INTERNATIONAL CONFERENCE OF AGRICULTURAL ECONOMISTS
ICAE
29th | Milan Italy 2015
UNIVERSITÀ DEGLI STUDI DI MILANO AUGUST 8 - 14
AGRICULTURE IN AN INTERCONNECTED WORLD

Validation of a non-parametric farm level crop choice simulation method

Louis Mahy, Bérénice E.T.I. Dupeux, and Jeroen Buysse
Ghent University, Belgium

Mahy et al. (2014) have developed a non-parametric methodology to predict land use choices of farmers in the context of the crop diversification measure. The methodology uses simulation at the micro level because crops cultivated by one farm cannot compensate for a lack of diversity of crops at another farm. A key difficulty of simulating at farm level is that the crop choice of an individual farmer is very difficult to predict because it depends on more factors than gross margins only, such as crop rotation, farmers experience, adaptability of machines and supply chain possibilities of the harvested products. The non-parametric methodology presented in this paper uses peer behaviour of farmers to identify choices of other farmers. This paper validates and compares the approach with improved versions on a regional case study in Flanders to show the possibilities and limitations of the methodology.

IAAE

1. Introduction

In response to the societal expectation for agriculture to contribute to environmental services (Van Huylenbroeck et al. 2007), the European institutions included a crop diversification requirement in the greening module of the latest Common Agricultural Policy (CAP) reform. To be compliant with this requirement and receive the associated payments, each year farmers need to have a minimum number of crops in given proportions on their arable land (European Parliament and Council of the European Union, 2013). This is an indirect way to stimulate crop rotation, which is in turn believed to be associated with biodiversity and general improvement of the resilience of soil and ecosystems (Westhoek, 2012) as well as reduced input demands for fertilizers and pesticides (Osman et al., 2015).

This measure confronts the agricultural economist community with a difficult problem: how to predict diversification behaviour. The lack of clear economic underpinning in a farmer's crop choice, defined as the self-selection problem by Paris (2001), makes diversification behaviour difficult to capture in economic models.

Since the crop diversification requirement operates at farm level, the impact assessment method should also operate at this level or should at least take the farm level heterogeneity into account. Among the existing farm-level models, there could be several candidates to solve the issue, although they show some problems. E.g. the symmetric positive equilibrium approach developed in Paris (2001) does not adequately reflect agent level heterogeneity (de Frahan et al., 2007, Rounsevell et al., 2011). Other models, such as typical farm models also are incompatible to model the specific diversification requirements since typical farms generally are more diverse than individual farms in reality (Louhichi et al., 2010).

Recently a new approach has been developed by Mahy et al. (2015). This model makes use of peer behaviour to predict reactions on the newly imposed diversification requirements of the CAP. In this non-parametric approach, the crop areas of the most resembling[1], complying peers are projected on the adapting agent's total farm area. Since the relative surfaces of the closest

[1] The resemblance is determined by the Euclidean distance of the crop areas.

peer comply with the diversification requirements, also formulated in relative terms, the projected crop configuration is by definition also compliant. Due to the use of the most resembling peer, the projected configuration deviates the least possible from the farm's original crop configuration. In this paper, this peer-based approach is submitted to a first ex-post validation and is further fine-tuned used past behaviour of the farm.

First, the validation process of the approach used in Mahy et al. (2015) is described. This is followed by a section with improvements on the approach and the comparison of both the basic and refined model. The final section draws the conclusions and indicates possible directions for future research.

2. Validation methodology and results

The most intuitive approach to validate the model described in Mahy et al. (2015) would be a comparison of the predicted and actual changes in crop areas caused by the crop diversification measure. However, at this point in time the post-reform data is not available yet. It would also be impossible to isolate the effect of the diversification measure from other influences. A feasible alternative is to look at general diversification behaviour and see whether the peer based model is able to predict the crop choice of voluntarily diversifying farmers.

Flemish land parcel data from 2013 (Agentschap voor Landbouw en Visserij, 2013) and 2014 (Agentschap voor Landbouw en Visserij, 2014) allow running such an ex-post validation. However, an increase of the number of crops from 2013 to 2014 at a specific farm is not motivated by regulation. In other words, it is voluntary. Nevertheless, in this paper we assume a similar peer-based predictive mechanism can be applied. The fundamental assumption of this peer-based approach is the following. A farmer optimizes his utility by making a decision he perceives as optimal. This decision is determined by many variables (monetary- and non-monetary variables, social- and psychological factors, etc.). We assume that a peer with a similar crop configuration shares on average more of these underlying decision-making variables with the diversifying farmer. Hence, a diversifying farmer should on average tend more in the direction of a close peer in the outcomes of his decision making process.

In Mahy et al. (2015) a minimum compliance with diversification requirements was assumed because of the external motivation. A farmer would change the least possible to the current crop configuration, since the latter is perceived as optimal. However, in case of voluntary diversification, the farmer might not per se pursue a minimum change to his crop configuration. Hence, we cannot assume a minimal change in his crop allocation. Therefore, the model described below is restricted to the prediction of the type of crop the diversifying farmer adopts. It leaves estimations of respective crop areas out of scope.

Similarity in terms of crop configurations is determined by the number of crops and crop areas. First, since we are talking about diversification behaviour, the peer needs to have at least one crop more than the diversifying farmer. On the other hand, since the average increase in crop numbers is 1.5, only farms with maximum 2 crops more than the diversifying farms are retained in the batch of possible peer farms. Second, within this subset of possible peer farms, the most resembling crop configuration is determined by the Euclidean distance of the individual farms' crop areas. The advantage of using crop areas is that it (in)directly covers three variables: which crops the farmers have in common, the area of those individual crops and the total farm area. Hence, the closest peer is determined by the following equations:

Minimize $$10^3\alpha_{n,peer} + \sum_c |\sigma_{c,n} - \sigma_{c,peer}| \quad (1)$$
s.t.
$$\sum_c \beta_{c,peer} \leq \sum_c \beta_{c,n} + 2 \rightarrow \alpha_{n,peer} = 1 \quad (2)$$

n farms have the possibility to grow c crops. Equation (1) identifies the closest peer for farm n, referred to as peer. Variables are represented by Greek symbols. σ is the crop allocation. α is a dummy variable with value 1 if the conditions in (2) regarding the number of crops are met, and where the presence of a crop on a farm is accounted for by β a dummy variable. The crop(s) $\beta_{c,peer}$ not present in the original crop configuration, composed by $\beta_{c,n}$, are then predicted as the crops the diversifying farmer would adopt. Applying these equations on the Flemish crop data of 2013 should, if the model is valid, allow predicting which crops the diversifying farms have adopted.

Three benchmarks could be used to evaluate the performance of this simplified version of the approach followed in Mahy et al. (2015). A first possible benchmark is the random probability of picking the right crop among the 146 crops in the model, which is 0.7%. However, not all crops are equally represented. A second benchmark takes into account that the presence of a rare crop does not have the same probability of being correct as predicting the adoption of a very common crop. Accounting for the distribution of the different crops on the farms can be done via the following formula:

$$\left(\sum_{c} \beta_{c,n} / \sum_{c,n} \beta_{c,n}\right)^2 = P_{bal} \qquad (3)$$

Where P_{bal} represents the probability of randomly choosing the right crop when accounted for the differences in distribution of the individual crops on the farms. This probability equals 12.7 %. A last possibility is always choosing the most common crop present at farm level. This way the probability increases to 20,9 %. Solely considering the farm-level, the last approach has the best 'random results'. However, at macro-level this approach reinforces the dominance of the most prevailing crop. Hence, from a macro perspective the second approach (P_{bal}) is more neutral since it maintains the status quo of the total aggregated areas of the individual crops. To account for agent-level heterogeneity and the interesting macro-level results, P_{bal} is used as benchmark in this paper[2].

Table 1 summarizes the results. On 35,460 farms, there are 5,486 farms with more crops in 2014 than 2013. We define the latter group as the adopting farms. From those adopting farms, the peer based model manages to predict a 100% correct crop configuration of 479 farms, or 8.8%. They are considered 100% correct if all crops present on the farm in 2014 coincide with the predicted crops, i.e. the crops present on the closest peer's farm.

Table 1 Results peer-based model – p.5

	Absolute number	% correct predicted
Farms	35,460	/
Adopting farms	5,486	8.8 %
Adopted crops in conservative adopting farms	5,277	22.4 %

[2] It should be noted that the calculated probability of 12.7% does not take into account the crops already present in the diversifying farmer's crop configurations. If the diversifying farms already have the most dominant crops in their crop configuration, diversification behavior involves less dominant or rare crops, which implies that P_{bal} is an overestimation.

One of the reasons why this estimate is so low is the general dynamic nature of annual crop configurations. Besides increasing the number of crops, there can also be changes among the crops present in the original crop configuration. Therefore, to isolate the prediction of the diversification aspect, one should look at conservative adopting farms. These are farms making no changes in the crops they already had, while adopting an additional crop. When considering only the conservative adopting farms, 22.4% of the predicted crops coincide with actually adopted crops. In other words, the peer-based model performs up to 1.8 times better than one could expect purely based on probability theory.

Besides farm and crop scale predictions there is also the macro-scale prediction. Even if crops are wrongly predicted at farm level, the predictions might still reflect correct tendencies on an aggregated level. The following formula shows whether the model allows capturing this:

$$1 - \left\{\sum_c \left(\left| \sum_n \beta_{c,n,predicted} - \sum_n \beta_{c,n,adopted} \right| / \left(2 * \sum_n \beta_{c,n,adopted}\right)\right)\right\} = P_{tendency}$$

Where $P_{tendency}$ is the proportion of crops correctly predicted at a fully aggregated level. By fully aggregated level is meant irrespective of the farm on which the adoption/ prediction occurs. The proportional difference between the number of times a certain crop is predicted to be adopted and the number of actual adoptions of this crop, is the error rate. Hence, $P_{tendency}$ is 1 minus the error rate. If this figure would be 0, none of the predictions would correspond to an adoption. If this figure would be 1, all individual crop predictions would coincide with an individual adopted crop. The latter would mean macro-level changes are fully captured. In case of the peer-based model, the result is 70 %. Hence, the number of times individual crop types are adopted is on average over- or underestimated by 30%.

3. Model improvement – past based

An alternative to using cross-sectional information to predict behaviour, is to use past behaviour. When a farmer had a crop in the past he did not have in the base year, we assume there is a high probability he will turn back to this crop when diversifying. Of course this is not diversification behaviour in the absolute sense, it is strongly related to the crop rotation applied on the farm. Nevertheless, Table 2 shows that the correct predictions increase using crop information from a crop in year X-1 crop if it is not present in the crop plan of year X. The past-based approach predicts 23% of the predicted crop configurations at farm-level fully correct. When considering the adopted crops in conservative adopting farms, 48.2% is correctly predicted.

Table 2 Results of the past based model

	Absolute number	% correct predicted
Farms	35,460	/
Adopting f with suitable past behaviour*	3,022	23 %
Adopted crops in conservative adopting farms	2,792	48.2 %

* Note this is only a subsample of the adopting farms.

Also on an aggregated level, again independent from farm level, 85% of the predicted crops match with an adopted crop (i.e. the $P_{tendency}$). Compared to a purely probability based approach, it allows to predict four times better the cropping plan at regional level. Hence, the farmer's own past would be an important improvement of the model of Mahy et al. (2015).

4. Model improvement – hierarchical

The past based model performs approximately two times as good as the peer based model. However, it is only applicable on a subset of the diversifying farms, those which have a crop in year X-1, not present in year X. This suggests that a hierarchical model, combining the past and the peer based model, would fit the overall data best. If there was a crop on the farm in previous year, not present in the base year, this information should serve as first choice in the prediction. If there is no such crop in year X-1, the peer based suggestion should serve as prediction. This approach can be formalised as follows:

$$(\beta_{x,c,peer} - \beta_{x,c,n}) * \varepsilon + \gamma_{x-1,c,n} + \beta_{x,c,n} = \beta_{x+1,c,n} \tag{1}$$

s.t.

$$\text{Minimize } (10^3 \alpha_{x,n,peer} + \textstyle\sum_c |\sigma_{x,c,n} - \sigma_{x,c,peer}|) \tag{2}$$

$$\textstyle\sum_c \beta_{x,c,peer} \leq \sum_c \beta_{x,c,n} + 2 \rightarrow \alpha_{x,n,peer} = 1 \tag{3}$$

$$\beta_{x,c,n} < \beta_{x-1,c,n} \rightarrow \gamma_{x-1,c,n} = 1 \tag{4}$$

$$\textstyle\sum_c \gamma_{x-1,c,n} = 0 \rightarrow \varepsilon = 1 \tag{5}$$

The hierarchical approach predicts the crop configuration of 17.8% of the adopting farms fully 100% correct, while 36% of the adopting crops in conservative adopting farms are correctly predicted. At macro-level 81.8% ($P_{tendency}$) of the predicted crops coincide with an adopted crop, which means for the individual crops, the predicted number of adoptions is on average an 18.2 % over- or underestimated.

Table 3 Results hierarchical model

	Absolute number	% correct predicted
Farms	35,460	/
Adopting farms	5,486	17.8 %
Adopted crops in conservative adopting farms	5,277	36.0 %

5. Conclusion

The model described in Mahy et al. (2015) has a probability of predicting the correct crop which is double as high as we could expect purely based on probability theory. However, the results also show much room for improvement. This improvement has been sought in the farmers' past behaviour. Such a past based model performs four times better than the probability based model and two times better than the peer based model.

Since not all farmers had crop configurations that allow predictions based on past based, a hierarchical, combining past and peer based model performs best if a prediction should be made for all farms. The past approach is used on farms with a crop present in the crop plan of year X-1 while non-existent in year X. The peer based approach is applied on farms without past diversification behaviour. This model achieves 36% accuracy at crop level and 22% at farm level. Macro-level crop plan decision are captured with 82% accuracy with the hierarchical model. The latter is calculated as the overlap between the number of times individual crop types are being adopted and predicted, independent from the farm on which they are adopted or predicted. These results are very intuitive because they confirm that a panel-data based approach is also preferred in non-parametric models as they are in parametric models. It also indicates that further improvements are possible using data with more time lags in the panel data based approach.

In addition, further validation of the model should provide insights on the validity of the model in case of imposed diversification behaviour and the accuracy of crop area predictions. Most interesting would be a comparison with other existing models. This includes the models using cropping plan decision mainly on regional economic parameters such as FFSIM (Louhichi et al., 2010) and the symmetric positive equilibrium approach (Paris, 2001) or others such as the machine learning approach developed by Osman et al. (2015).

7. References

Agentschap voor Landbouw en Visserij, 2013. Algemene landbouwgebruikspercelen ALV 2013. Agentschap voor Landbouw en Visserij, Vlaamse Overheid, Brussel.

Agentschap voor Landbouw en Visserij, 2014. Algemene landbouwgebruikspercelen ALV 2014. Agentschap voor Landbouw en Visserij, Vlaamse Overheid, Brussel.

De Frahan, B.H., Buysse, J., Polomé, P., Fernagaut, B., Harmignie, O., Lauwers, L., Van Huylenbroeck, G. and Van Meensel, J., 2007. Positive mathematical programming for agricultural and environmental policy analysis: review and practices. In Weintraub, A., Romero, C. Bjorndal, T and Epstein R. (ed.). Handbook of Research in Natural Resources. Springer Science, New York

European Parliament and Council of the European Union, 2013. Regulation (EU) No 1307/2013 of the European Parliament and of the Council of 17 December 2013 establishing rules for direct payments to farmers under support schemes within the framework of the common agricultural policy and repealing Council Regulation (EC) No 637/2008 and Council Regulation (EC) No 73/2009, O.J. L 347, 20.12.2013, p. 608–670.

Mahy, L., Dupeux B.E.T.I., Van Huylenbroeck, G., Buysse, J., 2015. Simulating farm level response to crop diversification policy. Land Use Pol., 46, 36-42

Louhichi, K., Kanellopoulos, A., Janssen, S., Flichman, G., Blanco, M., Hengsdijk, H., Heckelei, T., Berentsen, P., Oude Lansink, A., Van Ittersum, M., 2010. FSSIM, a bio-economic farm model for simulating the response of EU farming systems to agricultural and environmental policies. Agric. Syst. 103(8), 585–597

Osman, J., Inglada, J., Dejoux, J.-F., 2015. Assessment of a Markov logic model of crop rotations for early crop mapping. Computers and Electronics in Agric. 113, 234-243

Paris, Q., 2001. Symmetric Positive Equilibrium Problem: A Framework for the Rationalizing Economic Behaviour with Limited Information. Am. J. Agric. Econ. 83(4), 1049-1061

Rounsevell, M.D.A, Robinson, D.T. and Murray-Rust, D., 2011. From actors to agents in socio-ecological systems models. Phil. Trans. R. Soc. 367

Stoddard, F., Murphy-Bokern, D., 2012. Reform of the Common Agricultural Policy to improve the environment. A Statement from the 12th Congress of the European Society for Agronomy. European Society for Agronomy, France: Montpellier.

Van Huylenbroeck, G., Vandermeulen, V., Mettepenningen, E., Verspecht, A., 2007. Multifunctionality of Agriculture: A Review of Definitions, Evidence and Instruments. Living Rev. Landscape Res. 1, 3.

Westhoek, H., Van Zeijst, H., Witmer, M., van den Berg, M., Overmars, K., ven der Esch, S., van der Bilt, W., 2012. Greening the CAP An analysis of the effects of the European Commission‘s proposals for the Common Agricultural Policy 2014-2020. PBL Netherlands Environmental Assessment Agency.