Measuring the risk of non-compliance in Italian organic farms with parametric and non parametric models

D. GAMBELLI[1], *F. SOLFANELLI*[1], *R. ZANOLI*[1]

Abstract

Certification is an essential feature for organic products, and is based on controls/inspections to verify compliance with EU Reg. CE 834/07. Implementing a risk based inspection system can contribute to an effective and efficient organic control system, hence improving organic farming relative competitiveness. This study is part of the CERCOST EU project, and analyses data from inspections on a sample of Italian organic farms, aiming at the identification of management and structural risk factors contributing to the occurrence of non-compliances. Risk analysis is based on the estimation of both parametric and non-parametric models using negative binomial panel estimators and Bayesian network modelling.

Keywords: *non-compliance, risk-based certification, Bayesian network, count data models.*

Introduction

The main objective of an organic certification systems is to assure that organic products are compliant with the organic standards. Certification is a key element of organic farming systems today, because only certified organic products may be labelled as such, thereby gaining access to the organic market and earning premium prices (Dabbert et al., 2008). The certification system introduce transaction costs that reduce relative competitiveness of organic farming. A more efficient certification system may significantly contribute to increase organic product competitiveness while maintaining the benefits of trustable organic labelling. Article 27(3) of the EU Reg. 834/07 provides general guidelines for control visits and inspections, which should be based on a risk evaluation of non-compliance occurrence. The need for a risk-based control system is also indicated as key action in the European Action Plan for organic food and farming. While the potentials for a risk based inspection system for organic certification is recently being discussed by stakeholders (see Zanoli, 2010; Padel, 2010), few studies have considered the implementation of this approach from an empirical perspective. Gambelli et al. (2011) provide an implementation of risk based inspection system with a statistical approach to Italian data. This paper can be considered a continuation of these studies, and exploit the availability of a dataset for Italian organic farm with a wider time-span. This allows the implementation of a new methodological approach and the evaluation of results over a multiple-year period.

Other studies concerning the analysis of the organic certification system in Italy are Gambelli and Solfanelli (2009), and De Gennaro and Roselli (2008). The first developed a probabilistic model for the evaluation of non-compliance risk factors for organic farmers, the latter propose an efficiency and effectiveness analysis of the certification system in an Italian region. A risk based inspection system could support the control bodies particularly in the definition of the unannounced visits scheme, hence contributing to a more cost-effective system. The study reported here investigates the factors that affect the probability of non-compliances occurrence in a three-year sample (2007-2009) of Italian organic producers. A general description of the sanction distribution is provided, and the main outcomes of the statistical and probabilistic analyses are discussed, and compared with the indications from the technical report from ACCREDIA (2009) concerning the guidelines that control bodies follow for the present definition of operators' risk class.

Material and Methods

Data are obtained from a sample of farms certified by ICEA[1] in the period 2007-2009. The dataset represents about 20% of the Italian organic farms with a rather homogeneous geographical distribution over the country. The data contains information on farm structural-managerial characteristics and sanctions imposed on the farms according to results from control visits. Since data about non-compliance were not completely available, we have used the number of sanctions imposed on an operator as a proxy. In order to simplify our analysis, we have aggregated sanctions into two categories: slight and severe (Table 1), according to the definition of the different sanction types in ACCREDIA (2009). We assumed that a severe sanction is issued when "severe non-compliance" is detected, and a less severe sanction is issued when a "slight non-compliance" is

[1] Dipartimento di Scienze Agrarie, Alimentari e Ambientali – Università Politecnica delle Marche, Ancona-Italy, E-mail: d.gambelli@univpm.it. This paper is the outcome of a joint research effort. However, D. Gambelli wrote the 'Results' section, F. Solfanelli the 'Materials and methods' one, while R. Zanoli wrote the 'Discussion' section. Introduction and Conclusions are shared.

[2] ICEA (Ethical and Environmental Certification Institute) is among the oldest Italian certification bodies and has the largest share of inspected farms.

Table 1 - *Frequencies of sanctions by type and year.*

Nr of sanctions	Slight Sanctions 2007	2008	2009	Severe Sanctions 2007	2008	2009
0	8,024	8,082	7,778	8,665	8,436	8,034
1	494	407	302	67	142	142
2	216	116	113	26	46	24
> = 3	29	24	15	5	5	8
Total farms	8,763	8,629	8,208	8,763	8,629	8,208
Total sanctions	1,032	724	579	136	250	215
(%)	11.78	8.39	7.05	1.55	2.90	2.62

Source: own elaborations on ICEA data.

detected. In other words, we assumed that non-compliance is followed by sanctions with the appropriate level of severity.

In this study we want to estimate the risk of non-compliance (considered as probability of a non-compliance to happen) as dependent from a set of potential risk factors. We have no information concerning the reason why a non-compliance occurred, so we can only infer the risk from a list of potentially relevant risk factors related to the farmers. Risk factors can be classified into three categories:

- general risk factors: non-compliance co-dependence (the risk of slight non-compliance is affected by the occurrence of severe non-compliance, and vice versa) and non-compliance path-dependence (the risk of non-compliance is affected by the occurrence of non-compliance in the past); number of other certification schemes in the farm; farmer's experience as organic (years).
- structural/managerial risk factors: presence of non organic land, presence of licensee for direct marketing, farm size (ha), herd size (livestock units) farm and livestock complexity[2]; processing activities in addition to the ordinary farming activities.
- crop/livestock specific risk factor: thirteen crop types have been considered in the model (dummy variables): arable crops (*cereals, industrial crops, dry pulses, root crops*), fodder crops (*grasslands, green fodder*), permanent crops (*olives, grapes, fruits, citrus*), *vegetables* and *unutilized land*. For what concerns livestock, five types have been considered (dummy variables): *bovine, pigs, sheep, goats* and *poultry*.

For the analysis we have used two approaches: a parametric approach based on econometric count data models, and a probabilistic network approach based on Bayesian networks. The econometric approach provides an evaluation of the statistical significance of risk factors for the non-compliance occurrence, at the cost of some distributional and functional form assumptions; the probabilistic network approach is more flexible, and can handle inference under uncertainty, but does not provide specific testing for risk factor relevance. The parametric approach uses a count panel data model based on negative binomial distribution (see Cameron et al., 1998). The model explains the number of non-compliance at the farm level as a function of a set of risk factors. The panel model specification allows managing the individual heterogeneity due to individual effects. In our case these could refer to latent individual effects, due to non-measurable variables and to potential under reporting of non-compliances. We have used a random effect negative binomial model (see Greene, 2007 and Greene, 2008 for more details on the model).

The Bayesian network approach (see among others Horvitz et al., 1988) builds up a network of connections among variables, and the links among variables are measured in terms of conditional probabilities. Bayesian networks can be used to make inference concerning the conditional probability of non-compliance given a set of risk factors. For instance, we can infer the probability of getting an infringement if a farmer cultivates a specific crop, or runs a large farm, etc. The networks have been "learned" from data, for 2008 and 2009, and data on non-compliance for 2007 have been used to model the path-dependence of non-compliance through the three years time span. We have used PC algorithm for defining the structure of the network (see Spirties et al., 2000; Spirties and Meek, 1995), and the expectation-maximisation procedure for the computation of the conditional probability estimation (Dempster et al., 1977).

Results

Results for slight and severe non-compliance are reported separately in Table 2, together with the indication of the main risk factors considered in the technical guidelines-RT16 for Italian organic certification bodies (ACCREDIA, 2009). Relevant risk factors increasing the risk of non-compliance are shown with a +, while those decreasing the risk are shown with a –; blanks indicate no relevant impact on risk. When a risk factor is found as relevant in the econometric model only, it is marked with (EC), while if it is found relevant in the Bayesian network model only models, it is marked with (BN).

Relevant risk factors in the econometric model are those with coefficients above the 95% significance level. Relevant risk factors in the Bayesian networks are those showing an impact of at least 10% in absolute terms on the risk of non-compliance, either in 2008 or in 2009. We have considered in the RT16 column an adaptation of the risk factors taken into consideration by control bodies. Though a direct comparison is not possible as there is no bi-univocal correspondence between the risk factors we included in the analysis and those considered in the RT16, this information is useful to benchmark our findings. Diagnostics for the negative binomial model have referred to: overdispersion test (see among others Winkelmann and Zimmermann, 1995; Ismail and Jemain, 2011) which showed the preference for the negative binomial specification over the Pois-

[2] A Shannon index (see Shannon, 1948) was used in the parametric model, and the number of crop/livestock types in the non-parametric model.

Table 2 - R*elevant risk factors for non-compliance (NC) results from negative binomial and Bayesian networks models: a comparison with risk factors considered by control bodies (RT 16).*

	Slight NC	Severe NC	RT 16
General Risk factors			
Co-dependence of NC	+	+	+[b]
Path dependence of NC (BN model only)	+(BN)	+(BN)	+[b]
Number of other certification schemes	+(BN)		n.a.
Structural/managerial risk factors			
Farmer is a licensee		+(BN)	+
Farm size (ha)	+	+	+
Non organic land	-(EC)	+	n.a.
Farmers Experience as organic			n.a.
Processing activities	+	+(BN)	
GMO-risk crops	+/-[a]	+(BN)	+
Complexity of crops production	+	+	n.a.
Complexity of livestock production	+(BN)		+
Livestock units (LU) (BN models only)	+(BN)	+(BN)	+
Livestock density: LU/ha > 2 (BN models only)			n.a.
Crop/livestock specific risk factors			
Cereals	+(BN)	+	-
Industrial crops		+	n.a.
Dried pulses	+(BN)	+(BN)	n.a.
Root crops		+(BN)	n.a.
Grassland	+	+	n.a.
Green fodder	+	+	-
Unutilised land		+(EC)	n.a.
Other arable crops		+(BN)	n.a.
Vegetables		+(EC)	
Fruits	-(EC)	-(EC)	+
Olives	+(EC)	-	
Grapes	+(EC)	+(EC)	+
Citrus	-(BN)	-(BN)	
Bovine	+	+(BN)	+[c]
Goats	+(BN)		+[c]
Pigs	+(BN)	+(BN)	+[c]
Poultry	+(BN)	+(EC)	+[c]
Sheep	+		+[c]

[a] controversial results between EC and BN models; b only for severe NC; c livestock in general; n.a. not applicable.

son alternative specification; Hausman test (Hausman, 1978) for the random effects specification, which supported the choice for the random effect model over the fixed effect one.

Discussion and Conclusion

The main results arising from our analysis are that the *path-dependence* and *co-dependence of non-compliances* play a crucial role in the risk evaluation, though *path-dependence* could be not confirmed in the econometric model due to the impossibility to use a dynamic specification. As we have no information at the farmer level for what concerns personal aspects that could lead to non-compliant behaviours, *path-dependence* and *co-dependence of non-compliance* can be considered as a general proxy for the personal attitude of farmers to fraud. *Farm size* and *complexity of crops production* are also emerging as factors increasing the risk of non-compliances in both type of non-compliance and model. Finally, livestock production in general, though with results not confirmed by both models, can be considered as critical risk factor, which is indirectly confirmed by the results concerning fodder and grassland crops, usually related to livestock production. Few factors are found as reducing risk, and they are only referring to Mediterranean crops. These results are consistent with those found in Gambelli et al. (2011), which were referring to the same sample of farms but for only one year, and with those found in Gambelli and Solfanelli (2009) that were computed on a different farms sample. They are also generally consistent with the risk factors indicated in the RT16, particularly for what concerns the occurrence of severe non-compliance, *farm size*, and livestock production in general. Some differences however emerge for what concerns *cereals*, *green fodder*, and *fruit*. For the first two cases the RT16 assumes a negative impact on risk, which is not confirmed by our results, particularly for what concerns *green fodder*. Contrasting results for *fruit* between our findings and the indications in the RT16 are somehow less crisp as they are arising from one of the two models only.

The findings emerging from this study are consistent with those available from the (limited) literature on this issue, and provide an empirical verification of the risk factors considered by Italian control bodies. The results show the need to explore further the issue of non-compliances occurrence, advocating more research in collecting adequate data. Data on structural and managerial aspects are available from control bodies in sufficient details, but they fail in providing sufficient information at farmers' level. More detailed information like personal crime records, market price and market channel, liabilities and debts, financial information could improve considerably the explanatory power of both statistical and probabilistic models, supporting risk-based inspection schemes. Also, the availability of data for additional years would allow for a dynamic specification of the models, and enforce the empirical base for results. Further work could also consider the application of zero-inflated models handle the large share of farms without non-compliance, particularly for the severe case.

Acknowledgment

This publication was generated as part of the CERTCOST Project, agreement no. 207727 (http://www.certcost.org), with financial support from the European Community under the 7th Framework Programme. The publication reflects the views of the author(s) and not those of the European Community, which is not to be held liable for any use that may be made of the information contained.

References

ACCREDIA (2009): Directives for accreditation of 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, Availableat:http://www.accredia.it/extsearch_documentazione.jsp?area=7&ID_LINK=57&page=13&IDCTX=168&id_context=168.

Cameron, A.C., Trivedi ,P.K. (1998): Regression Analysis of Count Data, Cambridge University Press, New York, 413 p.

Dabbert, S., Lippert, C., Schulz, T., Zorn, A. (2008): Economic Analysis of Certification Systems for Organic Food and Farming at EU level. Proceeding of the 16th IFOAM Organic World Congress, Modena-Italy, June 16-20 2008. Available at: http://orgprints.org/view/projects/confe conference.html.

De Gennaro, B., Roselli, L. (2008): Un'analisi del sistema di controllo e certificazione delle produzioni da agricoltura biologica. In: evoluzione della normativa sull'agricoltura biologica e implicazioni sul sistema dei controlli, Programma Interreg III Grecia – Italia 2000-2006. Ideaprint, Bari, 166, pp. 83-126.

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.

Gambelli, D., Solfanelli, F. (2009): Una Rete Bayesiana per migliorare l'efficienza della fase di ispezione del processo di certificazione biologica. In Crescimanno, M., Schifani, G. (editors): Agricoltura Biologica: sistemi produttivi e modelli di commercializzazione e di consumo. Atti del IV Workshop GRAB-IT, Palermo, 26-27 Ottobre 2009. Università degli Studi di Palermo – Dipartimento di Economia dei Sistemi Agro-Forestali, Palermo, 405, pp. 91-96.

Gambelli, D., Solfanelli, F., Zanoli, R. (2011): Non compliance in organic certification: determinants for Italy. In Neuhoff, D., Halberg, N., Rasmussen, I.A., Hermansen, J., Ssekyewa, C., Sohn, SM. (editors): Organic is Life – Knowledge for Tomorrow. Proceedings of the third scientific Conference of ISOFAR, Seoul-Korea, 28 September - 01 October 2011. ISOFAR, Seoul-Korea, Vol. 2, 366 p., 112-115.

Greene, W.H. (2007): LIMDEP 9.0 Reference Guide. Econometric Software Inc., Plainview, N.Y.: E26-2 – E26-55.

Greene, W.H. (2008): Econometric Analysis, 6th editions. Prentice Hall, Upper Saddle River, N.J., 1178 p.

Hausman, J. A. (1978): Specification Tests in Econometrics, Econometrica, 46 (6): 1251-1271.

Horvitz, E.J., Breese, J.S., Henrion, M. (1988): Decision theory in expert systems and artificial intelligence, International Journal of Approximate Reasoning, 2: 247-302.

Ismail, N., Jemain, A.A. (2007): Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. Casualty Actuarial Society Forum - Winter 2007, 404 p., 103–158. Available at: http://www.casualtyactuaries.com/pubs/forum/07wforum/07w109.pdf.

Padel, S. (2010): The European Regulatory Framework and its Implementation in Influencing Organic Inspection and Certification Systems in the EU. Certcost project: deliverable D14/11. Available at: http://certcost.org/Lib/CERTCOST/Deliverable/D14_D11.pdf.

Shannon, C. E. (1948): A mathematical theory of communication, Bell System Technical Journal, 27: 379-423.

Spirties, 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. Montreal (Quebec Canada), Morgan Kaufmann, 348p., 294-299.

Spirties P., Glymour C., Scheines R. (2000): Causation, Prediction and Search. Cambridge (MA), MIT Press, 543p.

Winkelmann, R., Zimmermann, K. F. (1995): Recent Developments in Count Data Modeling: Theory and Applications, Journal of Economic, 9: 1-24.

Zanoli, R. (2010): Affidabilità ed efficacia del sistema nazionale di controllo e globalizzazione dei mercati. Atti del seminario "La certificazione dei prodotti biologici nel mercato globale", Bologna 25-26 marzo 2010. Available at: www.interbio.it/share/integra.../UNIVPM%20-%20Raffaele%20Zanoli.pdf.