

Feasibility of risk-based inspections in organic farming: results from a probabilistic model

Abstract

A risk-based inspection system might improve the efficiency of the organic farming certification system and ultimately provide a basis for increased competitiveness of this sector. This requires the definition of an effective inspection procedure that allows statistical evaluation of critical risk factors for noncompliance. Here, we present a study based on data from selected control bodies in five European countries that is aimed at determining the feasibility of risk-based inspections in the organic sector according to the data that are currently routinely recorded. Bayesian networks are used for identification of the factors that can affect the risk of noncompliance. The results show that previous/concurrent noncompliant behavior explains most of the risk, and that the risk increases with farm size and the complexity of their operations. The data currently recorded by control bodies appear to be insufficient to establish an efficient risk-based approach to these inspections.

Keywords: Organic farming certification, risk-based inspections, Bayesian networks

1. Introduction

The organic certification system allows consumers to recognize organic products and to be certain that they are actually produced according to specific rules. This system therefore represents a crucial tool for the differentiation of organic products and the generation of comparative advantages for organic producers. However, certification systems need to be effective without the generation of high implementation costs that are specific to organic producers, and that therefore might reduce their competitiveness with respect to nonorganic competitors. A risk-based inspection system, i.e. a planning tool for the development of the optimum system for the carrying out of inspection activities, might increase the efficiency and effectiveness of the organic certification system.

The concept of risk can be interpreted from different perspectives. European Council (EC) Regulation 178/2002 (European Council, 2002), which refers to the general principles relating to food safety, defines the risk in the food sector as “a function of the probability of an adverse health effect and the severity of that effect, consequential to a hazard” (Article 3(9)). The approach of the European Council (EC) Regulation 834/2007 (European Council, 2007), which refers to organic production and labeling, refers to the risk in a narrower sense, as the probability of not fulfilling the requirements laid down in the organic regulations. In the present study, we use the latter approach, using the terms of probability or risk of noncompliance as equivalent, to investigate the potential for a risk-based approach for inspection procedures in the organic sector.

The organic farming regulations explicitly consider that inspections and control procedures of organic operators should be based on risk evaluation (European Council, 2007, Article 27(3)). The need for a risk-based approach is also indicated in the European Action Plan for Organic Food and Farming, as a key action to be undertaken (Commission of the European Communities, 2004, Action 13): “Improve the performance of the inspection bodies and authorities by introducing a risk-based approach targeting operators presenting the highest risk in terms of fraudulent practices, and by requiring cross-inspections under Regulation (EEC) No 2092/91”.

Risk-based inspection management for the organic certification system might produce benefits in three areas: cost effectiveness, standardization, and audit potential of the inspection procedure across the control bodies. Cost effectiveness, in particular, can be achieved through a more effective design of the inspection scheme that is based on probabilistic approaches that can indicate the crucial risk factors in terms of the potential for infringement. This would reduce ‘unnecessary’ inspections and concentrate resources on the operators who present greater risk of noncompliance. Albersmeier et al. (2009) proposed a risk-oriented approach for the improvement of the efficiency of the certification system in the agro-food system.

Risk assessment should be based on scientific process (European Council, 2002, Regulation 178/2002, Article 3(11); European Council, 2004, Regulation 882/2004, article 3.1), and a risk-based inspection system should therefore use findings from a formal risk analysis, according to the defined criteria for inspection planning and the physical inspection procedures. This requires an empirical analysis of the potential risk factors, based on recorded evidence. See van Asselt et al. (2012) for a review of the methods for the implementation of risk-based controls in a food safety regulatory framework.

The aim of the present study is to determine whether a risk-based inspection procedure for the organic sector could be operatively implemented with the existing inspection data. The data used here, as the basis for our analysis, are those routinely recorded from five organic control bodies in Europe. The need for a higher harmonization in data collection emerges clearly when approaching the problem of comparing inspections records from different countries and /or control bodies. In fact, at present the situation cannot be considered as fully satisfactory, as also the European Commission (2007) comments. The definition of noncompliance in Regulation (EC) 834/2007 (European Council, 2007) is not particularly precise and lacks a clear indication about how noncompliance should be classified in terms of type and severity.

The homogenized datasets from the control bodies are analyzed using a probabilistic risk-modeling approach that is based on Bayesian networks. We analyzed the risk patterns that relate to noncompliance with the organic certification rules according to a set of potential risk factors. The purpose was to single out the impact on the probability of noncompliance due to the individual variables considered, and the joint effects on the probability of noncompliance due to different combinations of risk factors. To our knowledge, no previous similar study has been performed on such a vast dataset, with a comparison of results from different countries. Zorn et al. (2012) analyzed the differences in the German organic control bodies, Zorn et al. (2013) provide an analysis of noncompliance based on data from a German control body, while Gambelli et al. (2011) and Gambelli et al. (2012) analyze the noncompliance of Italian organic operators; finally Zanoli et al. (2012) provide an analysis of noncompliance in the Turkish organic sector.

The structure of this report is as follows: in the second section, a description of the data is provided; the third section describes the methodology used; the fourth section shows the results; the fifth section discusses the main results, and outlines the main conclusions from the study.

2. Data and variables

2.1 Description of the data

The data used in this study were extracted from the standard inspection databases of the following control bodies in five European countries: Bioinspecta (Switzerland), the Institute for Marketecology (IMO; Germany), the Danish Plant Directorate/ Danish Veterinary and Food Administration (Denmark), the Institute for Ethics and Environmental Certification (ICEA; Italy), the Soil Association (SA; UK). These control bodies, with the exception of IMO, are those with the largest share of certified organic farmers in the respective countries in 2009, ranging approximately between 24% (ICEA) to 91.3%¹ (Danish Plant Directorate). IMO accounts for about 10%² of certified organic farmers in Germany, and though it is not the leading control body in this country, it was chosen due to its international relevance and expertise, which helped us in the definition of a common framework of analysis across countries. Though statistical representativeness for all countries cannot be assured, the available dataset provides however detailed information for a considerable share of organic farmers.

We used the standard data that are routinely recorded by these control bodies to examine the development of risk-based inspections according to these data: no extra information was requested. The datasets contain information concerning the level and type of noncompliance and the related sanctions.

These refer to the time between 2007 and 2009 at the farm level, together with a description of the farm structures in terms of the physical size, and the type of crops and livestock, with some limited information on farm management with respect to their organic operations. The number of farms certified by control bodies is of course varying through time. We have however preferred to balance the panel, to have a more stable dataset and to avoid missing data problems concerning farmers that are not in the dataset for all the years of the analysis. The total number of observation for all countries is 15,779 per year.

As these data came from different sources and countries, a homogenization of the labels and variable definitions across the sources and countries was needed: the EUROSTAT coding for structural, crop and livestock data was taken as the reference (Commission of the European Communities, 1999).

With respect to noncompliance data, we only have information about the *detected* noncompliance, which will not necessarily coincide with the actual total level of noncompliance. The data on noncompliance are recorded as the result of the inspections performed by the control bodies. Each organic operator undergoes at least one inspection visit per year, while additional inspections are carried out on a sub-sample of operators on the basis of inspection schemes

¹ Though Denmark has only one national control body, some organic operators could ask to be certified by other foreign control bodies; also, note that in Denmark inspections and certification are managed by public control authorities.

² 22 control bodies are operating in Germany; IMO represents the third most important control body in terms of share of organic farmers.

decided independently by each control body, without a uniform approach to decide the frequency and share of unannounced or follow-up inspection visits. Before any statistical analysis can be performed, the data relating to noncompliance required further transformation and standardization.

Unfortunately, these different control bodies do not have a univocal way of recording and classifying these data. On the one hand, noncompliance is not coded in a way that allows univocal identification of the degree of severity in each of the countries. EC Regulation 834/07 (European Council, 2007) only refers to irregularities and infringements, where the former are less severe and the latter are more severe violations of the Regulation, and no other classifications are provided at the regulatory level. On the other hand, noncompliance information was not available for all of these control bodies or for all of these years. Therefore, the comparability of the data from different sources is limited without a correct appraisal of the extent of agreement of the data from these different sources.

However, most of the control bodies recorded the sanctions in a similar way, and this information can be ranked easily according to the severity. The UK control body did not record the data on the sanctions in the database, although the data on noncompliance is at least classified according to the degree of severity.

To proceed with the analysis, we therefore assumed that a severe sanction is issued when severe noncompliance (or an ‘infringement’) is detected, and a less severe sanction is issued in response to correspondingly less severe noncompliance (or an ‘irregularity’). In other words, the sanctions were used as a proxy for noncompliance for all of the control bodies (except in the UK), which means that we assumed that noncompliance was followed by sanctions with the appropriate level of severity.

At the same time, we had to standardize these sanction data across the control bodies, as they do not follow a common sanction scheme. A standardization of the coding was implemented for Switzerland, Germany, Denmark and Italy. Here the sanctions were classified either as ‘slight’ or ‘severe’, to maintain the concept of ‘irregularities’ and ‘infringements’ as used in the EC Regulation. The sanction scheme agreed upon by all of the control bodies in Italy served as the basic framework for the standardization and recoding of the sanctions. This was based on a clear correspondence between sanctions and noncompliance, which was structured around a limited number of sanction types (SINCERT, 2009), while the Swiss, German and Danish data have a higher number of sanction types. These were thus aggregated according to the severity of the noncompliance that they refer to. At this stage, the expertise from qualified personnel of the control bodies had a crucial role. In the UK, in accordance with the practice used by the data provider, we considered minor and major noncompliance as ‘slight’ (i.e. ‘irregularities’), while the ‘severe’ classification was used for all of the critical noncompliance recorded in the database.

Table 1 gives the different numbers of original sanction types that were behind the two harmonized sanction categories. The situation across countries is different, with Switzerland and Denmark showing higher numbers of sanction types, while Italy and Germany share a more similar sanction structure.

<Place Table 1 about here>

In the following, we will refer to slight and severe noncompliance, as indirectly measured according to the recorded sanctions for all of these countries, except the UK. Also, for ease of discussion, we refer to the results in terms of the specific countries, although we have to make clear that the results are only those that arise from the analysis of the selected control bodies, and therefore they cannot be generalized at the respective country level.

The distribution of the noncompliance shown in Table 2 illustrates some of the important differences across these countries.

<Place Table 2 about here>

The level of noncompliant farms decreases with respect to the severity of the noncompliance for all of the countries except Switzerland, and this is particularly low for Denmark. The level of slight noncompliance, on the other hand, is particularly high for Germany, which showed that between 43.4% and 50.1% of farmers had received at least one sanction of this type. For the UK, there was a similar situation for the slight noncompliance, and an increase in the relative level of severe noncompliance in 2009.

2.2 The variables

It is important to consider that the information relating to the level and type of noncompliance is associated to each farm for each year, with no more detailed information available in terms of what generated the noncompliance. In other words, we know, for example, that a farm committed “severe noncompliance” maybe due to the use of forbidden inputs, but we do not know what crop or livestock type these forbidden inputs were related to. Therefore, there was the need to associate the level and type of noncompliance with the other *farm-specific* or *farmer-specific* variables recorded in the datasets, to uncover the specific factors that contribute to the risk of noncompliance.

Of course, the list of potentially relevant risk factors is limited by the data availability. For example, we might believe that information on the solvency of a farmer is relevant, but this sort of data is not collected nor recorded by the

control bodies. Unfortunately, only a limited number of *farmer-specific* variables were available, while most of the potential risk factors refer to the farm operations and structure.

For the purpose of our analysis, the potential risk factors were coded as dummy variables (1, presence of the risk factor; 0, lack of the risk factor) or as discrete variables that assumed a limited number of values.

The potential risk factors that we considered in our analysis are the following:

Path-dependence of noncompliance (dummy variable): noncompliance that occurred in the past. Occurrence of noncompliance in 2007, either slight or severe, is considered in the 2008 models, while occurrence of noncompliance, either slight or severe, in 2007 and/or in 2008 is considered in the 2009 models;

Co-dependence of noncompliance (dummy variable): cross-effects between slight and severe noncompliance in the same year.

Participation in other certification schemes (dummy variable): other certification schemes besides the organic scheme, like ISO environmental schemes, Demeter certification, and others, are considered.

Farmer experience in the organic sector (discrete variable): in our database we only have information relating to how long a farmer has been certified with the control body that provided the dataset (number of days since joining the control body), so we do not know if the farmer was previously certified by other control bodies: this *proxy* variable might, therefore, underestimate the actual farmer experience. Three states are considered: low-, average- and high-experience farmers (thresholds vary according to country-specific distributions).

Licensee (dummy variable): license for the selling of organic products. Selling organic products on the organic market increases the potential impact of undetected noncompliance. In two countries (Italy and Denmark) the certified farmers need to require a specific license from the control body to sell their products, and this flags them as of greater risk, thus increasing the inspection effort. However data are available in a usable form only for Italy.

Farm size (discrete variable): this potential risk factor refers to the number of hectares of utilised arable area (UAA) of the farm. Three states are considered: small, intermediate and large farm sizes (thresholds vary according to country-specific distributions).

Herd size (discrete variable): this potential risk factor refers to the number of livestock units in the farm. Three states are considered: small, intermediate and large size of livestock production (thresholds vary according to country-specific distributions).

Number of crops (discrete variable): this potential risk factor is a proxy for the complexity of the farm structure, which can impact on its organic operations. Three states are considered: low, average and high number of crop categories (thresholds vary according to country-specific distributions).

Number of livestock species (discrete variable): this potential risk factor is another proxy for the complexity of the farm structure. Four states are considered: zero, low, average and high number of livestock types (thresholds vary according to country-specific distributions).

Livestock density (dummy variable): the EC organic Regulation imposes a maximum of two livestock units per hectare. A farm can derogate from this limit if it contracts forage and feed-land outside, and this might increase the risk of noncompliance.

Presence of specific crops and livestock types on the farm (dummy variables): given their sparseness, the data on crops were aggregated. The crop categories considered are: cereals, dried pulses, root crops, industrial crops, fresh vegetables, green fodder, other arable crops, permanent grassland, fruit, citrus, grapes, olives, and other permanent crops. Not all of these crop types are available for all of the countries (e.g. olives are available only for Italy). The livestock types considered are: bovines, pigs, sheep, goats, and poultry. These variables are used to determine whether some crops or livestock are more at risk.

On-farm processing (discrete variable): farmers with on-farm processing have more obligations to which they need to comply, and the risk might increase according to the number of different processed products. Three states are considered: zero, low and high number of processed products (dimensional thresholds vary according to country-specific distributions).

Genetically modified organism (GMO)-risk crops (dummy variable): maize and/or soya bean might increase the risk of noncompliance, as for these crops (accidental) contamination of seeds can occur.

Conversion and/or conventional land (dummy variable): the presence of non-organic land on the farm is considered.

Table 3 shows descriptive statistics for the potential risk factors considered in the analysis. Means and standard deviations are provided for discretised variables, and relative frequencies are provided for dummy variables³.

<Place Table 3 about here>

3. Statistical methods

Probabilistic networks were originally proposed by Horvitz et al. (1988), Howard and Matheson (1981), Pearl (1986, 1988), and Lauritzen and Spiegelhalter (1988), among others. These are considered as a feasible solution for the analysis of complex models under uncertainty, particularly without complete knowledge of the causal structure of the system itself, or without complete information about the potential effects of an action to be taken (Russell and Norvig, 2011; Cowell et al., 2007). Probabilistic networks can be used to learn about causal relationships in a system, and to associate probability values to a causal semantic. This results in an effective and flexible tool for making inferences, as it allows consideration of the change in the probability of the state of a variable when one or more other variables change. Gambelli and Bruschi (2010) and Kristensen and Rasmussen (2002) used probabilistic networks for simulating the exit of organic farmers from the sector, and for decision support in low input farming, respectively. Gambelli et al. (2011) used a probabilistic network to model the risk of noncompliance using a smaller and different dataset of Italian organic farms.

Here we use a discrete Bayesian network, i.e. a probabilistic network based on a finite set of discrete random variables that form a directed acyclic graph (DAG). A DAG is a graph with directed edges that join a set of nodes, or vertices (Figure 1). It is acyclic in the sense that there is no path starting from a node x_v and leading back to it. If there is a directed edge from $x_{pa(v)}$ to x_v , then given $x_v \in X$ we define $x_{pa(v)}$ as the parent node of x_v .

<Place Figure 1 about here>

Given a set V of nodes, X is the set of our random variables, and can be indexed by V : $X=(X_v)$ where $v \in V$.

³ Dichotomous variables follow a Bernoulli distribution with mean = p , and variance = $p(1-p)$, i.e. the variability is the higher, the closer the frequency is to 0.5.

The set of potential risk factors described in the previous section represents the set of ‘nodes’ that define the network. In our case, we ‘learned’ the network starting from a dataset of observed cases, mainly due to the general lack of prior information about the type of connections among the various noncompliance and risk factors.

The learning process goes through a two-step procedure. First, the network structure is determined (i.e. the connections among the nodes); secondly, the parameters of conditional probabilities (i.e. the ‘strength’ of the connections) are estimated. For the structural learning we used the PC algorithm, which is a robust and efficient solution, particularly for large datasets and sparse networks (Cooper and Herskovits, 1992; Kalisch and Buhlmann, 2007). For details of the PC algorithm, see Sprites et al. (2000), and for a discussion of the learning procedure for discrete data networks, see Sprites and Meek (1995). The analysis was performed with the Hugin 7.5 software. The parameters of the conditional probabilities were estimated with the expectation maximization algorithm (Dempster et al., 1977, Lauritzen, 1995). This is a two-step procedure: a first parameter estimate (expectation step) is followed by a new estimation of the parameter once the complete data set is considered (maximization step): this procedure is particularly robust, even in the case of parameter estimation with missing data. A Bayesian network was built for each country and year (2008 and 2009), and used as a basis for inference on the risk of slight and severe noncompliance. In the structural learning process we imposed as a constraint that the nodes representing noncompliance should only receive, and not generate, inputs from potential risk factors. Also, time direction was imposed to consistently model the path dependence of noncompliance. No other constraint was imposed in the learning procedure. Inference is performed on the basis of *evidence*, i.e. that a random variable X_v actually takes one of the possible states: $X_v = x_v$. In other words, the evidence shows that node X_v is actually in the state x_v .

If we define E as the set of such evidence, then we can calculate the posterior probability distribution $P(X|E)$ for all of the nodes X in the network (Jensen, 1996):

$$P(X|E)=P(X,E)/P(E) \tag{1}$$

Evidence can also be used to evaluate the probabilities of a set of variable combinations. Basically, two types of inference can be considered (Russell and Norvig, 2011): a diagnostic inference (given a set of evidence for child nodes, we go back to the most likely causes, i.e. the most likely configuration of the parent nodes); and a causal inference (given a set of evidence for parent nodes, we infer the most likely effects on the child nodes). For instance, if we would like to know the probability of severe noncompliance for a farm that produced green fodder and bovines, and that had already committed a slight noncompliance in the past, the evidence set would be: e_1 : “Green fodder = yes”; e_2 : “Bovines = yes”; e_3 : “Slight noncompliance in the past = yes”. It is not necessary to define the evidence set for all of the

variables; this allows the consideration of a generic farm type (e.g. bovine farm with noncompliance commitment in the past) regardless of all of the other variables, like size, other crop production, etc. Of course, the higher the level of evidence imposed, the lower the sample size on which the final conditioned probabilities are computed, and *vice versa*.

In our study, we are particularly interested in causal inference, i.e. identification of the risk factors that provide the most significant impacts in terms of probability variation of noncompliance. Hence, two types of analysis were considered. The first is a single-parameter sensitivity analysis that measures the effects of each risk factor individually according to the variation of the risk of noncompliance. The initial probabilities of noncompliance are taken as the reference, and a variation larger than 10% and 50% are labeled as + and ++, respectively (and - and - - for the equivalent negative variations). The second type of analysis is a multiple-evidence simulation, which refers to the cumulative effects for the risk of noncompliance due to a set of risk factors jointly considered, known as the **common effects**. A common effect is in place when a child node (the effect) has two or more causes (parent nodes). The parents, therefore, are marginally independent, but conditionally dependent (Korb and Nicholson, 2004, p. 41). Different sets of risk factors can be defined, with each ideally referring to a farm type. As the number of possible risk factor combinations is very high, for each country we consider only the two farm types associated with their highest probabilities of noncompliance. A conflict indicator was considered to check the consistency of the evidence imposed in the simulations (Jensen et al., 1990). Conflicts can occur among variables when the data do not support the evidence imposed. For instance, if we consider two pieces of evidence, e_1 : “Green fodder = Yes” and e_2 : “Bovines = No” and only a few cases in the dataset match this combination, then a conflict occurs. Given a set of evidence $E = (e_1, \dots, e_n)$, then the level of conflict can be measured as:

$$C(E) = \log\left(\prod_{i=1}^n (prob(e_i)/prob(e_1, \dots, e_n))\right) \quad (2)$$

Negative values of C indicate a lack of conflict in the evidence set, i.e. high ‘compatibility’ of the evidence set E (Jensen et al., 1990).

The performance of the networks as good classifiers for noncompliance was measured using receiver operating characteristic (ROC) curves, which can be considered as binary classifiers (Hanley and McNeil, 1982). As specific values can be computed for each state of noncompliance (i.e. noncompliance = yes, or noncompliance = no), the average values are considered. A ROC curve allows the inspection of the performance of a given variable as a classifier for the dataset, plotting the X-axis as the false-positive rate, and the Y-axis as the true-positive rate. The area under the ROC curve (AUC) provides a measure of the classification performance of the model. Meaningful AUC values range

between 0.5 (purely random classification of cases) and 1 (perfect classification of cases). Swets (1998) suggested the following levels of AUC values for model diagnostics: 0.5: non-informative classification; 0.5-0.7: low-accuracy classification; 0.7-0.9: average-accuracy classification; 0.9-1 highly accurate classification; 1: perfect classification.

4. Results

4.1 Model diagnostics

A separate Bayesian network model was developed for each of the five countries, with the 2008 data and replicated in 2009⁴. The 2007 noncompliance data were used as lagged variables, for testing path-dependence in both years, while the 2008 noncompliance entered the 2009 model for the same reason. Table 4 gives the relevant model diagnostics. The AUC values in our models range between 0.6 and 0.9, which are not particularly high, although they are above the critical 0.5 threshold.

<Place Table 4 about here>

4.2 Single risk factors

Table 5 shows the results of the impact of single risk factors on the probability of noncompliance (parameter sensitivity). The impacts were evaluated for each risk factor in terms of the largest variation either in 2008 or in 2009 with respect to the initial probabilities of (slight and severe) noncompliance, i.e. the expected likelihood of noncompliance that results when no evidence is imposed in the model. The results refer to the relevant impacts on the noncompliance probabilities that occurred in 2008 and 2009, to list the relevant risk factors over a two-year time span. The impacts due to each risk factor are labeled as positive if they show a directly proportional variation of probabilities of noncompliance, and *vice versa*⁵. Entries labeled as not applicable (n.a.) refer to risk factors where data are not available, or could not be used in the analysis because they are not discretizable (e.g. risk factors with 0%, or 100% of cases), or have an extremely sparse distribution.

<Place Table 5 about here>

⁴ Ten Bayesian networks have been estimated. They are not shown here for reasons of space, they are however available for the interested readers upon request to the authors.

⁵ The largest impact in absolute value occurring between 2008 and 2009 is recorded: e.g. if the *i*-th risk factor has a positive impact on the (initial) probability of noncompliance of 10% in 2008 and of 50% in 2009, it is labelled as “++” in Table 5. Analogous approach is used for negative impacts. Note that no cases of changes of sign of the impacts across the two years are recorded.

A distinction can be made between the effects of path-dependence and co-dependence, and those of all of the other variables. The former effects are generally quite strong for all of the countries and the years. In particular, the co-dependence effect leads to impacts above 50% for all countries, while the effects of path-dependence are slightly more differentiated, although they emerge, however, as an important risk factor.

For other farm-specific risk factors, the effects on the risk of noncompliance are generally lower than those induced by co-dependence and path-dependence. A few common risk factors can be found in most of the countries; namely, farm size, herd size, number of livestock species, and number of crops.

The smallest farm and/or herd size is always associated with the minimum risk, although the probability of noncompliance does not necessarily increase linearly with size: e.g. in Germany, the farms showing greater risk are the largest in size (UAA), while in the UK the larger the herd size, the larger the risk. In the other countries, the intermediate farm/ herd size usually showed the highest occurrence of noncompliance.

In general, livestock production emerges as a common risk factor for all countries, although we can differentiate the impact according to the livestock species. Pigs and poultry production increase the risk for at least one category of noncompliance in all of these countries except Italy for what concerns pigs. The impact of poultry production is particularly relevant in the German case, and for pig production in Switzerland. The production of green fodder, which can be considered as strictly linked to livestock production, is a relevant risk factor for all of the countries except Denmark and Switzerland.

When considered individually, the number of crop types has a positive impact on the probability of both slight and severe noncompliance in Italy and Germany and UK in particular, and a positive impact on the probability of slight noncompliance in Switzerland and Denmark.

Also GMO-risk crops show a positive impact on severe noncompliance in Italy, Germany, and in Switzerland for slight noncompliance.

Finally, we note that fruit, sheep in Denmark, and citrus (available only in Italy), are the only factors that show negative impact on the risk of severe noncompliance.

4.3 Analysis of the common effects

The interpretation of Table 5 is quite straightforward, which shows the individual contributions of each risk factor to the risk of noncompliance, and this provides a first insight into an understanding of the risk patterns in the different countries. However, the advantage of using Bayesian network models is connected to their simulation of the impact of combinations of different variables on risk; in other words, their ability to analyze common effects.

To explore common effects, two farm types were considered for each country (Table 6). Farm types have been defined as consistent combinations of risk factors yielding high probabilities of severe noncompliance, still maintaining a decent sample size. In fact, the joint consideration of a set of risk factors may considerably reduce the sample size, particularly when co-dependence and path-dependence effects are considered (due to the low share of noncompliant farmers). The consistency of the combination of each risk factor is measured by the conflict scores. The farm-type simulation shows how the combination of multiple risk factors can lead to a considerable increase in the probability of noncompliance, with a multiplicative final effect. Also, some cases emerged where the individual consideration of the risk factors with moderate impact on the risk of noncompliance can contribute to a strong increase in risk when combined with other factors. The conflict scores are always zero or negative, which thus indicates the consistency for the combination of risk factors considered in the farm types.

The co-dependence effects between slight and severe noncompliance is particularly strong in most of the cases. The information about the occurrence of a noncompliance can be a guide for the control bodies both for planning unannounced inspections, and can represent an alert providing relevant additional information in the inspecting procedure also for annual visits. In Table 6 we explicitly show the effects of the occurrence of noncompliance in 2009 in the risk of severe noncompliance and in the sample size in the different countries; for each farm type the sample size and the conflict scores refer to the complete set of risk factors, including the 2009 occurrence of slight noncompliance.

<Place Table 6 about here>

The German case is characterized by a high share of slight noncompliance, and a low share of severe noncompliance. In absolute terms, therefore, in the simulations of the two farm types, the probability of slight noncompliance reaches very high values, but in relative terms the increase in the severe noncompliance probability is proportionally higher. Farm type 1 is a large arable farm with permanent grassland, cereals, some land under conversion, and noncompliance committed in the past. In this case, the impact of the co-dependence effects of noncompliance in 2009 is not as strong as in other countries. Farm type 2 refers to a medium-sized farm with pigs production, cereals and noncompliance in the past. Also in this case, the co-dependence effect is not particularly large.

For all of the other countries, the effects of co-dependence of noncompliance are more evident. The Danish case shows particularly low initial probabilities for severe noncompliance, and both farm types refer to livestock farms. Farm type 1 is a medium sized farm with large bovine production, and noncompliance in the past. The impact of the co-dependence of noncompliance produces, however, a dramatic increase in the risk of severe noncompliance. The

dominance of this effect is confirmed by the results of farm type 2, a poultry farm. The final probabilities reach the same values as farm type 1 once the co-dependence effects of noncompliance are considered.

For the Italian case, farm type 1 represents a large arable farm, with cereals, and green-fodder production, which has committed noncompliance in the past. Farm type 2 is a large bovine breeding farm, with green fodder production and noncompliance in the past. For both farm types, the co-dependence and path-dependence effects are particularly relevant.

Farm type 1 for Switzerland refers to a farm with pigs breeding that was noncompliant in the past. Farm type 2 is an arable crop farm, with noncompliance in the past. Also for these farm types the co-dependence effect is particularly evident.

The British farm types both refer to livestock production. Farm type 1 is a medium-size farm, with poultry production and noncompliance in the past, the effects of which account for the largest part of the final probability of severe noncompliance. A similar situation is described by farm type 2, which refers to a large farm with bovine production, green fodder and noncompliance in the past.

Summing up, these farm-type simulations confirm the critical roles of path-dependence and co-dependence as the most prominent risk factors in our analysis.

5. Conclusions

Our analysis is based on data from selected control bodies in five European countries, and it therefore represents only a part of the organic operators, and was based over a limited time span. It should thus not be considered as representative of the organic sector in Europe. Our results can therefore be considered as a first insight into the still little-investigated factors that could have an impact on the risk of noncompliance in the organic sector, without attempting to generalize the results.

Notwithstanding these caveats, for all of the countries and years considered, we can summarize the results of our analysis by saying that co-dependence and path-dependence of noncompliance are largely the most important risk factors measured by our models. Other relevant risk-increasing factors are, although to a lower extent, farm size and complexity, and livestock production in general. These results are consistent with those found in Zorn et al. (2013) and Gambelli et al. (2012).

Given the almost total lack of relevant personal information at the level of the farmers, path-dependence and co-dependence represent indirect indicators (*proxies*) of the operator propensity to noncompliance.

On the basis of these results, we can formulate an answer to our initial question, i.e. if the data currently recorded by control bodies provides an effective base for a feasible risk-based inspection system in the organic sector.

As the largest part of the risk of fraud is determined by information relating to the detection of noncompliance, the present structure of the data recorded by the control bodies, which are mainly oriented towards farm-specific structural information, is probably not ideal for an efficient and effective risk-based inspection system. On the one hand, other factors are probably more relevant to explain the risk of noncompliance, although these are currently not recorded in the control body archives, and in most cases will not even be known to the inspectors. On the other hand, the relevance of path-dependence and co-dependence in explaining the risk suggests that noncompliant behavior has, most likely, more to do with the farmer than with the farm. Efforts should be made to integrate the information at the operator level, like gender, age, personal crime record, financial information (e.g. solvency rate, liabilities, debts), which might contribute significantly to the explanation of the non-compliant behavior of some farmers. There is also an almost complete lack of data relating to the economic size of the farms, in terms of turnover, profit, and level of equity and investment.

We should also consider that the measurement of the risk of noncompliance is necessarily contextualized to the period it refers to. A constant update of the risk assessment could exploit the additional information deriving from the update of the inspections visits' archives. Regulation (EC) 882/2004 (European Council 2004) explicitly considers the need for information on past records and on controls carried out in the past for the organization of official controls.

Our paper has focused on the concept of risk, in terms of measurability of the probability of the occurrence of noncompliance. This type of information, though essential for the development of risk based controls, could be integrated with additional evaluations to develop a consequential prioritization of the inspections to be performed by the control body (see van Asselt et al., 2012). In particular we believe that inspections planning by control bodies, besides the evaluation of the probability of noncompliance's occurrence, should also consider the potential consequences of the noncompliance, on the basis of the characteristics of the operator and/or of the supply chain where she operates. Among the aspects that could be taken into consideration we mention for instance the economic size of the supply chain, the positioning of the operator in the supply chain structure (e.g. noncompliance in the feed sector might propagate effects in the entire livestock sector) or the consumers' attitude towards the supply chain (e.g. noncompliance in the baby food sector could have heavy consequences on consumers' trust). Our impression based on the data and information collected from the five European control bodies is that the organic certification system is still not prepared to implement risk-based controls completely. The crucial role played by control bodies in the certification system could benefit from a more explicit definition of risk at the regulatory level, which could lead to more effective and uniform procedures of risk evaluation.

References

- Albersmeier, F., Schulze, H., Jahn, G., Spiller A., 2009. The reliability of third-party certification in the food chain: from checklists to risk-oriented auditing. *Food Contr.* **20**, 927-935.
- Cooper, G. F., Herskovitz, E., 1992. A Bayesian method for the induction of probabilistic networks from data. *Machine Learn.* **9**, 309-347.
- Commission of the European Communities, 1999. Decision of 24th November 1999. Brussels: Commission of the European Communities. Accessed June 2013, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2000D0115:20070101:EN:PDF>.
- Commission of the European Communities, 2004. Communication from the Commission to the Council and the European Parliament. European Action Plan for Organic Food and Farming, COM(2004)415 final. Brussels: Commission of the European Communities. Accessed June 2013, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2004:0415:FIN:EN:PDF>.
- Cowell, R. G., Dawid, A. P., Laureitzen, S.L., Spiegelhalter, D. J., 2007. *Probabilistic Networks and Expert Systems*. Springer, New York.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Series B (Methodological)* **39**, 1-38.
- European Council, 2002. Regulation (EC) No 178/2002 of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Accessed June 2013, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002R0178:EN:PDF>.
- European Council, 2004, Regulation (EC) No 882/2004 of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules. Accessed June 2013, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:165:0001:0141:EN:PDF>.
- European Council, 2007. Regulation (EC) No 834/2007 of 28 June 2007 on organic production and labelling of organic products and repealing Regulation (EEC) No 2092/91. Accessed June 2013, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:189:0001:0023:EN:PDF>.
- European Commission, 2007. Report on the supervision of inspection bodies and authorities of the Member States according to Article 15 of council Regulation (EEC) No 2092/91 on organic production. Accessed June 2013, available at: http://ec.europa.eu/agriculture/organic/files/consumer-confidence/inspection-certification/EU-25_report_en.pdf.

- Gambelli D., Bruschi V., 2010. A Bayesian network to predict the probability of organic farms' exit from the sector: a case study from Marche, Italy. *Comput. Electron. Agricult.* **71**, 22-31.
- Gambelli D., Solfanelli F., Zanolini R., 2011. Un sistema di certificazione risk-based per i controlli in agricoltura biologica: un'applicazione tramite Bayesian networks. *Economia Agro-Alimentare* **3**, 37-56.
- Gambelli D., Solfanelli F., Zanolini R., 2012. Measuring the risk of non-compliance in Italian organic farms with parametric and non-parametric models. *New Medit*, **4**, 35-38.
- Hanley J., McNeil B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.
- Horvitz, E. J., Breese, J. S., Henrion, M., 1988. Decision theory in expert systems and artificial intelligence. *Intl. J. Approx. Reason.* **2**, 247-302.
- Howard, R. A., Matheson, J. E., 1981. *Influence Diagrams*. In: R.A. Howard & J.E. Matheson, eds., *Readings in Decision Analysis*. Strategic Decisions Group, Menlo Park, Ca.
- Jensen, F. V., 1996. *An Introduction to Bayesian Networks*. University College London Press, London.
- Jensen, F. V., Chamberlain B., Nordahl T., Jensen F., 1990. Analysis in Hugin of Data Conflict. In Bonissone, P.P., Henrion, M., Kanal, L.N., & Lemmer, J.F., eds., *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*. Elsevier, Cambridge, Ma.
- Kalisch, M., Beuhlmann, P., 2007. Estimating high dimensional directed acyclic graphs with the PC-algorithm. *J. Mach. Learn. Res.* **8**, 613-636.
- Korb, K. B., Nicholson, A. E., 2004. *Bayesian Artificial Intelligence*. Chapman and Hall, London.
- Kristensen, K., Rasmussen, I. A., 2002. The use of a Bayesian network in the design of a decision support system for growing malting barley without use of pesticides. *Comput. Electron. Agricult.* **33**, 197-217.
- Lauritzen, S. L., 1995. The EM Algorithm for Graphical Association Models with Missing Data, *Computational Statistics & Data Analysis*, **19**, 191-201.
- Lauritzen, S. L., Spiegelhalter, D. J., 1988. Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc., Series B*, **50**, 157-224.
- Pearl, J., 1986. Fusion, propagation and structuring in belief networks. *Artific. Intell.* **29**, 241-288.
- Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, Ca.
- Russell, S., Norvig, P., 2011. *Artificial Intelligence: a Modern Approach*, 3rd edition. Prentice Hall, Engelwood Cliffs, NJ.

- SINCERT, 2009. Directives for accreditation of control bodies issuing declarations of conformity of agricultural processes and products and food-stuffs produced by organic agriculture methods according to EC Regulations n. 834/2007 and following integrations and modifications. Accessed June 2013, available at: http://www.accredia.it/UploadDocs/100_1234RT_16rev_02.pdf.
- Sprites, P., Meek, C., 1995. *Learning Bayesian Networks with Discrete Variables from Data*. In: *Proceedings of First International Conference on Knowledge Discovery and Data Mining*, 20-21 August 1995. Morgan Kaufmann, Montreal.
- Sprites, P., Glymour, C., Scheines, R., 2000. *Causation, Prediction and Search*. 2nd ed., Cambridge (MA): MIT Press.
- Swets, J. A., 1998. Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.
- Van Asselt, E. D., Sterrenburg, P., Noordam, M.Y., Van der Fels-Klerx, H. J., 2012. Overview of available methods for risk based control within the European Union. *Trends Food Sci Tech*, **23**, 51-58.
- Zanoli R., Gambelli D., Bruschi V., 2012. Analysis of non-compliances in the organic certification system in Turkey. *New Medit*, **4**, 74-77.
- Zorn, A., Lippert, C., Dabbert, S., 2012. Supervising a system of approved private control bodies for certification: the case of organic farming in Germany. *Food Contr.* **25**, 525-532.
- Zorn, A., Lippert, C., Dabbert, S., 2013. An analysis of the risks of non-compliance with the European organic standard: a categorical analysis of farm data from a German control body. *Food Contr.* **30**, 692-699.

Table 1. Number of sanctions per category of noncompliance.

Type of noncompliance	Denmark	Germany	Italy	Switzerland
Slight noncompliance	7	4	2	9
Severe noncompliance	5	3	3	7

Table 2. Distribution of the farms by category of noncompliance.

Country	Year	Farms with slight noncompliance		Farms with severe noncompliance		Total farms N°
		N°	%	N°	%	
Denmark	2007	153	5.9	20	0.8	2,118
	2008	158	6.0	26	1.0	2,118
	2009	44	1.8	13	0.5	2,118
Germany	2007	722	50.1	32	2.2	1,442
	2008	676	46.9	22	1.5	1,442
	2009	626	43.4	14	1.0	1,442
Italy	2007	512	8.0	62	1.0	6,420
	2008	435	6.8	132	2.1	6,420
	2009	358	5.6	141	2.2	6,420
Switzerland	2007	63	1.5	168	4.0	4,210
	2008	51	1.2	124	2.9	4,210
	2009	74	1.8	111	2.6	4,210
United Kingdom	2007	567	35.7	11	0.7	1,589
	2008	678	42.7	8	0.5	1,589
	2009	630	39.7	38	2.4	1,589

Table 3. Descriptive statistics for structural and managerial potential risk factors: relative frequencies (means and standard deviation for non-dichotomous variables), 2008-2009 average values* (unless differently specified).

	Denmark	Germany	Italy	Switzerland	United Kingdom
	<i>mean / relative freq.</i>	<i>mean / relative freq.</i>	<i>mean / relative freq.</i>	<i>mean / relative freq.</i>	<i>mean / relative freq.</i>
Observations per year	2,118	1,442	6,420	4,210	1,589
Structural risk factors					
Conventional UAA	0.04**	0.01	0.10	0.00	0.23
Conversion UAA	0.23**	0.35	0.15	not available	0.17
	68.56	43.39	42.59	19.64	149.54
Farm size - UAA (ha)	(s.d.: 103.76)	(s.d.: 79.88)	(s.d.: 141.58)	(s.d.: 15.03)	(s.d.: 447.29)
Farmer's Experience (years)	not available	8.89 (s.d.: 3.57)	7.66 (s.d.: 3.79)	13.15 (s.d.: 6.65)	7.75 (s.d.: 2.40)
	84.57		8.23	24.94	73.24
Herd size (livestock units)	(s.d.: 345.22)	not available	(s.d.: 44.93)	(s.d.: 21.28)	(s.d.: 167.28)
Licensee	1.00	not available	0.42	not available	not available
Other certification schemes	0.00	0.85	0.05	1.00	1.00
Processing activity (nr of processed products)	0.02 (s.d.: 0.15)	0.12 (s.d.: 0.50)	0.18 (s.d.: 0.70)	0.79 (s.d.: 1.32)	0.10 (s.d.: 0.40)
Managerial risk factors					
<i>Crop production</i>					
Cereals	0.75	0.55	0.34	0.20	0.41
Citrus	0.00	0.00	0.08	0.00	0.00
Dried pulses	0.01	0.13	0.03	0.02	0.09
Fresh vegetables	0.19	0.20	0.25	0.05	0.26
Fruit	0.09	0.06	0.36	0.08	0.15
GMO-risk crops	0.01	0.06	0.03	0.02	0.01
Grapes	0.01	0.01	0.29	0.03	0.01
Green fodder	0.85	0.51	0.29	0.96	0.56
Industrial crops	0.04	0.12	0.13	0.06	0.07
	3.17	2.97	3.10	3.06	3.18
Number of crop types	(s.d.: 1.34)	(s.d.: 1.81)	(s.d.: 1.78)	(s.d.: 1.41)	(s.d.: 1.77)
Olives	0.00	0.00	0.60	0.00	0.00
Other arable crops	0.00	0.02	0.02	0.11	0.01
Permanent grassland	0.66	0.88	0.29	0.98	0.84
Root crops	0.09	0.16	0.05	0.10	0.10
<i>Livestock production</i>					
Bovines	0.46	0.52	0.09	0.79	0.52
Goats	0.03	0.05**	0.02	0.21	0.01
Pigs	0.07	0.13	0.03	0.11	0.08
Poultry	0.13	0.18	0.01	0.33	0.16
Sheep	0.14	0.08	0.06	0.28	0.35
Livestock density (LU/ha>2)	0.11	not available	0.01	0.30	0.06
	1.02	1.05	0.25	2.68	1.16
Number of species	(s.d.: 0.98)	(s.d.: 0.98)	(s.d.: 0.71)	(s.d.: 1.43)	(s.d.: 1.15)

*We use 2008-2009 average values for a matter of synthesis unless differently specified. The difference between the two years is however particularly low due to the balanced nature of the panel.

** 2008 only.

s.d.: standard deviation.

Table 4. Log-likelihood (LL) and average areas under the ROC curves (AUC) for the different models.

	Denmark		Germany		Italy		Switzerland		United Kingdom	
	Slight NC	Severe NC	Slight NC	Severe NC	Slight NC	Severe NC	Slight NC	Severe NC	Slight NC	Severe NC
AUC (2008)	0.71	0.79	0.69	0.83	0.68	0.78	0.74	0.62	0.71	0.82
AUC (2009)	0.78	0.88	0.72	0.78	0.58	0.70	0.69	0.61	0.79	0.81
LL(2008)	-17,360.84		-14,124.89		-67,455.43		-38,610.05		-16,662.93	
LL(2009)	-15,113.33		-13,230.99		-67,574.69		-39,677.14		-16,522.92	

NC: noncompliance

Table 5. Impact on probability of noncompliance (NC) by risk factor (2008-2009).

	Denmark		Germany		Italy		Switzerland		United Kingdom	
	Slight NCs	Severe NCs	Slight NCs	Severe NCs	Slight NCs	Severe NCs	Slight NCs	Severe NCs	Slight NCs	Severe NCs
General risk factors										
Co-dependence of NCs	++	++	++	++	++	++	++	++	++	++
Path dependence of NCs	++	+	+	+	++	++	++	++	+	++
Structural risk factors										
Conventional UAA	*	*		+			n.a.	n.a.		
Conversion UAA	++	++	+	+	+	++	n.a.	n.a.		
Farm size	++	++	+	+	++	+			+	+
Farmer's Experience	n.a.	n.a.					+			
Herd size	++	++	n.a.	n.a.	+	+	++		+	++
Licensee	n.a.	n.a.	n.a.	n.a.		+	n.a.	n.a.	n.a.	n.a.
Other certification schemes	n.a.	n.a.			+		n.a.	n.a.	n.a.	n.a.
Processing activity			+	+	+	+	++			
Managerial risk factors										
<i>Crop production</i>										
Cereals	+				+	++	+		+	
Citrus	n.a.	n.a.	n.a.	n.a.		-	n.a.	n.a.	n.a.	n.a.
Dried pulses					+	++			+	
Fresh vegetables						+	++			+
Fruit	-	-					++	+		++
GMO-risk crops				+		++	+		n.a.	n.a.
Grapes	n.a.	n.a.	n.a.	n.a.		+	++		n.a.	n.a.
Green fodder			+		+	+			+	
Industrial crops						++	+			+
Number of crop types	+		+	+	+	++	+		+	++
Olives	n.a.	n.a.	n.a.	n.a.			n.a.	n.a.	n.a.	n.a.
Other arable crops	n.a.	n.a.	+			++	++		n.a.	n.a.
Permanent grassland					+	+				
Root crops				+		++	+		+	+
<i>Livestock production</i>										
Bovines	++	+			+	+			+	+
Goats	++	++	*	++			+	++	n.a.	n.a.
Pigs	+	+	+	+			++	++	+	+
Poultry	+	+	+	++		+	+		+	+
Sheep	-	-	+	+			+		+	+
Livestock density	++	++	n.a.	n.a.					+	++
Number of species	+	+	+	++	+	+	++	+	+	+

n.a.: not applicable (data are not available, not discretizable, or extremely sparse);

*: 2008 only

+: impact on noncompliance probability between 10% and 50% (largest score either in 2008 and/or in 2009);

++: impact on noncompliance probability >50% (largest score either in 2008 and/or in 2009);

-: impact on noncompliance probability between -10% and -50% (largest negative score either in 2008 and/or in 2009).

Table 6. Farm types associated with a high probability of noncompliance (NC) in the year 2009.

	Denmark		Germany		Italy		Switzerland		United Kingdom	
	Slight NC	Severe NC	Slight NC	Severe NC	Slight NC	Severe NC	Slight NC	Severe NC	Slight NC	Severe NC
Initial probability	2.1%	0.6%	43.7%	0.9%	5.6%	2.2%	2.2%	2.7%	39.1%	2.4%
Farm Type 1	bovines; herd size: large; farm size: intermediate NC in the past		Cereals; permanent grassland; farm size: large; non organic land; NC in the past		cereals; green fodder; farm size: large; NC in the past		pigs; NC in the past		poultry; farm size: intermediate; NC in the past	
Final probability	8.5%	1.5%	71.9%	1.5%	12.4%	4.7%	6.1%	7.3%	53.2%	3.7%
Final probs. if slight NC in 2009 occurs	-	7.1%	-	2.0%	-	8.1%	-	22.4%	-	6.0%
Conflict score	-5		-2		-2		-1		0	
Sample size	61		70		75		55		40	
Farm Type 2	poultry; NC in the past		pigs; cereals; farm size: intermediate; NC in the past		bovines; green fodder; herd size: large; NC in the past		cereals; root crops; NC in the past		bovines; green fodder; farm size: large; herd size: large; green fodder; NC in the past	
Final probs.	6.1%	1.4%	64.5	1.4	12.4%	3.0%	3.1%	6.9%	59.3%	4.0%
Final probs. if slight NC occurs in 2009	-	7.1%	-	2.0 %	-	5.2%	-	18.6%	-	6.0%
Conflict	-3		-2		-4		-2		-3	
Sample size	40		37		42		26		77	

Figure 1. Example of directed acyclic graph

