USING MULTIPLE NEIGHBORING INTERACTION EFFECTS IN SPATIAL REGRESSION SPECIFICATIONS TO REDUCE OMITTED VARIABLE BIAS

Hugo Storm, Thomas Heckelei

Hugo.storm@ilr.uni-bonn.de

Institute for Food and Resource Economics (ILR), University of Bonn, Germany

2016

Paper prepared for presentation at the 56th annual conference of the GEWISOLA (German Association of Agricultural Economists)

„Agricultural and Food Economy: Regionally Connected and Globally Successful“

Bonn, Germany, September 28 – 30, 2016

USING MULTIPLE NEIGHBORING INTERACTION EFFECTS IN SPATIAL REGRESSION SPECIFICATIONS TO REDUCE OMITTED VARIABLE BIAS

Abstract[1]

A major challenge in the analysis of micro level spatial interaction is to distinguish actual interactions from the effects of spatially correlated omitted variables. We consider a spatially lagged explanatory model (SLX) employing two spatial weighting matrices differentiating between local and regional neighborhoods. We empirically analyze spatial interaction between individual farms in Norway and additionally perform Monte Carlo simulations exploring the model's performance under different data settings. Results show that including two spatial weighting matrices can indeed reduce the bias resulting from omitted variables. The empirical application identifies different local and regional spatial interdependencies of direct payments with opposite sign.

Keywords: farm growth, spatial competition, spatial interaction, omitted variables, spatially lagged explanatory model

JEL classification: C10, C31, Q12, Q18

[1] **Acknowledgement**: The research presented in this paper is funded by the German Research Foundation (DFG) under grant no. STO 1087/1-1. The authors are grateful to Grete Stokstad and Svein Olav Krogli from the Norwegian Institute of Bioeconomy Research (NIBIO, Ås, Norway) for providing the farm coordinates used in the analysis.

1 Introduction

The application of spatial regression approaches crucially relies on the definition of a 'neighborhood' using a spatial weighting matrix, **W**. One obstacle is that the true neighborhood relations are usually unknown. This is particularly problematic if the estimated spatial interaction effect is sensitive with respect to the definition of **W**. In the literature, the importance of the definition of **W** for the estimation result is controversial. LESAGE and PACE (2011) argue that in most cases, the results are less sensitive to the definition of **W** than is commonly believed. Others, such as HOLLOWAY and LAPAR (2007), found that the spatial correlation in a Spatial Autoregressive model (SAR) model depends heavily on the definition of **W**. STORM et al. (2015) compared three different definitions of **W** and found that the results are rather insensitive with respect to **W**. These conflicting observations suggest that the extent of results being sensitive to the definition of **W** strongly depends on the context.

In this paper we consider a spatially lagged explanatory variable model (SLX) of the form $\mathbf{y} = \mathbf{X\beta} + \mathbf{WX\theta} + \boldsymbol{\varepsilon}$. This type of model has recently been advocated as a more credible alternative to the commonly used SAR model (with the form $\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X\beta} + \boldsymbol{\varepsilon}$) with respect to the identification of the interaction effects (GIBBONS and OVERMAN, 2012 and VEGA and ELHORST, 2015). As shown by GIBBONS and OVERMAN (2012), the reduced from of the SAR model is identical to the SLX model except for higher orders of $\mathbf{w}_i{}'\mathbf{W}^{n-1}\mathbf{X}$ for $n > 1$ despite their very different theoretical motivation. Distinguishing between both specifications and identifying ρ in the SAR model crucially depends on the assumption that the neighboring relationships **W** are exactly known and that $\mathbf{w}_i{}'\mathbf{W}^{n-1}\mathbf{X}$ for $n > 1$ are validly excluded as explanatory variables (i.e. are valid instruments). Since the exact spatial relationship is usually unknown, this requirement is rarely met in empirical applications.

High sensitivity of the estimated interaction effects with respect to the definition of **W**, however, also limits the credibility of the SLX estimation results given that the specific definition of **W** is often to a large degree ad hoc. In this paper we hypothesize that one source of sensitivity of the estimated interaction effect, $\hat{\boldsymbol{\theta}}$, with changes in **W** might be omitted variables $\mathbf{Z}$ that are spatially correlated at a different scale than $\mathbf{WX}$ but nevertheless correlated to $\mathbf{X}$.

The empirical context for our hypothesis is the analysis of farm level spatial interaction in Norway. Farms are assumed to compete on the local land market leading to negative spatial feedbacks for farm development while network effects such as knowledge spillovers or an improved corporate network lead to positive spatial feedbacks. STORM et al. (2015) analyzed how these farm level spatial interactions affect farm survival and change the aggregate impact of farm subsidies. Here we generalize their study by considering farm growth in terms of arable land instead of just farm survival. Specifically, we employ the SLX model to analyze to what extent farm growth can be explained by own and neighboring farm characteristics.

Apart from the direct interaction effects mentioned above (local land market, knowledge spillovers) we expect that there are also, potentially unobserved, spatially correlated variables that affect both farm growth and neighboring characteristics. These variables are likely correlated on a larger spatial scale than the direct interaction. The case of direct payments[2] illustrates this issue: At a local level it may by hypothesized that neighboring direct payments have a negative effects on farm growth due to competition on the land market. However, as direct payments are coupled payments in Norway they are correlated with farm size and specialization. Neighboring direct payments at a regional level might thus reflect differences in the regional farm structure. In case farms grow more strongly in regions with a larger average farm size we thus expect opposite effects of neighboring direct payments at the local and regional level.

To address this problem in our empirical application, we propose to use the SLX model with two spatial weighting matrixes at different scales. The possibility to consider more than one spatial weighting matrix is an additional advantage of the SLX model compared to the SAR model where this is not easily possible (LESAGE and PACE, 2011). With this we aim to distinguish between local and regional spatial interdependencies. Specifically, we expect that the actual interaction between farms primarily takes place on the local level while the interdependencies arising from spatially correlated, omitted variables also takes place at the regional level.

[2] In Norway farms receive subsidies in form of various direct payments based on the number of animals and area under production as well as output produced. These subsidies account for a substantial amount of farm income.

Additionally, we explore this setup with an artificial data generating process (DGP) using Monte Carlo Simulations. Specifically, we consider a DGP with an actual interaction effect and an omitted spatially correlated variable, which also correlates with the interaction variable. We then explore if this setup indeed causes the estimated interaction effects to be sensitive to the neighborhood definition. Secondly, we analyze under which condition we find an omitted variable bias when not correcting for the omitted variable. Finally, we explore to what extent and under which conditions a second "regional" spatial interaction variable can reduce the omitted variable bias and the sensitivity of the estimates. The aim of the Monte Carlo Simulation is to provide some practical guidance under which condition the inclusion of a second interaction variable is helpful.

In the next section, the importance of spatial interaction for farm growth is discussed from a theoretical point of view. The design of the empirical application along with results is discussed in section 3. In section 4 the Monte Carlo Analysis is presented, including the specification of the data generating process (DGP), the simulation setup and results. The final section concludes.

2 Theoretical Framework

In a non-spatial context, the analysis of farm growth is extensively studied (see ZIMMERMANN et al., 2009 for a review). ZIMMERMANN and HECKELEI (2012) and AKIMOWICZ et al. (2013) categorize the determinants of farm growth along with their theoretical underpinning. The selection of control variables included in the growth model is guided by these theoretical considerations. Here we limit the discussion to own and neighboring farm size and direct payments as the main explanatory variables of interest.

Since one of the main hypotheses is that farms interact with each other on the land market, we define farm size in terms of arable land[3].

One of the main determinants of farm growth is technological innovation and economies of scale (COCHRANE, 1958; HARRINGTON and REINSEL, 1995; HALLAM, 1991). Technological innovations reduce per unit costs and with broader adoption also output prices, driving out farms not willing or able to innovate. Innovative farms can grow by picking up the resources released by the leaving farms. Due to better access to information and financing, larger farms tend to be more capable to innovate leading to a positive impact of size on farm growth (WEISS, 1999). With increasing farm size, it might also be possible to realize technological and market-related scale effects increasing total factor productivity and lowering input prices, respectively. These factors would contribute to a positive impact of farm size on farm growth. In the specific case of Norway, however, there are also several policies that differentiate payments by farm size, such that small farms receive relatively more subsidies than large farms (KNUTSEN, 2007: 28). Additionally there exist several upper limits on livestock production[4]. These size discriminating policies might limit the relative growth potential of farms that are already large. The final relationship between farm growth and own size is thus ambiguous.

---

[3] This also includes pasture and fodder production on arable but excluded fodder production on areas where mechanic harvest is not possible because of bushes, rocks etc. in the fields. We exclude these areas because we assume that it is not easily transferred to arable land such that there is no direct substitution between the two.

[4] For example, for dairy operations the total milk quota is limited or concession limits exist for poultry and pig production (KNUTSEN (2007)).

Analogously, the theoretical effects of neighboring farm size on own growth is also ambiguous. On the one hand, farms compete on the land market for the limited available arable land. Consequently, we expect to find an effect of neighboring size opposite to the effect of size on own growth. Specifically, if own size positively affects own growth due to scale effects and a higher rate of innovation we expect a negative effect of neighboring size on own growth due to competition on the land market. In reverse, is the growth potential lower for large farms due to size discriminating policies, we expect positive effects of neighboring size due to lower competition on the land market. Apart from the interaction on the land market, however, farmers are also part of a corporate network with other farmers important for technology adoption, knowledge transfer, and market scale effects (CASE, 1992; ROGERS, 1995; BERGER, 2001; HOLLOWAY et al., 2002; GEZELIUS, 2014; PADEL, 2001; LEWIS et al., 2011; SCHMIDTNER et al., 2012; LAPPLE and KELLEY, 2015; SCHMIDTNER et al., 2015). Under the assumption that larger farms are more innovative, these cooperation effects should lead to a positive effect of neighboring size on own growth. Similarly, larger neighboring farms might also be fostering growth by maintaining a corporate network of suppliers, wholesalers and processors (MOSNIER and WIECK, 2010). Further, GEZELIUS (2014) highlighted the importance of exchanges in labor and machinery between neighboring farms in Norway.

Another driver of farm growth discussed in the literature is the relation between on- and off-farm wages (HALLAM, 1991). Direct payments increase this ratio, which might encourage farmers to increase farm labor input. Similarly, higher direct payments increase the return to land and with it farmer's willingness to pay (WTP) for land and consequently encourage farm growth. Following the same logic in reverse, neighboring direct payments should increase competition on the land market and limit the possibilities for own growth. This is a similar argument as in STORM et al. (2015) with respect to farm survival. It is also reflected in the discussion to what extent government payments capitalize into the land price. Several recent studies (BREUSTEDT and HABERMANN, 2011; FEICHTINGER and SALHOFER, 2014; GUASTELLA et al., 2014) analyze this question empirically by using a spatial lag dependent variable (SAR) model to explain prices with several land characteristics as well as spatially lagged prices.

3 Empirical model

As discussed above we aim to distinguish between local and regional spatial interaction by considering a SLX model including two spatial weighting matrices. Specifically, we consider a model of the form, $\mathbf{y} = \mathbf{X\beta} + \mathbf{W}_L\mathbf{X\theta} + \mathbf{W}_R\mathbf{X\lambda} + \boldsymbol{\varepsilon}$. The intention of the model is that the spatial interaction term $\mathbf{W}_L\mathbf{X}$ primarily captures spatial interaction taking place on a local level, while $\mathbf{W}_R\mathbf{X}$ is more likely to be driven by regional interaction. In the empirical application the regional spatial weighting matrix, $\mathbf{W}_R$, defines neighbors as all farms within a ring from radius 30 km to 60 km around the farm. This distance is set arbitrarily but we assume that it is substantially larger than the distance relevant for competition on the land market or (space dependent) knowledge spillovers. A ring is considered here in order to clearly differentiate the different effects between to local and regional level. For the local spatial weighting matrix, $\mathbf{W}_L$, we vary the radius in order to analyze the sensitivity of the final estimation results ranging from 500 m to 30 km, $\mathbf{W}_L^{0.5km},...,\mathbf{W}_L^{30km}$. In both cases, neighboring definitions are defined as a binary variable with no distance weighting applied.

Both weighting matrices are row standardized. Appendix 7.1 visualizes the neighboring relationships for one exemplary observation.

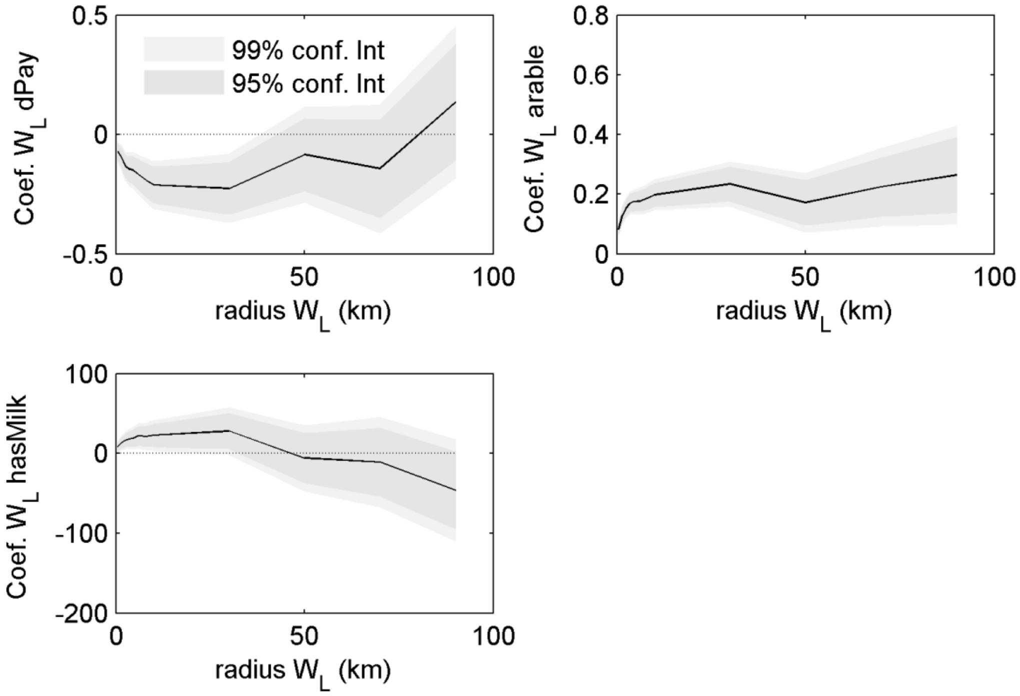
In our empirical application we aim to explain farm growth in terms of arable land between 1999 and 2009 (defined in *daa* = 1/10*ha*). For the analysis, we use a Norwegian data set providing individual, spatially explicit farm-level data of nearly all Norwegian farms in 1999 and 2009. Descriptive statistics for the dependent and the full set of explanatory variables, along with the variable codes, are provided in the appendix 7.2. For model specification, we start with a full model including all explanatory variables. Some insignificant variables are then excluded in cases they are not relevant for the research question.

3.1 *Sensitivity analysis with a single spatial matrix*

Before presenting the estimation results for our model including two spatial weighting matrices we start with a "classical" SLX specification including only one spatial weighting matrix. We vary the radius used for the neighboring definitions from 500 m to 90 km. The results of the model provide a reference for comparison and help illustrating the advantages of considering two spatial weighting matrices.

Figure 1 show the estimated coefficients for three selected spatially lagged variables for varying radii of the neighboring relationships. We observe that the effects of neighboring characteristics change quite substantially with changes in the definition of $\mathbf{W}$. For direct payments ($W_L dPay$) and the share of farms having milk cows ($W_L hasMilk$) we find a significant effect up to a radius of around 30km. Further increases in the radius lead to a change in the sign of the coefficient (even though not becoming significantly different from zero again). Only for arable land ($W_L arable$) the effect remains rather stable. Based on our discussion above, one explanation for the changes in estimated coefficients may be that our spatially lagged variables capture two different effects with different strength at different radii. First, the local interaction on the land market or via knowledge spillovers and second, the regional effect due to confounding variables that affect growth of all farms in the region and cause spatial correlation in our explanatory variables.

Figure 1 Estimated coefficients for the spatial lagged explanatory variables for varying neighboring definitions based on a radius from 0.5 to 90km.



Variable codes: W_LdPay = average neighboring direct payment; $W_L$arable = average neighboring arable land; $W_L$genChange = share of neighbors that had a generational transfer between 1999 and 2009; $W_L$hasMilk = share of neighbors that had milk cows in 1999)

3.2 *Sensitivity analysis of two spatial weighting matrices*

In order to distinguish the two effects we separate two different neighborhoods as discussed above. Appendix 7.3 shows that the correlation of characteristics between local and regional neighborhoods becomes increasingly positive with increasing radius of the local neighborhood. With a local radius of 30 km, the correlation coefficient is around 0.9. This finding supports the hypothesis that explanatory variables are indeed spatially correlated.

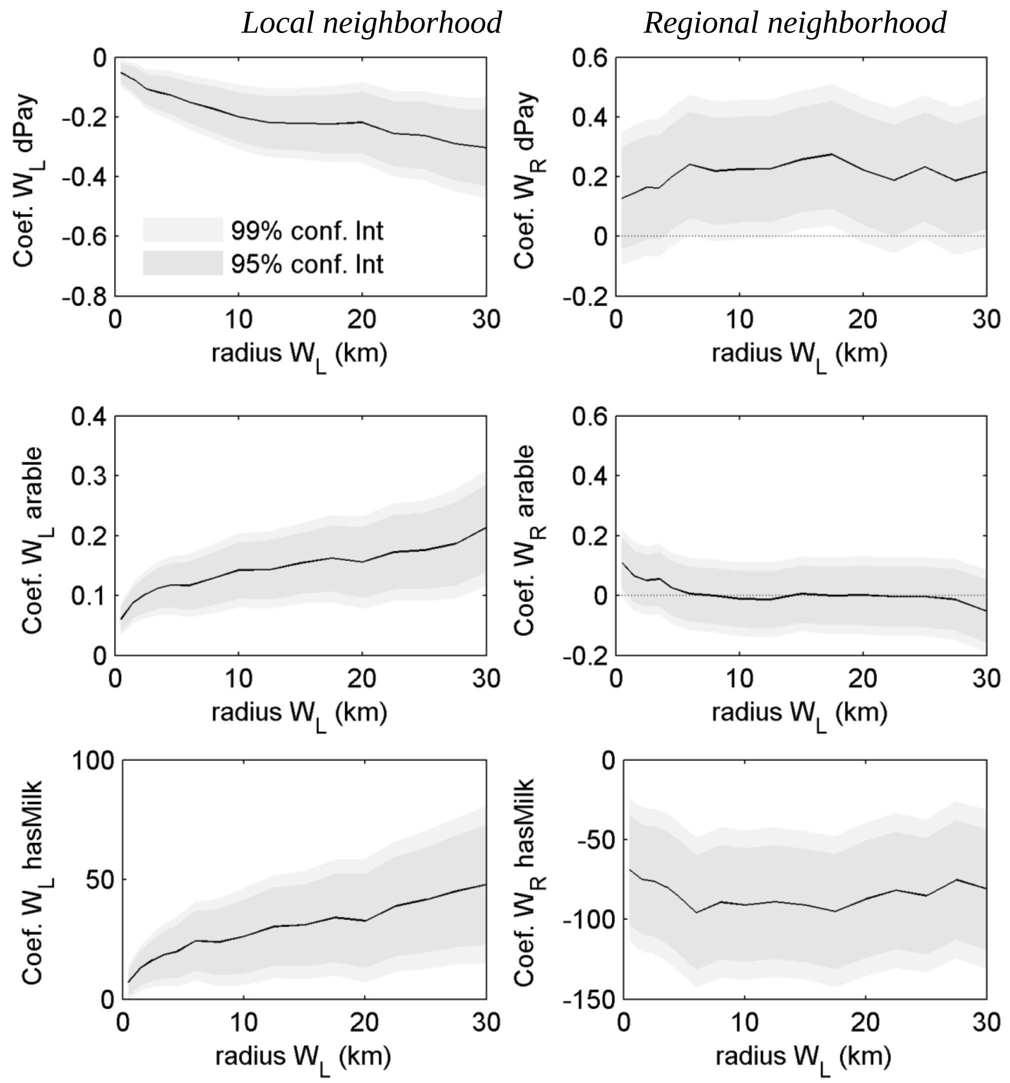
Figure 2 shows the estimated coefficient of selected variables for the local and regional neighborhood (the regression output for three radii of the local weighting matrix are provided in the appendix 7.4). The effects of the local neighborhood largely follow the neighboring effects for the range 0.5 to 30 km for just one spatial weighting matrix (compare figure 1). However, the effects between the local and regional neighborhood differ substantially despite the high spatial correlation of the explanatory variables. For example, consider the effect of neighboring direct payments on farm growth ($W_L dPay$). We find that neighboring payments are highly correlated between the local and regional neighborhood (for local 30km area: correl. coef. = 0.89; see appendix 7.3). Nevertheless, in figure 2 we find a fundamentally different effect of local ($W_L dPay$) and regional neighboring payments ($W_R dPay$). In the local neighborhood increasing direct payments significantly reduce farm growth while in the regional neighborhood increasing direct payments increase farm growth. We find a similar pattern for the milk cow share, with a significant positive effect in the local neighborhood ($W_L hasMilk$) and a significant negative effect in the regional neighborhood ($W_R hasMilk$). The differences might be explained be the fact that dairy farms are more likely to quit during the study period (STORM et al., 2015). On the regional scale, a high share of dairy farms might capture the effect that the farm is located in a dairy region where, on average, farm growth seems to be lower. The effect on the local scale, however, indicates that despite this effect, having a high share of dairy farms among the direct neighbors has a positive effect on growth. This again supports the hypothesis that spatial competition on the land market matters for farm growth. For average neighboring arable land ($W_L arable$) we found significant effects for the local neighborhood only. The regional characteristics seem to be irrelevant for farm growth. This is interesting when considering again Figure 1. There we concluded that the average neighboring arable land is always relevant, basically independent of the definition of the local radius. Here we find that only the close neighborhood matters but not the regional. The positive effect of the average neighboring size might either be explained by corporate network effects or the growth limiting effects of size discriminating policies as discussed in section 2.

These substantial and significant differences between local and regional effects for some variables despite high correlation between them strongly support the hypothesis of two different effects being captured with the spatially lagged variables. The negative effect of local direct payments, for example, supports the hypothesis that farm growth is negatively affected by competition on the land market that intensify as neighboring farms receive higher direct payment. The fact that the regional direct payments show an opposite effect might indicate[5] that the variable picks up regional characteristics which are associated with a higher consolidation and hence growth rate in the region. These characteristics could be, for example, the intensity of production or the productivity in a region. These opposite effect of direct payments between local and regional neighborhood together with the high correlation of the two is a strong indication that farms in the direct neighborhood indeed have a

[5] Note that lower bound of the 95% confidence interval is close to zero for all radii of the local neighbourhood indicating that the effect is weakly supported by the data. However, considering the high correlation between local and regional direct payments this is not surprising and the difference between the two effects is nevertheless rather substantive.

substantially different effect on farm growth, perhaps indicating a more direct interaction effect.

Figure 2 Estimated coefficients for the spatially lagged explanatory variables for varying local neighborhood definitions from a radius of 0.5 to 30 km



Note: The left column presents the coefficient of the spatially lagged variable with the local neighborhood (radius from 500m to 30km). The right column is the coefficient of the spatially lagged variables of all farms within a ring between a fixed radius of 30 to 60km (the coefficient is nevertheless changing due to the changing local neighborhood definition.

4 Monte Carlo Analysis

The hypothesis of two spatial weighting matrices at different spatial scales is based on our empirical observation. In the following, we aim to explore this setup with an artificial data generating process (DGP) using Monte Carlo Simulations. Specifically, we first aim to replicate the observed patterns regarding the sensitivity of the spatial interaction effects with respect to the spatial weighting matrix. Additionally, we use the Monte Carlo simulation to explore under which settings and to what extent a second spatial weighting matrix can improve estimation performance in a mean square error sense.

4.1 *Data generating process*

As outlined in the introduction we consider the DGP with an interaction term and a spatially correlated omitted variable that is also correlated with the included spatial interaction variable. Specifically, we consider the following DGP,

$$\begin{aligned}
\mathbf{y} &= \beta_0 + \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{W}_L\mathbf{x}_2\beta_3 + \mathbf{z}\lambda + \boldsymbol{\varepsilon}_Y \\
\mathbf{x}_2 &= \alpha_1 + \mathbf{z}\alpha_2 + \boldsymbol{\varepsilon}_X \\
\mathbf{z} &= \left(\mathbf{I} - \rho\mathbf{W}_R\right)^{-1}\mathbf{z}^* \\
\mathbf{z}^* &\sim N(\mu_{Z^*}, \sigma_{Z^*}) \\
\boldsymbol{\varepsilon}_Y &\sim N\left(0, \sigma_{\varepsilon_Y}\right) \\
\boldsymbol{\varepsilon}_X &\sim N\left(0, \sigma_{\varepsilon_X}\right)
\end{aligned} \tag{1}$$

where $\mathbf{y}$ is an $(N \times 1)$ dependent variable, $\mathbf{x}_1$ and $\mathbf{x}_2$ are $(N \times 1)$ explanatory variables and $\mathbf{z}$ an $(N \times 1)$ unobserved (and therefore later omitted) spatially correlated variable. The coefficients $\beta_1, \beta_2, \beta_3, \lambda$ specify the marginal effect of explanatory variables, interaction effect $\mathbf{W}_L\mathbf{x}_2$, and omitted variable, $\mathbf{z}$. The explanatory variable $\mathbf{x}_2$ is a linear function of $\mathbf{z}$ with α_2 specifying the correlation between $\mathbf{x}_2$ and $\mathbf{z}$.

We also draw for each observation $i = 1,...,N$ coordinates (ly_i, lx_i) in a Cartesian coordinate system with $lx_i, ly_i \sim U(0,R)$, with R specifying the size of the "landscape". Based on their location we then construct neighboring relationships specified by the $(N \times N)$ spatial weighting matrices $\mathbf{W}_L$ and $\mathbf{W}_R$. Neighbors are defined as all observations within a radius of size s_L and s_R, with $s_L < s_R$, for $\mathbf{W}_L$ and $\mathbf{W}_R$, respectively. Contrary to the empirical section where $\mathbf{W}_R$ is defined to be a ring around the farm, here $\mathbf{W}_R$ is defined as all farms within radius s_R. This definition is more suited to define a variable correlated across space. In the empirical section, the ring was only considered in order to highlight the different signs between the local and the regional interactions effects for some variables. Both spatial weighting matrices are row standardized.

For estimation we observe $\mathbf{x}_1$ and $\mathbf{x}_2$ while $\mathbf{z}$ remains unobserved. For $\mathbf{W}_L$ and $\mathbf{W}_R$ we assume that they might not be known exactly, which is usually the case in an empirical application. Specifically, we assume that we only have information about $\tilde{s}_L$ and $\tilde{s}_R$ defined as

$$\begin{aligned} \tilde{s}_L &= \theta_L s_L \\ \tilde{s}_R &= \theta_R s_R \end{aligned} \quad (2)$$

the parameters θ_L and θ_R therefore specify to what degree the true radius is observed correctly, with $\theta_L, \theta_R = 1$ implying an exact observation of the radius. The actually observed neighboring relationship is denoted by $\tilde{\mathbf{W}}_L$ and $\tilde{\mathbf{W}}_R$.

Summarizing, in order to generate a dataset from (1) we need to specify a set of coefficients $\{\beta_0;\beta_1;\beta_2;\beta_3;\lambda\}$ and parameters, $\{\alpha_1;\alpha_2;\rho;\mu_Z;\sigma_{Z^*};\sigma_{\varepsilon_Y};\sigma_{\varepsilon_{X_1}};\sigma_{\varepsilon_{X_2}};s_L;s_R;\ \theta_L;\theta_R;R;N\}$ of the DGP, then we draw the random variables $\{\varepsilon_Y,\varepsilon_X,\mathbf{z}^*,lx_i,ly_i\}$ to obtain the observables $\{\mathbf{y},\mathbf{x}_1,\mathbf{x}_2,\tilde{s}_L,\tilde{s}_R,\tilde{\mathbf{W}}_L,\tilde{\mathbf{W}}_R\}$ used for estimation.

Two different specifications are considered in the Monte Carlo Simulations. First a model including only the direct interaction

(M1) $$\mathbf{y}=\hat{\beta}_0+\mathbf{x}_1\hat{\beta}_1+\mathbf{x}_2\hat{\beta}_2+\tilde{\mathbf{W}}_L\mathbf{x}_2\hat{\beta}_3+\boldsymbol{\varepsilon}_Y\,. \qquad (3)$$

Since $\mathbf{z}$ is correlated to $\mathbf{x}_2$ it can be expected that the estimation results of the model suffers from an omitted variable bias. Alternatively, we consider an extended model including a second regional interaction term of the form

(M2) $$\mathbf{y}=\hat{\beta}_0+\mathbf{x}_1\hat{\beta}_1+\mathbf{x}_2\hat{\beta}_2+\tilde{\mathbf{W}}_L\mathbf{x}_2\hat{\beta}_3+\tilde{\mathbf{W}}_R\mathbf{x}_2\hat{\beta}_4+\boldsymbol{\varepsilon}_Y\,. \qquad (4)$$

The second regional interaction term $\tilde{\mathbf{W}}_R\mathbf{x}_2$ is intended to capture to some extent the effect of the omitted variable $\mathbf{z}$ due to the correlation between $\mathbf{x}_2$ and $\mathbf{z}$. This specification might still suffer from an omitted variable bias but we like to explore if and if yes how much we can reduce the bias.

4.2 *Monte Carlo Simulation setup*

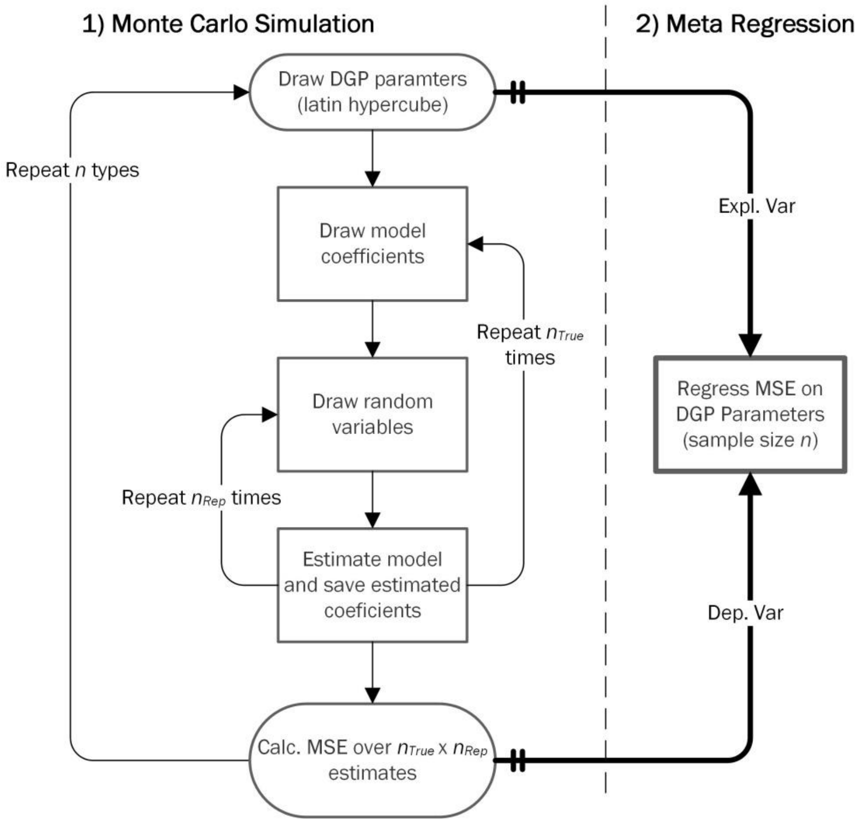
The Model Carlo simulation that we perform in the following is conducted in three steps. First we explore if the DGP in (1) can replicate the empirically observed pattern discussed above, additional we compare how M2 behaves in this respect. Following, we analyze to what extent model M1 suffers from on omitted variable bias under different conditions. Thirdly, we explore under which condition M2 can reduce the omitted variable bias and is superior to M1.

Regarding the first step, we found above that when systematically increasing the radius $\hat{s}_L$ the estimated coefficient for example for direct payments, changes from a negative effect for low values of $\hat{s}_L$ to a positive effect for high values of $\hat{s}_L$. Here we explore if this pattern can be replicated for a specific set of parameters of our DGP. With this set of parameter values, we generate one dataset used for estimation. We then perform estimation several times for different values for $\hat{s}_L$ and save the estimated coefficient $\hat{\beta}_3$. It can then be analyzed if the DGP and the specific set of parameter values results in similar pattern of estimation results as observed empirically. In order to capture sampling noise, several data sets are generated using the same set of parameters and estimation steps are repeated for each of them. The same procedure is repeated for model M2 for comparison.

In the second and third step, we conduct a simulation where we systematically vary the key parameters of the DGP and perform separate Monte Carlo Simulations for each parameter setting. The sets of parameters are created using a Latin hypercube sampling. The n Latin hypercube samples (design matrix) are obtained by drawing for each parameter one draw from each interval $(0,1/n), (1/n,2/n), ..., (1\text{-}1/n,1)$ and permuting these draws randomly. With these combined results from each single Monte Carlo Simulation we then perform a meta-analysis in which we explain the obtained MSE by the design matrix (i.e. the parameters of the DGP) in a linear regression. With this approach we can derive information if and to what extent the considered model suffers from an omitted variable bias under different settings. In each single Monte Carlo Simulation (i.e. fixed set of parameters) we draw n_{true} sets of the model coefficients $\{\beta_0, \beta_1, \beta_2, \beta_3, \lambda\}$. As such, we obtain a set of n_{true} "true" models, one for each set of parameters. These true models can then be used to generate outcomes from the DGP. For each true model we generate n_{rep} outcomes by drawing n_{rep} different sets of the random variables. For each dataset, estimation is performed and the MSE is calculated as $MSE = (3n_{ture}n_{rep})^{-1} \sum_{t=1}^{n_{true}} \sum_{r=1}^{n_{rep}} \sum_{k=2}^{3} (\beta_{ktr} - \hat{\beta}_{ktr})^2$ the difference between the true parameters β_2, β_3 and the estimated parameters $\hat{\beta}_2, \hat{\beta}_3$. The MSE is then averaged over each of the $n_{true} \times n_{rep}$ datasets. As such, we obtain one MSE value for each set of parameters which is then used in the following meta-analysis as the dependent variable. As the MSE is strictly positive, we use a standard Tobit model in the meta-analysis when explaining the MSE by the parameters of the design matrix. A schematic representation of the structure of the meta-analysis is provided in Figure 3.

This type of meta-analysis explaining the MSE is performed for model M1. Then a similar approach is used again for model M2. However, this time our main question is whether including a second spatial interaction term can help in reducing the omitted variable problem or more specifically in which circumstances M2 is/is not superior to M1. Therefore, the setup of the meta-analysis is slightly changed. During the Monte Carlo Simulation for each of the $n_{true} \times n_{rep}$ datasets we estimate both M2 and M1. Then we calculate for each estimate the difference in the obtained MSE, $\delta MSE = MSE_{M2} - MSE_{M1}$. As δMSE is no longer censored at zero we estimate an OLS model in the meta-analysis instead of the Tobit model considered previously.

Figure 3 Schematic representation of the Monte Carlo meta-analysis workflow



Note: Parameters of the Data generating process (DGP) are $\{\alpha_1;\alpha_2;\rho;\mu_Z;\sigma_{Z^*};\sigma_{\varepsilon_Y};\sigma_{\varepsilon_{X_1}};\sigma_{\varepsilon_{X_2}};s_L;s_R;\theta_L;\theta_R;R;N\}$, *model coefficients are* $\{\beta_0,\beta_1,\beta_2,\beta_3,\lambda\}$ *and random variables are* $\{\varepsilon_Y,\varepsilon_X,\mathbf{z}^*,lx_i,ly_i\}$.

Finally, using the results of this last step we apply a classification tree algorithm, a widely used approach in the area of statistical learning (HASTIE et al., 2009). It provides an intuitive way to illustrate the results of the model comparison. In order to simplify the interpretation and visualization of the results we take δMSE to construct a bivariate variable $m_r = \{M1, M2\}$ as

$$m_r = \{M1 \mid \delta MSE_r < 0\} \text{ and } m_r = \{M2 \mid \delta MSE_r \geq 0\}. \quad (5)$$

In the classification three m_r is the response variable that we aim to predict based on the setting of the DGP. The application is based on the MATLAB® Statistics and Machine Learning Toolbox routine "*fitctree*". We use a Gini's diversity index as split criterion. To prune the tree we calculate the 10-fold cross-validation error for each subtree (excluding the highest pruning level) and select the smallest tree whose loss is within one standard error of the minimum loss among all subtrees (routine "*cvLoss*").

4.3 *Monte Carlo Results*

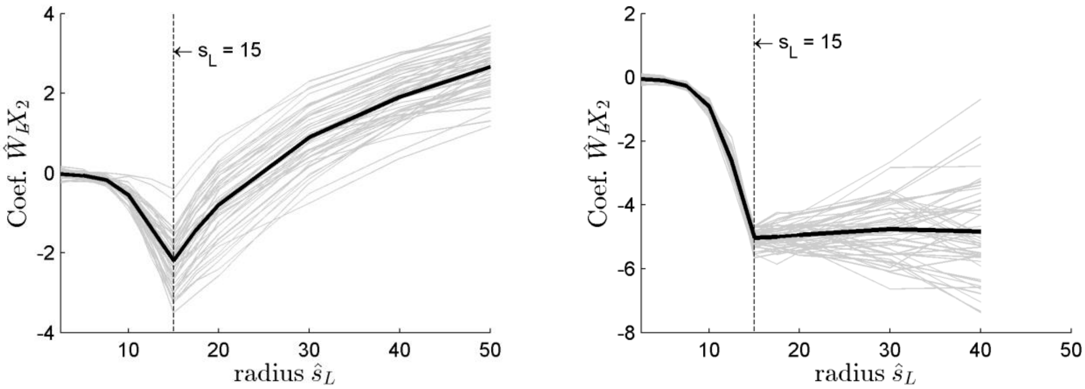
The presentation of the Monte Carlo results follows the structure outline in the empirical section. First we will discuss to what extent the estimated coefficient $\hat{\beta}_3$ is sensitive to the definition $\tilde{\mathbf{W}}_L$ or more specifically $\tilde{s}_L$. In the second and third section we present the results of the meta-analysis for model M1 and M2, respectively.

Sensitivity of the $\hat{\beta}_3$ with respect to $\hat{\mathbf{W}}_L$

Figure 4 (left panel) shows that the DGP mimics the empirical finding of a changing estimated coefficient for $\hat{\beta}_3$ under an appropriate choice of parameters. The chosen parameter setting of the DGP imply that we have a spatially omitted variable $\mathbf{z}$ that is highly spatially correlated $\{\rho = 0.9\}$ at a larger scale (regional neighboring radius equal to $s_R = 50$) than the interaction variable (radius $s_L = 15$). Additionally, with $\alpha_2 = 0.7$ we have a rather strong correlation between $\mathbf{z}$ and $\mathbf{x}_2$. The variation of the random parts in both $\mathbf{z}$ and $\mathbf{x}_2$ is rather large, $\sigma_{Z^*} = 5$ and $\sigma_{\varepsilon_{X_2}} = 5$, compared to the model error, $\sigma_{\varepsilon_Y} = 5$. The model coefficients $\beta_3 = -5$ and $\lambda = 9$ for the variables $\mathbf{z}$ and $\mathbf{x}_2$, respectively, are chosen in order to mimic the changing effect (from negative to positive) of $\hat{\beta}_3$ for varying $\hat{\mathbf{W}}_L$, similar to for the case of neighboring direct payments above (Figure 1). The full specification of the DGP is provided in Figure 4. As outlined above we generated 50 different datasets drawing different errors but using the same parameter specification. For each dataset we repeat the estimation for different values of $\hat{s}_L$. The obtained results are plotted in Figure 4.

For $\hat{s}_L < s_L$ the estimated coefficient decreases and approaches the true coefficient $\beta_3 = -5$. However, even for $\hat{s}_L = s_L$ the estimated coefficient $\hat{\beta}_3$ is still considerably larger than the true coefficient. When further increasing $\hat{s}_L$, we observe again an increasing estimation bias. In this area the coefficient is likely to pick up more and more of the effect of the omitted variable $\mathbf{z}$, at the expense of the effects of $\mathbf{W}_L\mathbf{x}_2$, resulting in an estimated coefficient ranging between $\beta_3 = -5$ and $\lambda = 9$. In the right panel of Figure 4 we repeat the exercise for model M2. For $\hat{s}_L < s_L$ the pattern looks similar as for M1. However, at the point $\hat{s}_L = s_L$, the estimated effect is close to the true value, $\beta_3 = -5$. Increasing the radius further has hardly any effect on the average estimated coefficient across each set of 50 runs and $\hat{\beta}_3$ remains around -5. At the level of a single runs, however, the variation in the estimates increases the further $\hat{s}_L$ deviates from s_L. This graphical inspection indicates that M2 indeed can reduce bias, but that deviations from the true neighboring relationships increase variance in the estimates. This issue is more rigorously explored in the following.

Figure 4: Estimated coefficients $\hat{\beta}_3$ for varying radii $\hat{s}_L$ for 50 different Monte Carlo runs for model M1 and M2. Solid line is the average estimated coefficients across all 50 runs. The dashed line indicates the true local radius $s_L = 15$.



Model M1 Model M2

Notes: DGP specification $\{\alpha_1 = 10; \alpha_2 = 0.7; \rho = 0.9; \mu_Z = 2; \sigma_{Z^*} = 5; \sigma_{\varepsilon_Y} = 1; \sigma_{\varepsilon_{X_1}} = 2; \sigma_{\varepsilon_{X_2}} = 5; s_L = 15; s_R = 50;$ $R = 500; N = 3000; \beta_0 = 1; \beta_1 = 2; \beta_2 = 3; \beta_3 = -5; \lambda = 9, \theta_L = 1, \theta_R = 1\}$.

Omitted variable bias in model M1

In order to analyze if and under which setting model M1 suffers from an omitted variable problem we perform a meta-analysis of a Monte Carlo Simulations as outlined above. Specifically, in the Latin hypercube sampling we considered the following value ranges for a subset of the parameters

$\alpha_2 = [0;1]; \rho = [0;0.95]; \sigma_{z^*} = [1;5]; \sigma_{\varepsilon_Y} = [1;5]; \sigma_{\varepsilon_{X_2}} = [1;5]; R = [200;400];$

$N = [1000;3000]; \theta_L = [0.5;2.5]$. The remaining parameters are kept at fixed values $\alpha_1 = 10; \mu_Z = 2; \sigma_{\varepsilon_{X_1}} = 2; s_L = 15; s_R = 50$. In the Latin hypercube sampling a design matrix of size 2000 is generated. For each of these samples we draw $n_{true} = 10$ sets of the model coefficients $\beta_1, \beta_2, \beta_{3,} z$, each from a uniform distribution in the range $[-10,10]$. For each true model we draw error $n_{rep} = 10$ times, resulting in $n_{true} n_{rep} = 100$ simulations for each of the

Latin hypercube samples. As described above we then explain the obtained MSE in a meta regression. Specifically, we consider the parameters of the DGP as linear and squared effects as explanatory variables. Instead of α_2 we considered the correlation coefficient between $\mathbf{x}_2$ and $\mathbf{z}$ as the relationship between the two also involves a random part ε_X. For some variables, cross effects are considered. We apply a model selection process based on the AIC. As outlined above the approach is applied twice, first, to explain MSE_{M1} (middle columns of Table 1) and secondly to explain $\delta MSE = MSE_{M2} - MSE_{M1}$ (right columns of Table 2). The precise specification along with the estimated coefficients and a description of the model selection is provided in Table 1.

With respect to the model M1 we find that MSE_{M1} increases with increasing variation ($\sigma_{\varepsilon_{Z^*}}$) and increasing spatial correlation ρ of $\mathbf{z}$. On the other hand, increasing variation in $\mathbf{x}_2$ ($\sigma_{\varepsilon_{x2}}$) decreases MSE_{M1}. Additionally, we find a negative effect of the cross term between $\sigma_{\varepsilon_{Z^*}}$ and $\sigma_{\varepsilon_{x2}}$, an indication that the increases in the variation of $\mathbf{z}$ are lower the larger the variation in $\mathbf{x}_2$. Increasing correlation between $\mathbf{z}$ and $\mathbf{x}_2$, however, amplifies both the (increasing) effect of variation of $\mathbf{z}$ and the (decreasing) effect of the variation $\mathbf{x}_2$. These results are intuitive as increases in $\sigma_{\varepsilon_{Z^*}}$ and ρ worsen the omitted variable problem which is counteracted by any (ceteris paribus) increase of variation in $\mathbf{x}_2$. For the correlation coefficient between $\mathbf{z}$ and $\mathbf{x}_2$ we find an almost bell shaped relationship with a maximum at around 0.5 (with all other variables at their means). The increasing part (up to a correlation coefficient of 0.5) can be attributed to an increase in the omitted variables bias. The decreasing part for values above 0.5 is less clear. One explanation might be that with increasing correlation, $\hat{\mathbf{W}}_L\mathbf{x}_2$ is more and more capable of capturing the effect of the $\mathbf{z}\lambda$. This might imply a higher bias for $\hat{\beta}_3$ but may result in a reduction of the bias for $\hat{\beta}_2$. The combined effect might be a reduction in MSE_{M1} for increases of the correlation beyond a correlation coefficient of around 0.5. The number of observations has a negative effect on MSE_{M1}, which might come from two effects. Increasing N increases both the degrees of freedom and the number of neighbors. Similarly, an increase in the regional size has an increasing effect on MSE_{M1}. An explanation might be that the number of neighbors decrease with increasing R (and constant s_L). Having fewer neighbors implies that $\mathbf{W}_L\mathbf{x}_2$ is calculated from fewer observations which might reduce the precision with which the effect of β_3 is measured. The correct definition of the local neighborhood is also decisive. For the coefficient θ_L, which defines the relative error in guessing the neighboring radius, we estimate a U-shaped relationship with a minimum at around 1.2. This is somewhat larger as the expected minimum at 1. These results indicate that it is particularly problematic if the radius is chosen smaller than the true radius, i.e. if $\hat{\theta}_L < 1$ while choosing the radius larger is less of a problem[6]. For empirical application this results is interesting as it suggests that when the radius is only approximately known, the chosen value should rather be at the upper end of a plausible range.

[6] The estimated relationship indicates that (with all other variables at their means) approximately the same MSE_{M1} is incurred for $\hat{\theta}_L = 0.9$ and $\hat{\theta}_L = 1.4$. This implies that chosen $\hat{s}_L$ around 10% lower as the true value has the same negative effect as setting it around 40% higher as the true value.

Table 1: Model estimates explaining MSE_{M1} using a Tobit model and $\delta MSE = MSE_{M2} - MSE_{M1}$ using an OLS model by the settings of the DGP as covariates. The sample of explanatory variables is constructed using a Latin hypercube design (see description in section 3).

Dep. Variable	**MSE_{M1}**		**MSE_{M2} -MSE_{M1}**	
Model	**Tobit**		**OLS**	
Variable	Coef	p-value	Coef	p-value
c	36.2614	0.0000	3.3386	0.0000
N	-0.0008	0.0000	2.9E-05	0.3313
N^2	---	---	---	---
R	0.0362	0.0000	-0.0062	0.0779
R^2	-4.6E-05	0.0000	1.0E-05	0.0879
$corr(\mathbf{x}_2, \mathbf{z})$	38.7826	0.0000	-2.3751	0.0000
$corr(\mathbf{x}_2, \mathbf{z})^2$	-38.2029	0.0000	2.1564	0.0000
ρ	-3.5217	0.0108	3.0219	0.0000
ρ^2	5.7911	0.0000	-3.6472	0.0000
σ_{z^*}	3.4695	0.0000	0.0676	0.5167
$\sigma_{z^*}^2$	---	---	-0.0385	0.0225
$\sigma_{\varepsilon_{x2}}$	-6.5622	0.0000	0.0682	0.5347
$\sigma_{\varepsilon_{x2}}^2$	1.4593	0.0000	-0.0407	0.0077
$\rho \times corr(\mathbf{x}_2, \mathbf{z})$	---	---	---	---
$\rho \times \sigma_{z^*}$	---	---	-0.4670	0.0000
$\rho \times \sigma_{\varepsilon_{x2}}$	---	---	0.2401	0.0000
$\sigma_{z^*} \times \sigma_{\varepsilon_{x2}}$	-1.0798	0.0000	0.0605	0.0001
$corr(\mathbf{x}_2, \mathbf{z}) \times \sigma_{z^*}$	6.7939	0.0000	0.2085	0.0154
$corr(\mathbf{x}_2, \mathbf{z}) \times \sigma_{\varepsilon_{x2}}$	-7.7527	0.0000	---	---
$\hat{\theta}_L$	-67.6650	0.0000	-1.3701	0.0000
$\hat{\theta}_L^2$	28.6740	0.0000	---	---
$\hat{\theta}_R$	***	***	-1.7242	0.0667
$\hat{\theta}_R^2$	***	***	0.6116	0.1417
R^2	0.9999		0.3829	
adj. R^2	0.9999		0.3772	
N	1960		1960	
optimization	bfgs		---	

Notes: Before estimation 1% of observations are excluded each from above and below, in order to eliminate the influence of outlier. Model selection is performed by selection the specification with the lowest AIC. The selection is performed in blocks in order to limit the number of combinations. First, all combinations of squared effects are considered while including all main effects and cross terms. Secondly all combination of cross terms are considered while including all main effects and the best specification for the squared effects obtained in the first step.

Model comparison between M2 and M1

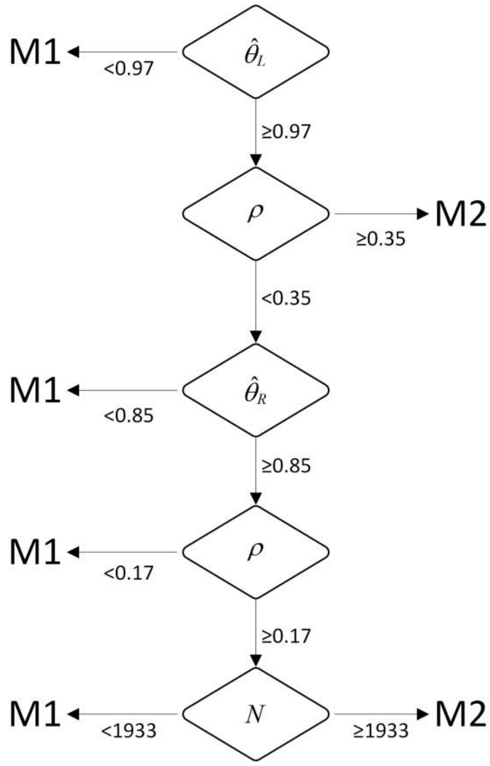
In the next step we present results for a model explaining $\delta MSE = MSE_{M2} - MSE_{M1}$. A positive/negative coefficient in this model implies that M1 is becoming relatively better/worse compared to M2.

An increasing variation of $\mathbf{z}$ has a negative effect, with an increasing rate, on δMSE favoring model M2. The variation in $\mathbf{x}_2$ on the other hand has a decreasing effect, with decreasing rate, favoring model M1. Both effects are amplified by increases in ρ, i.e. the spatial correlation of $\mathbf{z}$. Additionally, increases in $\sigma_{\varepsilon_{x2}}$ increase the effect of σ_{z^*} implying that the positive effect of $\sigma_{\varepsilon_{x2}}$ is larger the larger σ_{z^*} or in reverse that the negative effect of the variation of $\mathbf{z}$ is lower the larger the variation in $\mathbf{x}_2$ (see appendix 7.5). Similarly higher correlation between $\mathbf{x}_2$ and $\mathbf{z}$ increases the effect of σ_{z^*}, i.e. lowering its negative effect. These results indicate that M2 is preferable with high variation of $\mathbf{z}$ and low variation of $\mathbf{x}_2$, particularly when $\mathbf{z}$ is strongly spatially correlated. The effect of the spatial correlation itself (ρ) follows an inverted U-shape with a maximum of around 0.3. Up to that point, increases in ρ favor model M1, while further increases favor M2. For the correlation between $\mathbf{x}_2$ and $\mathbf{z}$ in contrast we find an U-shape relationship. Model M2 is preferred for a correlation coefficient of around 0.4 while increases or decreases from that point on favour M1. This might mirror the observed effect of $corr(\mathbf{x}_2, \mathbf{z})$ on MSE_{M1} following a bell shape. Above we argue that beyond a certain point further increases in $corr(\mathbf{x}_2, \mathbf{z})$ reduce the overall MSE as $\hat{\beta}_2$ is becoming less bias as $\hat{\mathbf{W}}_L \mathbf{x}_2$ is more and more capable of capturing the effect of the $\mathbf{z}\lambda$. As this effect is already included in M2 it might be less effected from changes in

$corr(\mathbf{x}_2, \mathbf{z})$ resulting in the observed U-shape pattern. Also for the regional size, R, we found a U-shape relationship with minimum around 310. This relates to roughly 10 local neighbors and 100 regional neighbors on average. Increasing or decreasing the regional size from that point favors Model M1. The sample size N does not have in significant effect. Increases in the local neighbourhood radius in the range from 0.5 to 1.5 of the true radius lead to a linear decrease in δMSE. Above we concluded for M1 that the radius should not be chosen to narrowly, the results here indicate that this is even more important for M2 as we observed a decrease in δMSE with increasing radius. Similarly, for the regional neighboring radius we find a decreasing effect (with a diminishing rate) on δMSE in the range from 0.8 to 1.5 of the true radius. This again indicates that if model M2 is considered also the regional radius should be chosen rather too large than to narrow.

The decision tree classification approach described above allows exploring and illustrating the same Monte Carlo results in a different way (Figure 5). The approach provides a simply binary classification between model M1 and M2. The first node differentiates based on $\hat{\theta}_L$. For $\hat{\theta}_L < 0.97$ M1 is preferred otherwise we go the next node where we chose M2 if $\rho \geq 0.35$. This implies that, as long as the local radius is not chosen to narrow, M2 is a superior in cases with relatively high spatial correlation in the omitted variable, $\mathbf{z}$. As this case of a highly spatially correlated omitted variable is the starting point for your analysis, the result supports the hypothesis that M2 is indeed a valid extension of M1. The further braches indicate that even in cases where the spatial correlation of the omitted variable is modestly low, $0.17 < \rho < 0.35$, model M2 can be superior as long as the regional radius is not chosen to small ($\hat{\theta}_R \geq 0.85$) and a sufficient sample size is available ($N \geq 1933$) otherwise M1 remains superior. Interestingly, with the pruning level shown in figure 5 which is determined as discussed in section 3, the classification does neither relate to the variation of $\mathbf{x}_2$ nor to the correlation between $\mathbf{z}$ and $\mathbf{x}_2$. The importance of both variables seems to be lower compared the selected variables ($\hat{\theta}_L, \hat{\theta}_R, \rho, N$).

Figure 5 Decision tree comparing model choices between model M1 or M2 based on a MSE comparison in the Monte Carlo Simulation.



5 Conclusion

In this paper we have analyzed the importance of farm level spatial interaction for farm growth. One of the main challenges in the analysis of spatial interaction is to distinguish between spatial interaction as well as spatial correlation arising due to spatial correlation of omitted variables affecting both outcomes and explanatory variable. We approached this challenge by estimating an SLX model with two different spatial weighting matrices in order to distinguish between local and regional interaction effects. Additionally, we systematically analyzed the sensitivity of our results with respect to varying neighboring definitions. Our empirical application, using a Norwegian dataset, indicates that despite high spatial correlation in the explanatory variables the neighboring effects of the explanatory variables differ substantially between local and regional neighborhood. This result provides strong empirical support for the hypothesis that individual farm growth depends substantially on the behavior of directly neighboring farms i.e. that direct interaction occurs. Given that we found a negative effect of the amount of direct payments farms receive in the direct neighborhood, while the effect in the regional neighborhood was positive, indicates that farms compete on the local land market in order to grow and direct payment matter here. This finding contributes to the literature where empirical results concerning spatial farm level interaction and their roll for farm growth are lacking.

Based on this empirical finding we perform Monte Carlo simulations analyzing to what extend we can replicate the empirically observed patters and to what extent a second neighboring interaction variable defined at a larger spatial scale can improve the model. Specifically, we consider a DGP with a spatial interaction variable and a spatial correlated omitted variable which is also correlated to the included interaction variable. Results show that the DGP can indeed reproduce the empirical finding of highly sensitive estimation results with respect to different neighboring definitions. Further, we show that under specific settings a second spatial weighting matrix can indeed improve the models MSE. The simulation result provide practical conclusion for empirical application. Specifically we found that the results crucially depend on a definition of the true neighboring radius. Setting the neighboring radius of the interaction to narrowly has a stronger adverse effect then defining it to broadly. The proposed model with a second spatial interaction variable at regional scale is particularly superior when the spatial correlation of the omitted variable is high. But even for modest spatial correlation it can be superior if the regional neighboring radius is not chosen to narrowly and a sufficient sample size is available.

6 References

AKIMOWICZ, M., MAGRINI, M.-B., RIDIER, A., BERGEZ, J.-E. and REQUIER-DESJARDINS, D. (2013): What Influences Farm Size Growth? An Illustration in Southwestern France. In: Applied Economic Perspectives and Policy 35(2):242–269.

BERGER, T. (2001): Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. In: Agricultural Economics 25(2-3):245–260.

BREUSTEDT, G. and HABERMANN, H. (2011): The Incidence of EU Per-Hectare Payments on Farmland Rental Rates: A Spatial Econometric Analysis of German Farm-Level Data. In: Journal of Agricultural Economics 62(1):225–243.

CASE, A. (1992): Neighborhood influence and technological change. In: Regional Science and Urban Economics - Special Issue Space and Applied Econometrics 22(3):491–508.

COCHRANE, W. (1958): Farm Prices: Myth and Reality. University of Minnesota Press, Minneapolis.

FEICHTINGER, P. and SALHOFER, K. (2014): The common agricultural policy of the EU and agricultural land prices - a spatial econometric approach for Bavaria. Paper presented at the 2014 EAAE International Congress, August 26-29, 2014, Ljubljana, Slovenia.

GEZELIUS, S.S. (2014): Exchange and Social Structure in Norwegian Agricultural Communities: How Farmers Acquire Labour and Capital. In: Sociologia Ruralis 54(2):206–226.

GIBBONS, S. and OVERMAN, H. G. (2012): Mostly Pointless Spatial Econometrics? In: Journal of Regional Science 52(2):172–191.

GUASTELLA, G., MORO, D., SCKOKAI, P. and VENEZIANI, M. (2014): The capitalization of fixed per hectare payment into land rental prices: a spatial econometric analysis of regions in EU. Paper presented at the 2014 EAAE International Congress, August 26-29, 2014, Ljubljana, Slovenia.

HALLAM, A. (1991): Economies of Size and Scale in Agriculture: An Interpretive Review of Empirical Measurement. In: Review of Agricultural Economics 13(1):155–172.

HARRINGTON, D.H. and REINSEL, R. D. (1995): A Synthesis of Forces Driving Structural Change. In: Canadian Journal of Agricultural Economics 43:3–14.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. H. (2009): The elements of statistical learning, 2 ed.

HOLLOWAY, G. and LAPAR, M. L. A. (2007): How Big is Your Neighbourhood? Spatial Implications of Market Participation Among Filipino Smallholders. In: Journal of Agricultural Economics 58(1):37–60.

HOLLOWAY, G., SHANKAR, B. and RAHMANB, S. (2002): Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. In: Agricultural Economics 27(3):383–402.

KNUTSEN, H. (2007): Norwegian Agriculture - Status and Trends 2007. Norwegian Agricultural Economics Research Institute (NILF), OLSO.

LAPPLE, D. and KELLEY, H. (2015): Spatial dependence in the adoption of organic drystock farming in Ireland. In: European Review of Agricultural Economics 42(2):315–337.

LESAGE, J.P. and PACE, R. K. (2011): Pitfalls in Higher Order Model Extensions of Basic Spatial Regression Methodology. In: The Review of Regional Studies 41(1):13–26.

LEWIS, D.J., BARHAM, B. L. and ROBINSON, B. (2011): Are There Spatial Spillovers in the Adoption of Clean Technology? The Case of Organic Dairy Farming. In: Land Economics 87(2):250–267.

MOSNIER, C. and WIECK, C. (2010): Determinants of spatial dynamics of dairy production: a review. Discussion Paper 2010:2, Institute for Food and Resource Economics, University of Bonn.

PADEL, S. (2001): Conversion to Organic Farming: A Typical Example of the Diffusion of an Innovation? In: Sociologia Ruralis 41(1):40–61.

ROGERS, E.M. (1995): Diffusion of innovations, 4th ed. Free Press, New York, NY.

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

SCHMIDTNER, E., LIPPERT, C., ENGLER, B., HARING, A. M., AURBACHER, J. and DABBERT, S. (2012): Spatial distribution of organic farming in Germany: does neighbourhood matter? In: European Review of Agricultural Economics 39(4):661–683.

STORM, H., MITTENZWEI, K. and HECKELEI, T. (2015): Direct Payments, Spatial Competition, and Farm Survival in Norway. In: American Journal of Agricultural Economics 97(4):1192–1205.

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

WEISS, C.R. (1999): Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. In: American Journal of Agricultural Economics 81(1):103–116.

ZIMMERMANN, A. and HECKELEI, T. (2012): Structural Change of European Dairy Farms - A Cross-Regional Analysis. In: Journal of Agricultural Economics 63(3):576–603.

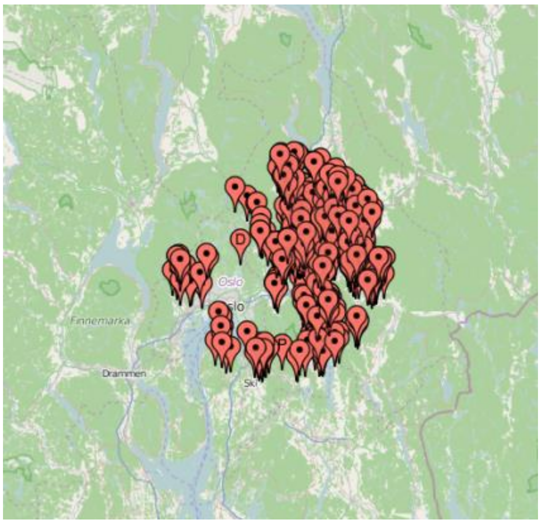
ZIMMERMANN, A., HECKELEI, T. and DOMÍNGUEZ, I. P. (2009): Modelling farm structural change for integrated ex-ante assessment: review of methods and determinants. In: Environmental Science & Policy 12(5):601–618.

7 Appendix

7.1 *Appendix: Local and regional neighboring farms for one exemplifying observation (point A).*

Local neighborhood, $\mathbf{W}_L^{30km}$

Regional neighborhood, $\mathbf{W}_R$ (ring 30 to 60km)



Note: Only a random sample of 500 neighboring farms are shown per maps. The total number of neighboring farms is 1540 and 5122 for the local and regional neighborhood, respectively. Source Maps: http://gpso.de/maps/ Map Data: 2015 OpenStreetMap.

7.2 *Appendix: Descriptive statistics, variable definition and variable codes*

	Code	Unit	Mean	Median	min	max	std
Change in Arable land 1999 to 2009	delArable	daaa	32.34	3.00	-1247.00	5061.00	122.01
Age of the farm holder	age	year	48.20	48.00	7.00	90.00	10.99
Arable land	arable	daa[a]	158.19	125.00	0.01	2994.00	136.11
Observed labor input	obsLabo	hour	2642.67	2500.00	8.00	52330.00	1881.48
Estimated labor requirement	reqLabo	hour	2391.60	2107.97	17.42	42873.53	1805.32
Total direct payments	dPay	1000 Nkr	204.95	195.71	0.00	1252.47	133.23
Total market return	mRet	1000 Nkr	-40.45	-37.41	-2168.29	1403.76	72.81
Ratio observed over estimated labor requirement	laboObs/Req	ratio	1.28	1.11	0.00	65.77	0.95
Dummy if farm has milk cows	hasMilk	binary	0.42	0.00	0.00	1.00	0.49
… has sheep	hasSheep	binary	0.33	0.00	0.00	1.00	0.47
… has sows	hasSows	binary	0.06	0.00	0.00	1.00	0.24
… has poultry	hasPoultry	binary	0.01	0.00	0.00	1.00	0.09
Tot. Direct pay. per total farm area	dPayUaar	1000 Nkr / daa[a]	1.28	1.28	0.00	40.74	0.80
Regional dummy[b] for "Other regions in Eastern Norway"	argR12	binary	0.19	0.00	0.00	1.00	0.39
… "Jæren"	argR21	binary	0.04	0.00	0.00	1.00	0.21
… "Other regions in the counties of Agder and Rogaland"	argR22	binary	0.09	0.00	0.00	1.00	0.28
… "Western Norway"	argR32	binary	0.21	0.00	0.00	1.00	0.41
… "Lowlands in Trøndelag"	argR41	binary	0.08	0.00	0.00	1.00	0.27
… "Other regions in Trøndelag"	argR42	binary	0.08	0.00	0.00	1.00	0.27
… "Northern Norway"	argR52	binary	0.09	0.00	0.00	1.00	0.28

[a]daa = 1/10 ha. [b] reference region is "Lowlands in Eastern Norway"

7.3 *Appendix: Correlation coefficients between spatially lagged variables in the direct neighborhood (radius 500m to 30km) and the farm in a ring between a radius of 30 to 60km.*

Radius of direct Neighbourhood	dPay	age	hasMilk	arable
500 m	0.3054	0.0164	0.3526	0.3056
2 km	0.5020	0.1151	0.5056	0.6063
3 km	0.5990	0.2104	0.5780	0.6832
4 km	0.6519	0.2987	0.6215	0.7239
5 km	0.6886	0.3720	0.6525	0.7492
6 km	0.7247	0.4700	0.6841	0.7745
8 km	0.7566	0.5453	0.7133	0.7952
10 km	0.7794	0.6145	0.7370	0.8113
13 km	0.8026	0.6693	0.7611	0.8257
15 km	0.8218	0.7100	0.7797	0.8377
18 km	0.8375	0.7361	0.7956	0.8478
20 km	0.8509	0.7664	0.8097	0.8579
23 km	0.8631	0.7963	0.8230	0.8679
25 km	0.8736	0.8155	0.8362	0.8776
28 km	0.8838	0.8310	0.8494	0.8875
30 km	0.8941	0.8430	0.8623	0.8980

Note: See Appendix 7.2 for variable codes.

7.4 *Appendix: Regression results for 3 radii and selected variables*

	W_km2		W_km15		W_km30	
Variable	**Coef**	**p-value**	**Coef**	**p-value**	**Coef**	**p-value**
const	71.2361	0.0443	26.8659	0.4994	-0.7593	0.9859
age	-4.4820	0.0000	-4.5359	0.0000	-4.4265	0.0000
age^2	0.0429	0.0000	0.0431	0.0000	0.0423	0.0000
arable	-0.1021	0.0000	-0.1092	0.0000	-0.1073	0.0000
obsLabo	0.0009	0.2078	0.0012	0.0871	0.0008	0.2352
reqLabo	0.0042	0.0000	0.0034	0.0007	0.0038	0.0001
dPay	0.1095	0.0000	0.1257	0.0000	0.1182	0.0000
mRet	-0.0283	0.0552	-0.0236	0.1175	-0.0242	0.1106
laboObs/Req	1.5565	0.1203	1.2219	0.2310	1.3784	0.1825
hasMilk	-13.1539	0.0000	-14.5153	0.0000	-13.8421	0.0000
….						
W_L dPay	-0.0742	0.0006	-0.2210	0.0000	-0.3030	0.0000
W_L arable	0.0889	0.0000	0.1542	0.0000	0.2134	0.0000
W_L reqLabo	-0.0028	0.0345	0.0002	0.9452	0.0003	0.9325
W_L hasMilk	13.0853	0.0003	31.0686	0.0004	47.9661	0.0002
W_L age	0.2026	0.0000	0.2995	0.4409	1.2183	0.0380
….						
W_R dPay	0.1451	0.0982	0.2574	0.0041	0.2178	0.0257
W_R arable	0.0645	0.1427	0.0061	0.8996	-0.0517	0.3331
W_R reqLabo	0.0053	0.2887	0.0061	0.2381	0.0105	0.0700
W_R hasMilk	-74.8880	0.0000	-90.9625	0.0000	-81.0408	0.0000
W_R age	0.5744	0.4234	1.3895	0.0608	1.0868	0.1577
….						
n	32043	---	30940	---	30257	---
AIC	8.5973	---	8.6324	---	8.6547	---
rsqr	0.0424	---	0.0438	---	0.0425	---
rbar	0.0413	---	0.0426	---	0.0413	---

Note: The definition of the variable codes is provided in Appendix 7.3. Before estimation 1% of observations are excluded each from above and below, in order to eliminate the influence of outlier. This elimination of outliers cause the difference in the sample size across the different runs.

7.5 *Appendix: Predicted values for* $\delta MSE = MSE_{M2} - MSE_{M1}$ *based on the estimated model presented in Table 1 (right columns). All other variables at kept at their respective means.*