Heckman two-stage spatial models for agricultural land use change

Davide Martinetti, Ghislain Geniaux

HECKMAN TWO-STAGE SPATIAL MODELS FOR AGRICULTURAL LAND USE CHANGE

Davide Martinetti[*±1] – *Ghislain Geniaux*[1]

[1] UR Ecodeveloppement, INRA PACA, Avignon, France
∗ Speaker
± Corresponding author: (e-mail address)

Introduction

Modelling agricultural land use and land use change is a remarkable challenge in terms of statistical analysis. Often, the scarcity of data hampers an exhaustive analysis of the subject, while, instead, its abundance complicates the estimation of a statistical model, since there exists no suitable technique than can cope with the complexity and the size of the problem, especially when spatial dependence and spatial autocorrelation are taken into account. This issue is strongly related to the spatial scale of the study region: large areas tend to have less refined scales, while smaller areas support finer ones and allow to reach the agent decisional scale (farms or plots) for which microdata are relevant. The Provence region in Southern France offers the unique opportunity to access fine scale data at farm level.

Materials and Methods

We considered a spatial database containing the agricultural choices of 1077 farms in 2007 and 2008 that cover all farms of Vaucluse County (Southern France). The cover choices are grouped into 28 categories, but we further cluster them into 9 larger sets representing the main agricultural productions of the study region: “cereals”, “oleaginous”, “industrial”, “vegetables and flowers”, “wine”, “olives”, “orchards”, “meadows” and “others”. Each farm is then described by a vector of shares of the corresponding nine types of productions (and the corresponding surfaces in hectares over the total farmed area), plus other variables representing information on the farm structure (number of employees and work load), the farm owner (age and land portfolio) and the farm physical characteristics (size, exposure, slope distance to cities and main roads). This type of models is usually referred as fractional or compositional models. The aim is to model agricultural land-use shares as a function of the observed characteristics of the farm.
The most common strategy for such problems is to use multicrop econometric models derived from a profit maximization problem where land is considered as an allocable fixed input (Fezzi and Bateman 2011) and/or Multinomial logit (Wu and Segerson (1995), Carpentier and Letort 2015). However, when technology production are considered, very detailed information for each farms is needed and it limits such analysis to small samples or survey data: such models are then a-spatial (do not consider spatial dependence) and can not be considered as proper land use models. On the other hand, land-use models including both agricultural covers and spatial dependence of agricultural land-use choices generally consider only one or two classes of crops, and are typically applied to grid data. Finally, farmers’ crop allocation within their farm can also be studied using fractional model (there exists a variety of them, see, amongst others the Fractional logit model (Papke and Wooldridge, 1993) and the Dirichelet model (Mullahy, 2010).

In this contribution we propose a statistical framework with spatial dependence for predicting agricultural choices with a large variety of agricultural covers (9 classes representing main production types). Delbecq and Florax (2010) and Ay et al (2014) consider spatial model of land use share by introducing spatial dependence in the logit-transformed aggregate models. Here, we propose a different strategy using a two-stage procedure, a.k.a. Heckman two-stage method or Heckit, in which spatial dependence can be introduced in both stages. Typically, Heckit models are used for correcting the bias originated by non-randomly selected samples: in the first stage, the selection bias is estimated by means of a probit model, that can be used to predict the probability of an observation to belong to the non-randomly selected sample. In the second stage, a transformation (Inverse Mills Ratio or IMR) of these predicted individual probabilities is added as explanatory variable to the final model.

In our proposal, we reinterpret the Heckman two-stage method to the case of compositional data, in order to accommodate for the presence of multiple dependent variables: in the first stage we estimate 9 probit models (one for each type of production) in order to predict the probability of observing the different types of productions in each of the 1077 farms. To improve the predictions of the first stage, we adopted a spatial version of the classic Probit model, i.e. the Spatial Autoregressive Probit model (SAR probit, see Martinetti and Geniaux, 2016), in which the probability of a certain type of production is also explained by its presence in neighbouring farms. In the second stage, we model the surface of each type of production, separately, by means of different continuous regression models: linear regression (OLS), linear regression with spatial dependence of the dependent variable (SAR), local linear model (GWR) and locally weighted SAR (GWRSAR, see Geniaux and Martinetti 2016), with the precaution of adding to each of the nine models the corresponding IMR, computed from the previous Spatial SAR probit estimation. It has to be noticed that these second stage models are fitted only on those observations in which the corresponding type of production has a strictly positive value.

Results and Discussion

Preliminary results of the Heckit framework (with OLS regression for the second stage) show that: in the first stage, a) Spatial probit models outperform classic probit in terms of right predictions of presence/absence of crops in farms b) The spatial autocorrelation parameters are always positive, but sometimes non-significant; In the second stage a) The selection bias corrections were significant for "cereal", "industrial", "wine", "orchards" and "others" productions.

In general: a) the list of variables that are significant for explaining the presence/absence of a certain types of production are almost never used to explain its corresponding share; this legitimize the use of a two stage method. In particular, the choice of allocating crops is better explained by geo-physical variables, while the share within the available arable land of the farm is better modelled by socio-economical variables. b) The sum of squares of the residuals on the test sample of our method is 6 to 8 times smaller if compared with classic fractional regression methods.

References

Ay J.S., Chakir, R., and Le Gallo, J. (2014). Individual vs. aggregate models of land use changes: Using spatial econometrics to improve predictive accuracy? *Working Paper*.

Carpentier, A. and Letort, E. (2014). Multicrop production models with multinomial logit acreage shares. *Environmental and Resource Economics*, 59(4):537–559.

Delbecq, B. A., Florax, R., et al. (2010). Farmland allocation along the rural- urban gradient: the impacts of urbanization and urban sprawl. In *Agricultural and Applied Economics Association 2010 Annual Meeting*, July, pages 25–27.

Fezzi, C. and Bateman, I. J. (2011). Structural agricultural land use modeling for spatial agro-environmental policy analysis. *American Journal of Agricultural Economics*, 93(4):1168– 1188.

Geniaux, G. and Martinetti, D. (2016). Spatial econometrics models un- der misspecification the case of spatially-varying covariates and spatial weight misspecification. In *15th International Workshop on Spatial Econometrics and Statistics*, Orleans, France.

Martinetti, D. & Geniaux, G. (2016). ProbitSpatial: Probit with spatial dependence, SAR and SEM models*. R package version 1.0*.

Murteira, J.M. & Ramalho, J.J. (2014). Regression analysis of multivariate fractional data. *Econometric Reviews*, **35**, 1-38.

Mullahy, J. (2010). Multivariate Fractional Regression Estimation of Econometric Share Models. *NBER Working Papers 16354, National Bureau of Economic Research, Inc.*

Papke, L. E. and Wooldridge, J. (1993). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates.

Wu J, Segerson K (1995) The impact of policies and land characteristics on potential groundwater pollution in Wisconsin. *Am J Agr Econ* 77(4):1033–1047