based on Thornthwaite's aridity index in the Agroclimatic Atlas of Castilla y León (Álvarez-Arias

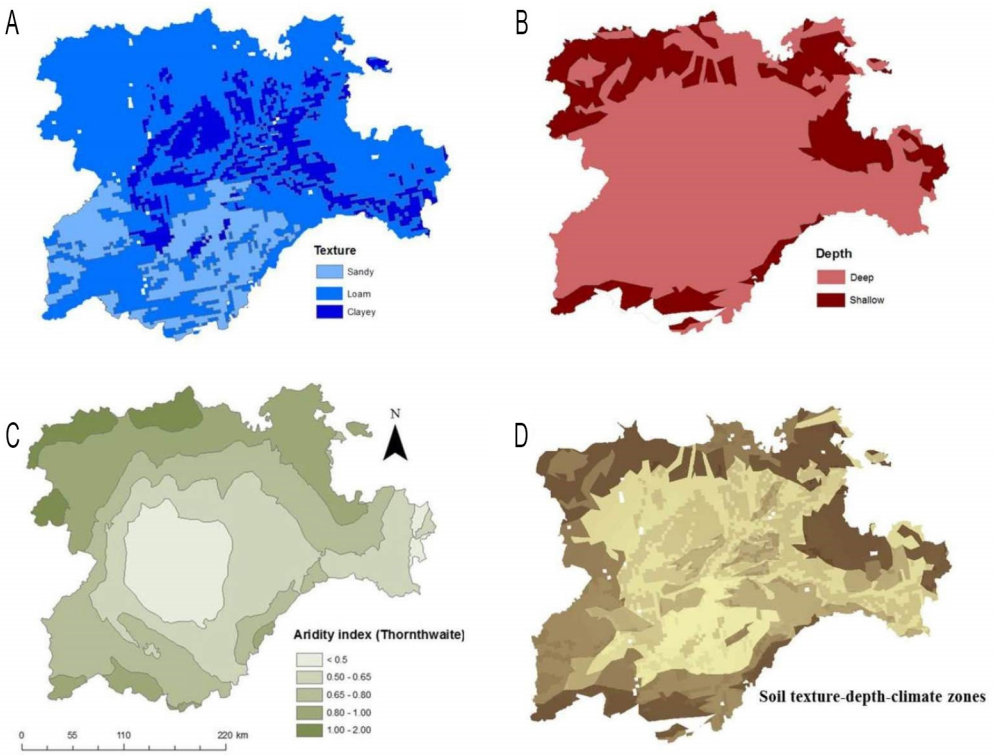

Figure 1. (A) Soil texture, (B) soil depth, (C) Thornthwaite aridity index and (D) soil texture-depth-climate zones in the study region (Castilla y León, northern central Spain). Adapted from the Agroclimatic Atlas of Castilla-y-León (Álvarez-Arias *et al.*, 2013).

et al., 2013) was used to assign a climate zone to each sub-area (Fig. 1C). Recorded weather data assigned to each climate zone was obtained from the System of Agroclimatic Information for Irrigated crops (SiAR), covering 14 years (2000-2013) (Table S4 [suppl.]). The map contained polygons assigned with a number defining a single combination of soil texture-depth-climate zone (Fig. 1D).

Models' statistical performance was evaluated using the root mean square error (RMSE) (Eq [1]):

$$RMSE = \sqrt{\frac{\sum_i^n (S_i - O_i)^2}{n}} \quad [1]$$

where Si and Oi are the simulated and observed yields, respectively, and n is the number of observations used.

The model was run for each combination of soil texture-depth-climate zone in the region. CERES-Wheat simulates wheat biomass production (dry matter). Simulated yields were converted into yields at harvest assuming 13% humidity in order to make them comparable with actual, expected and insured yields. Average simulated rainfed yields (13% humidity) (Yw) were assigned to the corresponding polygon in the map.

Yw includes spatial and temporal variability of simulated attainable rainfed yield data aggregated at municipality level using the mean yield in 531 municipalities all along the study region and per year (2000-2013).

Yield gap analysis

Five yield gaps (GapZ, GapB, GapE, GapEC, GapW) were defined by the following yield differences:

$$GapZ = Y_{insZ} - Y_{exp} \quad [2]$$
$$GapB = Y_{insB} - Y_{exp} \quad [3]$$
$$GapE = Y_{insE} - Y_{exp} \quad [4]$$
$$GapEC = Y_{insEC} - Y_{exp} \quad [5]$$
$$GapW = Y_{w} - Y_{exp} \quad [6]$$

where Yexp is the county average expected yield; YinsZ is the county average zonal maximum insurable yield; YinsB is the county average insured yield in the option Basic per year, and *YinsE, YinsEC* are the county average insured yield each year in the option *Extended* before and after contracting the complementary insurance, respectively. Lastly, *Yw* is the county average simulated attainable rainfed yield.

YinsZ is considered to have a perfect fit when it is equal to *Ya*; indicating that the insurance system would have a perfect knowledge of the actual average yield in the region. Therefore, *GapZ* was assessed as a measure of the accuracy of the insurance parameter *YinsZ*. *GapB, GapE* and *GapEC* indicate the distance between insured and expected yields; they were evaluated as preliminary indicators of asymmetric information for each of the insurance options, reflecting farmer's information advantage with respect to the insurer. *GapE* and *GapEC* were also contrasted in order to evaluate how farmers increase insured yield when uncertainty decreases (insured before sowing *vs*. insured at midseason). Lastly, *GapW* provides insight into how similar or dissimilar simulated rainfed yields are from expected yields, indicating the appropriateness of using crop models simulated data in crop insurance parameterization.

Asymmetric information

An asymmetric information indicator was obtained as the difference between the probability of indemnity based on the risk expected by the insurer, in which possible asymmetric information is assumed, and the one based on the actual risk (no asymmetric information). The procedure to calculate the asymmetric information indicator is illustrated in Fig. 2. The likeliness of indemnity was calculated as the probability of the actual yields being lower than a threshold yield (y_{thres}) (Eq [7], Fig. 2A).

$$Probability\ of\ indemnity = \int_0^{y_{thres}} f(y)dy \quad [7]$$

A probability density function of actual yields, *f(Ya)*, left-truncated at zero, was fitted for each of the counties using the software package @-Risk (Palisade Corp., 2011), selecting the distribution with the lowest Akaike Information Criterion (AIC) statistic. Threshold yield (y_{thres}) was defined as the insured yield (*YinsB, YinsE* or *YinsEC*) after subtracting a 30% deductible, as this is the most common deductible level selected by the farmers. To compute the probability of indemnity based on the actual risk (non-asymmetric information), y_{thres} was calculated from the average actual yield *f(Ya)* (*Ya*)(Fig. 2B). For the insurance options *Basic* and *Extended*, y_{thres} was calculated from the average insured yields (*YinsB, YinsEC*) (shown in Fig. 2C for *YinsEC*).

Lastly, the asymmetric information indicator was defined for the insurance options *Basic* (*AsymB*) and *Extended* before (*AsymE*) and after contracting the complementary insurance (*AsymEC*), as shown in Eqs [8], [9] and [10] and represented in Fig. 2D.

$$AsymB = \int_0^{0.7 \times YinsB} f(Ya)dy - \int_0^{0.7 \times \acute{Y}a} f(Ya)dy \quad [8]$$

$$AsymE = \int_0^{0.7 \times YinsE} f(Ya)dy - \int_0^{0.7 \times \acute{Y}a} f(Ya)dy \quad [9]$$

$$AsymEC = \int_0^{0.7 \times YinsEC} f(Ya)dy - \int_0^{0.7 \times \acute{Y}a} f(Ya)dy \quad [10]$$

where *AsymB* and *AsymEC* are the differences between the probability of indemnity based on the insured yields

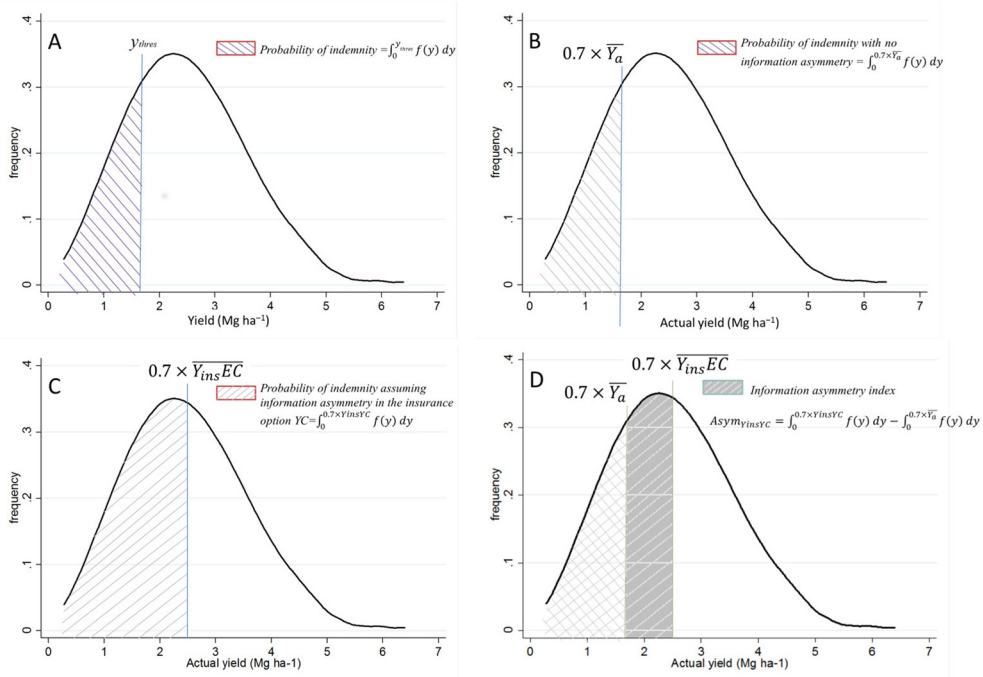

Figure 2. Procedure to calculate the asymmetric information indicator: (A) the probability of indemnity calculated as the cumulative probability of yield being lower than a threshold yield (y_{thres}), (B) the probability of indemnity with no asymmetric information where y_{thres} is the average actual yield after subtracting a 30% deductible, (C) the probability of indemnity with asymmetric information for the insurance option Extended including the complementary insurance (*AsymEC*), where where y_{thres} is the average insured yield in insurance option EC after subtracting a 30% deductible,and lastly, (D) the asymmetric information indicator calculated as the difference between (C) and (B).

($y_{thres} = 0.7 \times YinsB$; and $y_{thres} = 0.7 \times YinsEC$) and the one based on observed actual yields ($y_{thres} = 0.7 \times \acute{Y}a$).

Factors influencing insurance demand

Yield gaps and asymmetric information can potentially influence farmers' insurance demand. Statistical analyses were performed to test their significance in explaining insurance penetration rates. In this work, insurance demand at the county level is thus an aggregation of all farmers' individual decisions about whether insuring or not.

A factor analysis was performed in order to constrain explanatory information to non-correlated variables to be used later on in a regression model exploring the influence on insurance demand.

For that, in a first step, the suitability of data for factor analysis was evaluated using the Kaiser-Meyer-Olkin (KMO) test. The value of KMO was 0.743, enough to conclude that data was suitable for such analysis (Kaiser, 1974). Second, the number of extracted factors was determined using the eigenvalues. We found that restraining factors with eigenvalue higher than 1 was too restrictive for such a model with a high number of variables. Therefore, we used the alternative approach of creating a scree plot (graphing the eigenvalues of all factors listing them in decreasing order of their eigenvalue) and restraining the number of factors above the inflection point (Cattell, 1966). The resulting number of factors was 4. Lastly, factor loadings were computed by an iterated principal factors' algorithm until convergence and rotated using the varimax rotation method.

Factor analysis permitted selecting a single variable from each group (the one with the higher loading). Selected variables were used as independent variables in two linear regression models fitted to investigate the impact of yield gaps and county wheat production characteristics on the decision of farmers in a given county and year to insure their wheat production ($Insured_{wheat}$) and the proportion of insured area under the option basic ($Insured_B$ / $Insured_{wheat}$). The variable $Insured_{wheat}$ was calculated as the fraction of total wheat cultivated area (JCyL, 2014) that was actually insured in any of the insurance modules, per county and year. It thus represents an index of insurance penetration measured at the county level. Models were fitted with Stata v12 (StataCorp, 2011).

Results

CERES-Wheat model calibration and validation

The calibrated crop parameters reproduce a wheat cultivar with high vernalization requirements (60 days), and intermediate photoperiod sensitivity (105 % reduction in rate per 10 h drop in photoperiod), a short grain filling period (400 °C-days), and an intermediate phyllochron for wheat (95°C-days). Parameters related to potential biomass production were 19.5 kernels per unit canopy weight (# g^{-1}), 30.5 kg grain weight, and 2.59 g per standard, non-stressed mature tiller, including grain (Table S5 [suppl.]). Table S5 [suppl.] shows observed and simulated data and the *RMSE* for phenology and grain yield for cultivar Marius. For calibration and validation, *RMSE* for anthesis and physiological maturity were 2.6 and 6.5, and 11.5 and 9.9 days, respectively. For grain yield, *RMSE* were 0.4 and 0.9 Mg/ha.

Yields and yield gap analysis

Figures 3 and 4 show the kernel density estimation of yields and yield gaps including all counties defined within this work. In both cases, they were estimated using Stata v12 and the Epanechnikov kernel. The y-axis has the units and dimensions of the reciprocal of the variable in the x-axis. Thus, density is not measured on a probability scale, and therefore it might exceed 1.

Figure 3 shows the kernel density estimation of the zonal maximum insurable yield (*YinsZ*), expected yield (*Yexp*), and water limited yields (*Yw*) (Fig. 3A), and the *GapZ* and *GapW* gaps, (Fig. 3B). The lowest yields were *YinsZ,* followed by *Yexp* and lastly, *Yw,* with the mean values at 1.96, 2.77 and 3.78 Mg/ha, respectively (Table 1, Fig. 3A). The largest variability was found for *Yw*, followed by *Yexp* and lastly *YinsZ* (Fig. 3A). *GapZ* had the mean at -0.85 Mg/ha and a low variability (percentiles 10 and 90, being -1.45 and -0.37 Mg/ha, respectively), while *GapW* had a mean of 1.01 Mg/ha and a high variability (percentiles 10 and 90, -0.36 and 2.13 Mg/ha, respectively) (Table 1 and Fig. 3B).

Fig. 4A represents the kernel density estimation of expected yields (*Yexp*), and the farmers' insured yield before sowing with option *Basic* (*YinsB*), before sowing in the option *Extended* (*YinsE*), and at midseason insured with option *Extended* and adjusted with the complementary insurance (*YinsEC*). Fig. 4B shows the gaps *GapB, GapE,* and *GapEC*. The lowest yields were *YinsE*, followed by *Yexp, YinsB* and lastly, *YinsEC*, with the mean at 2.22, 2.77, 3.54 and 3.63 Mg/ha, respectively (Table 1, Fig. 4A). *GapE* was the lowest, and had a negative mean at -0.56 Mg/ha, followed by *GapB* with a positive mean at 0.75 Mg/ha, and lastly *GapEC* with the mean at 0.86 Mg/ha (Table 1).

Asymmetric information

To calculate the probabilities of indemnity (Eq [7]) and the asymmetric information indicators (Eqs. [8], [9] and 10]), probability density distribution functions were fitted for actual yields, *f(Ya)*, in each of the 38 counties in the studied region. Out of the 38 distribution functions fitted, 12 were Weibull, 11 Triangular, 4 Gamma, 7 BetaGeneral, 2 Logistic and lastly, 2 Logistic distributions. Such distributions have been used earlier for crop yield modelling (Gallagher, 1987; Nelson & Preckel, 1989; Atwood *et al.*, 2003; Sherrick *et al.*, 2004; Tolhurst & Ker, 2015). Results of these density fits are available from the authors upon request.

Fig. 5 shows the kernel density estimation of the probability of indemnity (Fig. 5A) and the asymmetric information indicator (Fig. 5B). Again, the kernel density was estimated using Statav12 and the Epanechnikov kernel. The lowest probability

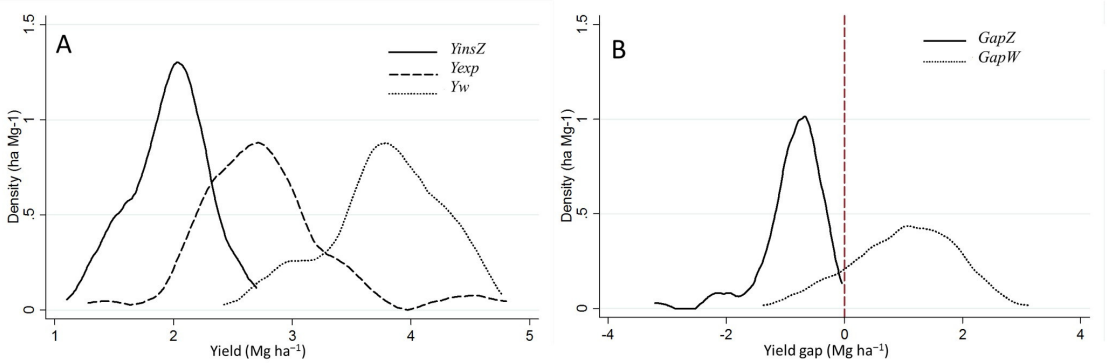

Figure 3. Kernel density estimation of yields and yield gaps: (A) Zonal maximum insurable yields (*YinsZ*), expected yields (*Yexp*) and water-limited yields (*Yw*); and (B) maximum insurable yields to expected yields gaps (*GapZ*) and water-limited to expected yields gaps (*GapW*).

Table 1. Notation, description and descriptive statistics of yields and yield gaps used within the yield gap analysis and the asymmetric information and insurance demand model for winter wheat in Castilla y León. Data is aggregated from the original data base of observations per county.

Variable	Definition	N	Mean	Min	p10
***Area*$_{wheat}$ (1000 ha)**[1]	Area cultivated with wheat in a given year and municipality	79	0.50	0.03	0.17
***Rainfed* (fraction)**	Fraction of the wheat area cultivated under rainfed conditions	79	0.91	0.30	0.78
Ya (Mg/ha)	Actual rainfed yield	79	2.76	1.29	2.18
***CV*$_{Ya}$ (–)**	Coefficient of variation of actual yields under rainfed conditions (1995-2013)	79	0.09	0.02	0.03
***Skewness*$_{Ya}$ (–)**	Skewness of actual yields under rainfed conditions (1995-2013)	79	0.08	-1.61	-1.11
Yexp (Mg/ha)	Expected rainfed yield	79	2.77	1.28	2.20
YinsZ (Mg/ha)	Zonal maximum insurable yield	79	1.96	1.10	1.48
***GapZ* (Mg/ha)**	Gap between YinsZ and the expected yield under rainfed	79	-0.85	-3.21	-1.41
Yw (Mg/ha)	Simulated attainable rainfed yield	79	3.78	2.43	2.99
***GapW* (Mg/ha)**	Gap between Yw and *Yexp*	79	1.01	-1.37	-0.36
YinsB (Mg/ha)	Average insured yield in option Basic	76	3.54	2.22	2.89
GapB (Mg/ha)	Gap between YinsB and *Yexp*	76	0.75	-0.76	-0.16
YinsE (Mg/ha)	Average insured yield in option Extended before sowing	79	2.22	1.17	1.65
GapE (Mg/ha)[1]	Gap between YinsE and *Yexp*	79	-0.56	-2.59	-1.09
YinsEC (Mg/ha)	Average insured yield in option Extended including the complementary insurance	79	3.63	1.82	3.29
GapEC (Mg/ha)	Gap between YinsEC and *Yexp*	79	0.86	-1.61	0.41
***Insured*$_{wheat}$ (-)**	Fraction of the area cultivated with wheat insured in a given year and a given county	79	0.26	0.02	0.17
Insured$_B$ (-)	Fraction of the area cultivated with wheat insured in the option Basic in a given year and a given county	76	0.10	0.00	0.01
Insured$_E$ (-)	Fraction of the area cultivated with wheat insured in the option Extended in a given year and a given county	79	0.17	0.01	0.05
***AsymB* (-)**	Asymmetric information indicator for the insurance option Basic	75	0.17	-0.10	-0.03
***AsymE* (-)**	Asymmetric information indicator for the insurance option Extended	78	-0.08	-0.26	-0.13
***AsymEC* (-)**	Asymmetric information indicator for the insurance option Extended including the complementary insurance	78	0.22	-0.10	0.07

[1]Variables included in the insurance demand models (in bold).

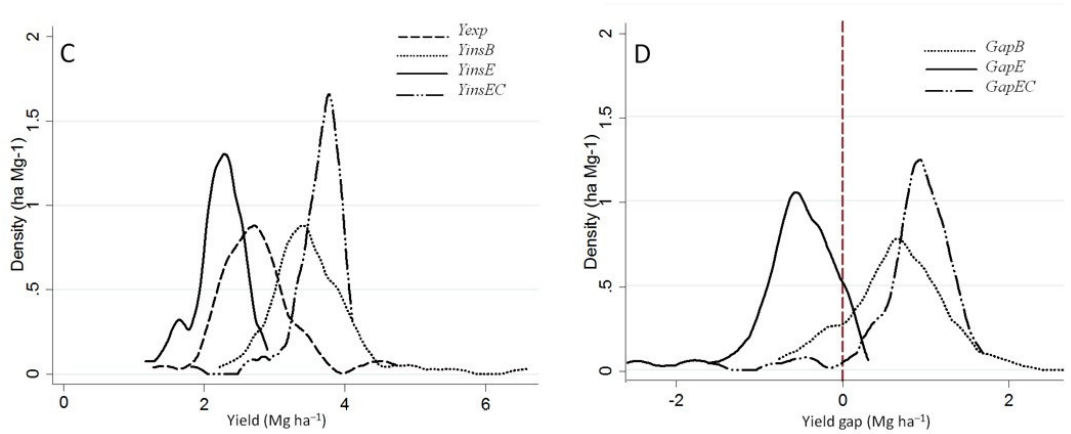

Figure 4. Kernel density estimation of insured yields and yield gaps: (A) expected yields (*Yexp*) and farmer's insured yields before sowing in option Basic (*YexpB*), before sowing in option Extended (*YinsE*), and adjusted at mid-season in option Extended (*YinsEC*), and (B) gaps between farmer's insured yields before sowing in option Basic and expected yields (*GapB*), before sowing in option Extended and expected yields (*GapE*), and the adjusted yields at mid-season in option Extended to expected yields (*GapEC*).

of indemnity was found when using the average insured yield for the option *Extended* before sowing (y_{thres}=*YinsE*), even lower than when no asymmetric information was considered (y_{thres}=average *Ya*). The probability of indemnity when using the average insured yield was higher for the option *Extended* including the complementary insurance (at mid-season, y_{thres}=*YinsEC*) than for the option *Basic* (y_{thres}=*YinsB*) (Table 1, Fig. 5A). The probability of indemnity when no asymmetric information was considered (y_{thres}=average *Ya*) showed the lowest variability. Consequently, the asymmetric information indicator was higher for the option *Extended* including the complementary insurance (y_{thres}=*YinsEC*) than for the option *Basic* (y_{thres}=*YinsB*), and both higher than the option *Extended* before sowing (y_{thres}=*YinsE*) (Table 1, Fig. 5B).

Insurance demand, yield gaps and asymmetric information

Table 2 shows the four factors of the factor analysis including loadings and the percentage of variance of each of them, once rotated. Position of relevant variables determines the factors influencing the demand of wheat insurance in the study region. As a result, 49.5% of the variance is explained by factor 1 variables, defined as "Expectations" in Table 2. Factor 2 variables “Yield asymmetric information” explains 19.5%, factor 3 variables “Yield variability” 17.5 % and lastly, factor 4 “Commercial effort”, 13.5%.

Factor 1 grouped three variables (*GapZ, GapW* and *AsymEC*) indicating a high correlation between them. Higher yield gaps between expected yields and maximum insurable yield (*GapZ*) and attainable rainfed yields (*GapW*) are then correlated with higher asymmetric information for the insurance option Extended when including the complementary insurance (*AsymEC*). This makes sense because when the farmer sets *YinsEC*, the certainty on final yields is higher at mid-season. *GapZ* was selected as the representative variable of factor 1 to be included in the regression models.

Two regression models were fitted using the most representative variables of each of the four factors to explain wheat insurance penetration and the prevalence of the insurance option *Basic* ($Insured_{wheat}$ and $Insured_B$ / $Insured_{wheat}$). Explanatory variables were, therefore, $Area_{wheat}$, *GapZ, CV_Ya* and *AsymB*. Table 3 shows the model parameters estimates of each of the explanatory variables and the significance of their contribution to explain the dependent variable, number of observations (N), and the coefficient of determination (R^2_{adj}).

The results show that wheat insurance demand ($Insured_{wheat}$) was higher in counties with higher wheat cultivated area ($Area_{wheat}$), higher *GapZ* and lower asymmetric information for contracts in option *Basic* (*AsymB*). At the same time, counties with a higher *AsymB* tend to have a higher proportion of insured area under the option Basic. Greater interest in contracting crop insurance would be expected in those counties with higher yield variability (higher CV_{Ya}) and in which crop yield distributions are skewed to the left (lower $Skewness_{Ya}$), with yields closer to the maximum observed more frequently than very low yields. However, model results discarded CV_{Ya} as an influencing factor on insurance demand.

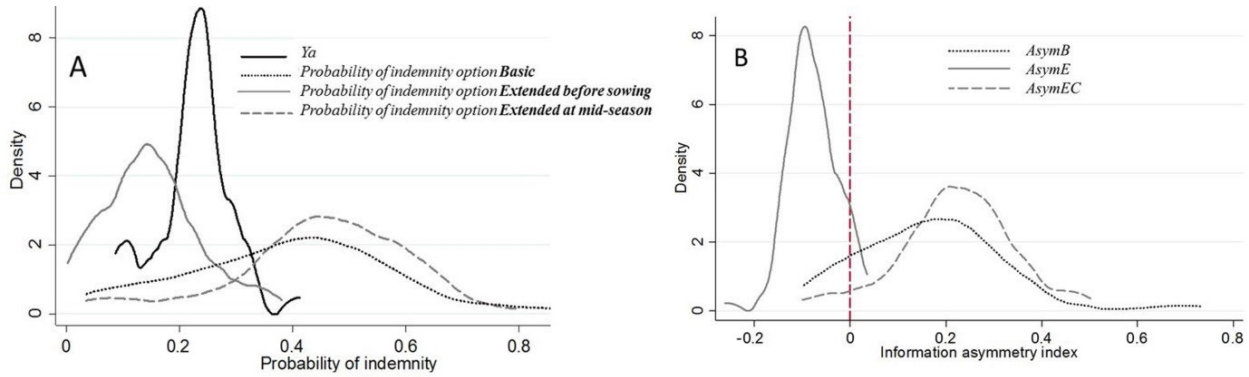

Figure 5. Kernel density estimation of the probability of indemnity and the asymmetric information indicator: (A) Frequency distribution of the probability of indemnity being y_{thres} calculated from average actual yield (solid line, Ya), from the farmers’ insured yields in option extended adjusted at mid-season (dashed grey line, YinsEC) and from the farmers’ insured yields in option Basic before sowing insured (dotted line, YinsB); and (B) Frequency distribution of the asymmetric information indicator for farmers’ insured yields in option Basic before sowing (dotted line, AsymB), and the asymmetric information indicator for farmers’ insured yields in option Extended adjusted at including the complementary insurance (dashed grey line, AsymEC).

Table 2. Factor analysis model including loadings and the percentage of variance of the rotated factors.

Variables	Factor 1 *Expectations*	Factor 2 *Asymmetric information*	Factor 3 *Yield variability*	Factor 4 *Commercial effort*
Area$_{wheat}$ (1000 ha)				0.81
Rainfed (fraction)				
CV Ya (–)			0.80	
Skewness Ya (–)				
GapZ (Mg/ha)	0.86			
GapW (Mg/ha)	0.69			
AsymB (-)		0.90		
AsymE (-)				
AsymEC (-)	0.86			
Proportion (%)	49.5	19.5	17.5	13.5

Blanks represent abs(loading) < 0.5. KMO test value = 0.73

Table 3. Parameter estimates in the regression models for the proportion of the wheat cultivated area that was insured (Insuredwheat) and the proportion of insured area in option basic with respect to the total insured area in 2011 and 2012 in Castilla y León. Observations are per county and year.

Model	Model – Explained variable	
Explanatory variables	**Insured$_{wheat}$**	**Insured$_B$** *option Basic* **/ Insured$_{wheat}$**
Area$_{wheat}$ (1000 ha)	0.08***[1]	---
GapZ (Mg/ha)	0.03*	---
CV$_{Ya}$ (–)	---	---
AsymB (-)	-0.28***	0.78***
N	72	72
R^2_{adj}	0.30	0.17

[1]--- Not significant; * p<0.1; ** p<0.05; *** p<0.01

Discussion

What is driving farmer's agricultural insurance choice?

Premium subsidies are the most common public intervention to incentive crop insurance demand (Babcock & Hart, 2005; Claassen *et al.*, 2005; Garrido & Zilberman, 2008). In the case of wheat insurance in Castilla y León, and for option *Extended* (the one with the largest guarantees), most of the insured area was gathered by Module 2, despite subsidies in Module 1 being higher than in Module 2. The second most selected insurance alternative was Module P, the option *Basic*, with the lowest premium subsidy (Table S1 [suppl.]). Both in Module 2 and Module P, loss compensation is evaluated per plot. This suggests that the accuracy in setting insured yield is decisive in the willingness of the farmers of contracting insurance, being subsidies less important in driving farmers' insurance demand.

Our results show that insurance demand was higher in counties with larger areas cultivated with winter wheat (Table 3). Moreover, higher insurance demand is expected in vast regions with low farmers' density where commercial efforts by insurance companies might be profitable; one successful sale would yield a higher selling commission. Moreover, regions showing stronger crop specialization, and thus lower income sources, incur farm larger revenue risks, being crop insurance an interesting alternative risk management tool (Niewuwoudt *et al.*, 1985; Cabas *et al.*, 2008). Enjolras *et al.* (2014) and Santeramo et al (2016) also found farm size to have an influence in farmers' willingness to select risk management tools, including crop insurance. They argue that that this is related to the higher exposition of smaller farms to changes in their income than larger ones, and to fixed costs associated to the enrolment in the insurance system.

Information asymmetry indicators (*AsymB, AsymE* and *AsymEC*) were also included in the analysis as a measure of farmers' information advantages. Results show higher insurance demand related to a lower *AsymB*, but a higher proportion of the insured area under the option Basic for higher *AsymB*. This suggests

that, in regions with lower asymmetric information and better adjusted insurance parameters, farmers have a higher tendency to contract insurance, and to do it under a wider coverage (option Extended). But the causality reversed: increasing insurance demand might help decreasing asymmetric information as the probability of participation of less risky producers increases, mitigating adverse selection (Shaik *et al.*, 2008). Properly calibrated insurance parameters would make insuring more attractive to a higher number of farmers.

Are crop models useful for crop insurance assessment?

GapW was the highest yield gap found among the gaps included in this analysis. This result suggests that, as expected, it is more accurate to calibrate insurance parameters based on actual reported yields instead of attainable yields (or simulated water limited yields), and therefore on historical data rather than using crop models. Otherwise farmers' yields would be overestimated.

Mean *GapW* in CyL was 1.01 Mg/ha. This differs from those reported in Boogaard *et al.* (2013), within an analysis for the whole of Europe using the crop model WOFOST. They calculated *Yw* between 5 and 6 Mg/ha and *GapW* between 3 and 4 Mg/ha in CyL. Therefore, mean *Yw* was about 1 Mg/ha higher and mean *Ya* was about 1.5 Mg/ha lower than reported in this work. These differences might be related to the accuracy in calibration and the scale of application.

Nevertheless, the use of crop models might be useful in designing new insurance packages when no historical data is available or to evaluate scenarios of expected changes, such as under expected climate change. In that case, it is suggested that yield gaps be estimated and considered when using simulated attainable yields.

Acknowledgements

We thank David A. Nafría-García, Miriam M. Fernández-Sánchez, and Alberto Gutiérrez-García (ITACyL) for the cooperation in soil data gathering, for clarifications on Castilla-y-León cereals farming and production regions and technical help. We also thank María José Martínez-Tineo (ENESA) for the clarifying discussions on cereals insurance, and the Spanish Ministry of Agriculture, Food and Environment (MAGRAMA) for meeting our request to use ESYRCE data.

References

Affholder F, Poeydebat C, Corbeels M, Scopel E, Tittonell P, 2012. The yield gap of major food crops in family agriculture in the tropics: Assessment and analysis through field surveys and modelling. Field Crop Res 143: 106-118. https://doi.org/10.1016/j.fcr.2012.10.021

Álvarez-Arias MV, Cubero-Jiménez D, Fernández-Sanchez M, Garrido-del Pozo N, Gutiérrez-García A, Nafría-García DA, Villarino-Barrera I, 2013. Atlas Agroclimático de Castilla y León. Agencia Estatal de Meteorología and Instituto Tecnológico Agrario de Castilla y León, Valladolid, Spain.

Atwood J, Shaik S, Watts M, 2003. Are crop yields normally distributed? A reexamination. Am J Agric Econ 85: 888-901. https://doi.org/10.1111/1467-8276.00495

Babcock BA, Hart CE, 2005. Influence of the premium subsidy on farmers´crop insurance coverage decisions. CARD Working Papers No. 05-WP 393, Ames, IA, USA.

Bannayan M, Crout NMJ, Hoogenboom G, 2003. Application of the CERES-Wheat Model for Within-Season Prediction of Winter Wheat Yield in the United Kingdom. Agron 95: 114-125. https://doi.org/10.2134/agronj2003.0114

BOE, 2013. Ministerial Order AAA/1629/2013, that defines the insurable goods and yields, minimal crop technical conditions, scope of application, guarantee periods, suscription dates and unitary prices for increasing ceverage insurance for the *Plan 2013 de Seguros Agrarios Combinados*. Boletín Oficial del Estado [España] No. 217, 02.09.2013.

Boogaard H, Wolf J, Supit I, Niemeyer S, van Ittersum M, 2013. A regional implementation of WOFOST for calculating yield gaps of autumn-sown wheat across the European Union. Field Crop Res 143: 130-142. https://doi.org/10.1016/j.fcr.2012.11.005

Cabas JH, Leiva AJ, Weersink A, 2008. Modeling exit and entry of farmers in a crop insurance program. Agr Resour Econ Rev 1: 92-105. https://doi.org/10.1017/S1068280500002173

Castañeda-Vera A, Leffelaar PA, Álvaro-Fuentes J, Cantero-Martínez C, Mínguez MI, 2015. Selecting crop models for decision making in wheat insurance. Eur J Agron 68: 97-116. https://doi.org/10.1016/j.eja.2015.04.008

Cattell RB, 1966. The scree test for the number of factors. Multivar Behav Res 1: 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Claassen R, Lubowski RN, Roberts MJ, 2005. Extent, location and characteristics of land cropped due to insurance subsidies. Am Agr Econ Assoc Annual Meeting. Providence, RI, USA, p. 27.

Coble KH, Knight TO, Pope RD, Williams JR, 1997. An expected-indemnity approach to the measurement of moral hazard in crop insurance. Am J Agric Econ 79: 216-226. https://doi.org/10.2307/1243955

Enjolras G, Capitanio F, Aubert M, Adinolfi F, 2014. Direct payments, crop insurance and the volatility of farm income: some evidence in France and in Italy, New Medit 13 (1): 31-40.

ESYRCE, 2013. Encuesta sobre superficies y rendimientos de cultivo. Madrid.

Eurostat, 2013. Agriculture, forestry and fishery statistics. Luxemburg.

Gallagher P, 1987. U.S. soybean yields: Estimation and forecasting with nonsymmetric disturbances. Am J Agric Econ 69: 796-803. https://doi.org/10.2307/1242190

Garrido A, Zilberman D, 2008. Revisiting the demand of agricultural insurance: The case of Spain. Agric Financ Rev 68: 43-66. https://doi.org/10.1108/00214660880001218

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C, 2010. Food Security : The Challenge of feeding 9 billion people. Science 327: 812-818. https://doi.org/10.1126/science.1185383

Godwin D, Ritchie J, Singh U, Hunt L, 1989. A User's Guide to CERES Wheat, V2.10. Alabama.

Goodwin BK, 1994. Premium rate determination in the Federal Crop Insurance Program: What do averages have to say about risk? J Agric Resour Econ 19: 382-395.

Hoogenboom G, Jones JW, Wilkens PW, Porter CH, Boote KJ, Hunt LA, Singh U, Lizaso JL, White JW, Uryasev, *et al.*, 2012. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.5 [CD-ROM].

JCyL, 2014. Anuario de estadística agraria de Castilla y León. http://www.jcyl.es/web/jcyl/AgriculturaGanaderia/es/Plantilla100/1284228463984/_/_/

Jones J, Hoogenboom G, Porter C, Boote K, Batchelor W, Hunt L, Wilkens P, Singh U, Gijsman A, Ritchie J, 2003. The DSSAT cropping system model. Eur J Agron 18: 235-265. https://doi.org/10.1016/S1161-0301(02)00107-7

Kaiser HB, 1974. An index of factor simplicity. Psychometrika 39: 31-36. https://doi.org/10.1007/BF02291575

Lobell DB, Cassman KG, Field CB, 2009. Crop yield gaps : Their importance, magnitudes, and causes. Annu Rev Environ Resour 34: 179-204. https://doi.org/10.1146/annurev.environ.041008.093740

MAGRAMA, 2014. Metodología de la encuesta anual sobre superficies y rendimientos de cultivos en España. Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid.

Mahul O, Stutley CJ, 2010. Government support to agricultural insurance. Challenges and options for developing countries. World Bank. © World Bank. License: CC BY 3.0 IGO. https://openknowledge.worldbank.org/handle/10986/2432

Nelson CH, Preckel PV, 1989. The conditional beta distribution as a stochastic production function. Am J Agric Econ 71: 370-378. https://doi.org/10.2307/1241595

Niewuwoudt WL, Johnson SR, Womack AW, Bullock JB, 1985. The demand of crop insurance. Agricultural Economics Report No. 1985-16.

Palisade Corp., 2011. @RISK: A Hands-On Tutorial. http://risk.jifsan.umd.edu/files/2013/07/RISKTutorial_ExpertsCorner.pdf

Santeramo FG, Goodwin BK, Adinolfi F, Capitanio F, 2016. Farmer participation, entry and exit decisions in the Italian crop insurance program. J Agr Econ 67 (3): 639-657. https://doi.org/10.1111/1477-9552.12155

Shaik S, Coble KH, Knight TO, Baquet AE, Patrick GF, 2008. Crop revenue and yield insurance demand : A subjective probability approach. J Agr Appl Econ 40 (3): 757-766. https://doi.org/10.1017/S1074070800002303

Sherrick BJ, Zanini FC, Schnitkey GD, Irwin SH, 2004. Crop insurance valuation under alternative yield distributions. Am J Agric Econ 86: 406-419. https://doi.org/10.1111/j.0092-5853.2004.00587.x

StataCorp, 2011. Stata Statistical Software: Release 12. StataCorp LP, College Station, TX, USA.

Thornthwaite CW, 1948. An approach toward a rational classification of climate. Geogr Rev 38: 55-94. https://doi.org/10.2307/210739

Tolhurst T, Ker A, 2015. On technological change in crop yields. Am J Agric Econ 97: 137-158. https://doi.org/10.1093/ajae/aau082

van Ittersum MK, Cassman KG, Grassini P, Wolf J, Tittonell P, Hochman Z, 2013. Yield gap analysis with local to global relevance. A review. Field Crop Res 143: 4-17. https://doi.org/10.1016/j.fcr.2012.09.009

van Wart J, Kersebaum KC, Peng S, Milner M, Cassman KG, 2013. Estimating crop yield potential at regional to national scales. Field Crop Res 143: 34-43. https://doi.org/10.1016/j.fcr.2012.11.018

Zhao H, Gao G, Yan X, Zhang Q, Hou M, Zhu Y, Tian Z, 2011. Risk assessment of agricultural drought using the CERES-Wheat model: A case study of Henan Plain, China. Clim Res 50: 247-256. https://doi.org/10.3354/cr01060