DOES AGGREGATION LEAD TO BIASED INFERENCES? AN EMPIRICAL ANALYSIS OF THE ADOPTION OF OIL-PUMPKIN CULTIVATION IN AUSTRIA AT THE FARM- AND MUNICIPALITY-LEVEL

Andreas Niedermayr and Jochen Kantelhardt

a.niedermayr@boku.ac.at

Institut für Agrar- und Forstökonomie, Universität für Bodenkultur Wien,
Feistmantelstraße 4, 1180 Wien, Österreich

2017

Posterpräsentation anlässlich der 57. Jahrestagung der GEWISOLA
(Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.)
und der 27. Jahrestagung der ÖGA
(Österreichische Gesellschaft für Agrarökonomie)
„Agrar- und Ernährungswirtschaft zwischen Ressourceneffizienz und gesellschaftlichen Erwartungen“
Weihenstephan, 13. bis 15. September 2017

DOES AGGREGATION LEAD TO BIASED INFERENCES? AN EMPIRICAL ANALYSIS OF THE ADOPTION OF OIL-PUMPKIN CULTIVATION IN AUSTRIA AT THE FARM- AND MUNICIPALITY-LEVEL

Andreas Niedermayr and Jochen Kantelhardt [1]

Summary

The aim of this study is to assess, whether estimation of the same innovation-adoption model at farm- and municipality-level results in an ecological fallacy, meaning that based on aggregated data, one would make inverse inferences about the driving forces influencing the adoption decision at the farm level. The adoption of an emerging alternative crop in Austria, the Styrian Oil Pumpkin, serves as an applied example. Our findings indicate the presence of an ecological fallacy. We therefore propose further research, which could consist of Monte Carlo simulations in order to analyse sensitivity of results with respect to the degree of aggregation.

Keywords

Innovation adoption, ecological fallacy, SLX-Tobit model, Styrian Oil Pumpkin PGI.

1 Introduction

Empirical innovation adoption studies are interested in estimating the effect of various driving forces on the adoption of innovations. As innovation adoption often occurs in spatial clusters, the notion of spatial spillover effects is of particular interest in this context. In order to analyse spatial spillover effects, spatially explicit data of the whole population of interest (e.g. farms) is needed. As complete census data at the farm level is hardly available and limited resources prevent large-scale surveys of the whole farm population, researchers mostly use aggregated data (e.g. GARRETT et al., 2013; NIEDERMAYR et al., 2016). However, this may result in an ecological fallacy (ANSELIN, 2002), meaning that the use of aggregated data to make inferences about a process happening at the farm level (the adoption decision) may lead to inverse inferences about the true relationship of interest. While limited research that compares the outcomes of such studies at different aggregation levels exists (e.g. SCHMIDTNER et al., 2015), we are not aware of any empirical analysis comparing aggregated- and farm-level results. The aim of this study is therefore to assess, whether aggregation could lead to an ecological fallacy. The adoption of oil-pumpkin cultivation in an Austrian case study region serves as an applied example.

2 Data and Methods

For the regression analysis, we use previously unavailable, spatially explicit cross-sectional data from 2010 of roughly 7,726 farms in a case-study region in Lower Austria (BMLFUW, 2016), where the implementation of a protected geographical indication for Styrian Pumpkin Seed Oil triggered a dynamic development of oil-pumpkin cultivation (NIEDERMAYR et al., 2016). Because of censoring in our dependent variable (share of arable land, cultivated with oil pumpkin), we estimate a Tobit model and further extend it to a Spatial Lag of X (SLX) Tobit model. In a SLX model, spatial lags of the independent variables (WX), reflecting for each observation the average value of neighbouring observations, are added as further independent variables (HALLECK VEGA and ELHORST, 2015). This allows estimating potential spatial spill-over effects of selected independent variables on adoption, which could in our case consist of

[1] Institut für Agrar- und Forstökonomie, Universität für Bodenkultur Wien, Feistmantelstraße 4, 1180 Wien. E-Mail: a.niedermayr@boku.ac.at, jochen.kantelhardt@boku.ac.at

local production- and marketing conditions for oil-pumpkin cultivation. The independent variables in our model describe natural conditions, availability of oil pumpkin specific infrastructure, production- marketing- and policy- related factors, social, temporal and spatial factors. We directly aggregate the farm-level data to the municipality-level, in order to rule out any other sources of influence on the results.

3 Preliminary results

Table 1 shows the partial effect at the average (PEA) of the independent variables. While most signs of the significant variables do not change, when comparing municipality- and farm-level results, there are also differences. We briefly illustrate the issue with the variable direct marketing, while noting that a similar line of argument is also possible for others (e.g. the spatial lag variables). Although, direct marketing is beneficial for oil-pumpkin cultivation from a theoretical point of view, the model based on municipality-level data shows a negative relationship. Most likely, at the municipality level, the presence of direct marketing farms, which do not cultivate oil pumpkin, leads to a bias of the true relationship of interest. Such potential ecological fallacies could also be present in comparable studies and are in our case overcome by an analysis at the farm-level. However, the scarce availability of spatially explicit farm-level data is not likely to change in the near future, ruling out this option as a general solution. We therefore propose further research, which could include Monte Carlo simulations in order to analyse the sensitivity of results with respect to the degree of aggregation.

Table 1: Comparison of marginal effects at the municipality- and farm level

Independent Variables	Municipality level		Farm level	
Soil-quality index	**0,04**	**n.s.**	**0,01**	*******
Distance to nearest drying facility for pumpkin seeds	-0,20	***	-0,10	***
Livestock density	-0,11	*	-0,01	**
Log(farm size)	-0,010	*	-0,003	**
Log(UBAG subsidy for arable land)	-0,000	n.s.	0,000	n.s.
Log(arable land)	**0,005**	**n.s.**	**0,006**	*******
Temporal lag of oil-pumpkin share	1,00	***	0,09	***
Direct marketing	**-0,06**	*****	**0,15**	*****
Organic farming	0,09	***	0,95	***
Agricultural education	0,03	**	0,08	*
WX of Direct marketing	-0,05	n.s.	-0,39	n.s.
WX of Organic farming	**0,003**	**n.s.**	**0,45**	******
WX of Agricultural education	**-0,01**	**n.s.**	**0,34**	*****

Source: own calculations, data from BMLFUW (2016). Note: ; ***, ** and * and denote significance at the 1%, 5% and 10% levels, n.s.=not significant

References

ANSELIN, L., 2002. Under the hood Issues in the specification and interpretation of spatial regression models. Agricultural Economics 27 (3), 247–267

BMLFUW, 2016. Integrated Administration and Control System Database.

GARRETT, R.D., LAMBIN, E.F., NAYLOR, R.L., 2013. Land institutions and supply chain configurations as determinants of soybean planted area and yields in Brazil. Land Use Policy 31, 385–396.

HALLECK VEGA, S., ELHORST, J.P., 2015. The SLX Model. Journal of Regional Science 55 (3), 339–363.

NIEDERMAYR, A., KAPFER, M., KANTELHARDT, J., 2016. Regional heterogeneity and spatial interdependence as determinants of the cultivation of an emerging alternative crop: The case of the Styrian Oil Pumpkin. Land Use Policy 58, 276–288.

SCHMIDTNER, E., LIPPERT, C., DABBERT, S., 2015. Does Spatial Dependence Depend on Spatial Resolution?: An Empirical Analysis of Organic Farming in Southern Germany. German Journal of Agricultural Economics 64 (3), 175–191.

Contact to author:

Andreas Niedermayr
E-Mail: a.niedermayr@boku.ac.at
Web: http://www.wiso.boku.ac.at/afo/

University of Natural Resources and Life Sciences, Vienna
Department of Economics and Social Sciences

57. Jahrestagung der GEWISOLA, 27. Jahrestagung der ÖGA

„Agrar- und Ernährungswirtschaft zwischen Ressourceneffizienz und gesellschaftlichen Erwartungen". Weihenstephan, 13. bis 15. September 2017

DOES AGGREGATION LEAD TO BIASED INFERENCES?

AN EMPIRICAL ANALYSIS OF THE ADOPTION OF OIL-PUMPKIN CULTIVATION IN AUSTRIA AT THE FARM- AND MUNICIPALITY LEVEL

Andreas Niedermayr* and Jochen Kantelhardt*

*Institute of Agricultural and Forestry Economics, University of Natural Resources and Life Sciences, Vienna

INTRODUCTION

- **Ecological fallacy:** Based on an analysis with aggregated data, one possibly makes inverse inferences about driving forces influencing a process happening at the individual level. (OPENSHAW 1984).
- **Aim of this study:** Assess, whether estimation of the same empirical innovation-adoption model at farm- and municipality-level results in an ecological fallacy.
- **Applied example:** Adoption of oil-pumpkin cultivation in an Austrian case-study region (see NIEDERMAYR et al. (2016) for more information).
- **Illustration** of an ecological fallacy in *Figure 1* (right panel): share of arable land cultivated with oil pumpkin varies with the presence of direct marketing on individual farms (difference between intercepts of dashed lines – **direct effect**), but also with the share of direct marketing farms in a municipality (slopes of dashed lines – **contextual effect**).

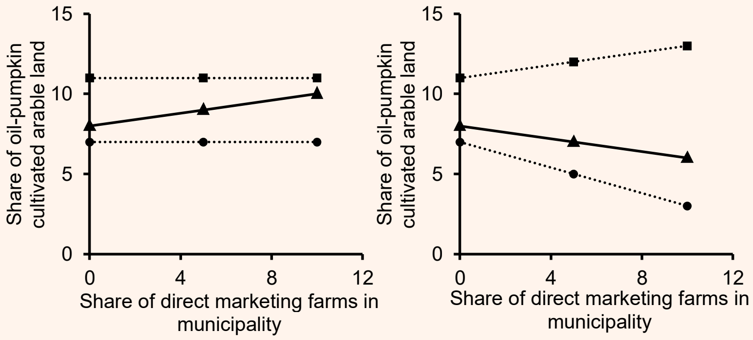

Figure 1: *Example for correct (left) and incorrect (right) inference from aggregated data. ▲ all farms (observed); ■ direct marketing farms (not observed); ● non direct-marketing farms (not observed). Source: own elaboration, modified from JARGOWSKY (2005).*

DATA AND METHOD

Cross sectional data (2010) of 7,726 farms in an Austrian case-study region (BMLFUW 2016)
► Empirical innovation adoption model

⬇

High proportion of zeroes in dependent variable
► Tobit model

⬇

Contextual effects of independent variables
► Spatial Lag of X (SLX) model
(HALLECK VEGA and ELHORST 2015)

⬇

Direct aggregation of data from farm- to municipality level
► Comparison of results

Figure 2: *Methodological approach. Source: own elaboration.*

- **Dependent variable:** share of oil-pumpkin cultivated arable land.
- **Independent variables** control for: natural conditions, proximity to oil-pumpkin specific infrastructure, production- marketing- and policy-related factors, social, temporal and spatial factors.
- Inclusion of **spatial lags of independent variables** (average value of neighbouring observations), allows a clear distinction between direct effects and contextual effects.

PRELIMINARY RESULTS

- **Direct marketing** beneficial for oil-pumpkin cultivation from a theoretical point of view.
- However, results show **negative relationship** at the municipality-level.
- A switch of the analysis to the **farm-level** possibly resolves this issue.
- Spatial lags of independent variables capture **spatial correlation**, but their effects need to be interpreted with care (contextual effect vs. omitted variable bias).

Table 1: *Regression results of SLX-Tobit model at municipality- and farm level*

Partial effect at the average (PEA) of independent Variables	**Municipality level**	**Farm level**
Soil-quality index	0,04	0,01 ***
Distance to nearest drying facility	-0,20 ***	-0,10 ***
Livestock density	-0,11 *	-0,01 **
Log(farm size)	-0,01 *	-0,003 **
Log(UBAG subsidy for arable land)	-0,00	0,00
Log(arable land)	0,01	0,01 ***
Temporal lag of oil-pumpkin share	1,00 ***	0,09 ***
Direct marketing	**-0,06 ***	**0,15 ***
Organic farming	0,09 ***	0,95 ***
Agricultural education	0,03 **	0,08 *
WX of Direct marketing	-0,05	-0,00
WX of Organic farming	**0,00**	**0,01 ****
WX of Agricultural education	**-0,01**	**0,003 ***

*Note: "Direct marketing", "Organic farming" and "Agricultural education" are dummies at the farm level and shares of farms at the municipality level; spatial-lag variables are denoted by the prefix "WX of"; ***, ** and * and denote significance at the 1%, 5% and 10% level. Source: own elaboration, data from BMLFUW (2016).*

OUTLOOK

- Potential ecological fallacies also in **comparable studies** possible.
- Analyses at the farm-level mostly **not feasible** (lack of data).
- Further research, based on e.g. **Monte Carlo simulations** in order to analyse the sensitivity of results with respect to the degree of aggregation (e.g. based on unified grids (SCHMIDTNER et al. 2015) with varying size).

REFERENCES

BMLFUW, 2016. IACS Database.

Halleck Vega, S., Elhorst, J.P., 2015. The SLX Model. J. Reg. Sci. 55 (3), 339–363.

Jargowsky, P.A., 2005. Ecological Fallacy. In: Kempf-Leonard, K. (Ed.) Encyclopedia of social measurement, vol. 1. Elsevier, Amsterdam, pp. 715-721.

Niedermayr, A., Kapfer, M., Kantelhardt, J., 2016. Regional heterogeneity and spatial interdependence as determinants of the cultivation of an emerging alternative crop: The case of the Styrian Oil Pumpkin. Land Use Policy 58, 276–288.

Openshaw, S., 1984. Ecological Fallacies and the Analysis of Areal Census Data. Env. and Plan. A 16 (1), 17–31.

Schmidtner, E., Lippert, C., Dabbert, S., 2015. Does Spatial Dependence Depend on Spatial Resolution?: An Empirical Analysis of Organic Farming in Southern Germany. GJAE 64 (3), 175–191.