Modelling binary outcomes in a classroom setting: comparing spatial and traditional estimations of peer effects

Annikka Lemmens*(376396)

Master thesis MSc Econometrics

Supervisor: dr. W. Wang

Second assessor: dr. A. Camehl

Date final version: August 10, 2022

Abstract

This thesis compares a traditional and a spatial approach to peer effects modeling to study the uptake of the supplementary grant in vocational education (MBO). This is a classroom setting with a binary outcome variable, in which endogenous peer effects are of particular interest. I show that a spatial autoregressive model and a traditional instrumental variable model are nearly identical in theory and practice. When outcomes are continuous and errors are exogenous, a spatial autoregressive combined model is only slightly different. Despite their similarities, the way modellers think about issues, such as the reflection problem and selection bias is different. Furthermore, spatial modellers have difficulty implementing of fixed effects when class sizes vary. On the other hand, the spatial literature on binary outcome variables is more advanced, especially for spatial logit models. I illustrate the similarities and the limitations of the respective approaches empirically by means of a Monte Carlo simulation study with three different data generating processes. Focusing on the supplementary grant, neither approach seems to be suitable to research peer effects in this cross-sectional setting. The instruments used by both models seem to not be exogenous, the exclusion restriction is hard to defend. Furthermore, using the spatial logit model turned out to be unfeasible for a data set of over 250,000 observations, due to computational issues.

*The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Contents

Acknowledgements

Firstly, I want to thank my colleagues at the Netherlands Bureau for Economic Policy Analysis, and especially dr. Maria Zumbuehl for all her help during the writing process. Secondly, thanks to my supervisor dr. Wendun Wang for his insightful remarks. Also thanks to prof. J.P. Elhorst, our discussion on spatial econometrics was greatly appreciated. Last but not least, thank you so much for supporting me, Kevin Scheurwater & fam Lemmens-Rooda

1 Introduction

Many Dutch students that are entitled to a supplementary study grant do not apply for one. A recent study on non-uptake of supplementary study grants by students in vocational education (MBO) shows that about 26% of entitled students do not apply for the grant, leading to an average missed income of about 190 euros per month(Zumbuehl and Magnée, 2022). Similarly, this holds for about 25% of students in higher education (HBO/WO) (Visser et al., 2020). At the same time, 41% of the same students apply for a study loan, which has worse loan conditions. Apparently, while students are in need of additional income, they do not apply for the supplementary grant.

There are multiple factors that might influence a student to apply for a supplementary grant. One explanation is that students are affected by the choices of their classmates. Knowing the role that a student's network plays in the decision to apply for a grant can give some indication on how to effectively approach the problem of non-uptake. If peer effects are prevalent, introducing a policy measure to increase uptake can have much a larger effect than originally expected, as information spreads from one student to the other. Researching peer effects is, however, quite difficult, especially using a cross-sectional data set without information on which students interact with each other.

The aim of this thesis is to answer the question: *"Can the traditional and spatial approach be used to discover whether there are peer effects in the uptake of the supplementary grant in MBO?"*. Traditional linear-in-means modelling as used in education economics is a well-established method to research peer effects. Spatial econometrics stems from a different econometric field and is up-and-coming in the area of peer effects. Lin (2010) and Abbiati and Pratschke (2021) have discussed the advantages of the spatial approach over the traditional approach when relationships within classrooms are known. On the other hand, Gibbons and Overman (2012), and Corrado and Fingleton (2012) have provided critical evaluations of problems that plague both approaches, but spatial modelers mostly seem to ignore. A full comparison of the spatial approach and the traditional approach in a classroom setting has not been done before. Furthermore, the issues plaguing the two approaches have not yet been discussed specifically for a binary outcome variable, like the uptake of the supplementary grant.

The thesis is structured as follows. Firstly, I provide an overview of what peer effects are, which problems plague modellers and empirical studies on peer effects. Next, I elaborate on the methodology of the traditional approach. This is followed by a discussion of spatial models in section

3.2 and a comparison in section 3.3. I show that the two approaches actually do not differ in theory in a basic setting, but the way modellers think about the identification issues differs. In section 4 I specifically look at binary outcome variables. Here, spatial modellers are more advanced than traditional modellers, with a much more well-established literature on logit models. Afterwards, I compare the models empirically in section 5 using Monte Carlo simulations and three different data-generating processes. In section 6 I focus specifically on the uptake of the supplementary grant, to see whether the two approaches are appropriate to research this issue. The thesis ends with a conclusion in section 7 and a discussion in section 8.

2 Literature review

Much of the current literature on peer effects can be traced back to a seminal paper by Manski (1993). He poses the following equation.

$$\mathbb{E}(y|x,z) = \alpha + \beta\mathbb{E}(y|x) + \mathbb{E}(z|x)'\gamma + z'\zeta + \mathbb{E}(u|x,z), \qquad \mathbb{E}(u|x,z) = x'\delta \tag{1}$$

The behaviour of interest is y. In Manski's model, x are attributes characterizing an individual's reference group, and (z, u) are attributes that directly affect behaviour. The attributes z are observable, whereas u are unobservable. A person's outcome y depends on their peer group in three different ways. There are endogenous peer effects: $\beta\mathbb{E}(y|x)$, exogenous/contextual effects: $\mathbb{E}(z|x)'\gamma$ and correlated effects: $x'\delta$. Endogenous effects occur when someone is influenced to behave differently due to how their group behaves. Contextual effects occur when someone's behaviour is influenced by the exogenous characteristics of other members in their group. And lastly, correlated effects occur because individuals in the same group act similarly, because they are alike in their unobservable characteristics or face a similar institutional environment (Manski, 1993).[1] Often, correlated effects are only seen as a nuisance, something that disturbs the estimation of exogenous and endogenous peer effects.

[1]Manksi illustrates these three different types of peer effects in the following way (p. 533): "*Consider the high school achievement of a teenager. There is an endogenous effect if, all else equal, achievement tends to vary with the average achievement of the students in his school, ethnic group, or other reference group. There is an exogenous effect if achievement varies with, say, the socio-economic composition of the reference group. There are correlated effects if youths in the same school tend to achieve similarly because they are for example taught by the same teachers.*"

2.1 Identification issues

Estimating peer effects is not straightforward. Researchers need to take into account different identification issues, that make it difficult to accurately describe the different types of peer effects. In the rest of this section I elaborate on three prominent problems.

Besides introducing the three different types of peer effects, Manski (1993) also introduced the reflection problem. This is one of the most fundamental problems in the peer effects literature. The reflection problem entails that endogenous and exogenous peer effects cannot be identified separately. First integrating over z leads to the following equation (p. 534):

$$\mathbb{E}(y|x) = \alpha + \beta\mathbb{E}(y|x) + \mathbb{E}(z|x)'\gamma + \mathbb{E}(z|x)'\zeta + x'\delta \tag{2}$$

This equation has a unique solution, as long as $\beta \neq 1$:

$$\mathbb{E}(y|x) = \frac{\alpha}{1-\beta} + \frac{\mathbb{E}(z|x)'(\gamma + \zeta)}{1-\beta} + \frac{x'\delta}{1-\beta} \tag{3}$$

If you insert (3) in (1), you obtain a reduced form model with composite parameters $\frac{\alpha}{1-\beta}$, $\frac{\gamma+\beta\zeta}{1-\beta}$, $\frac{\delta}{1-\beta}$. In this model ζ and the composite parameters are identified, as long as $\beta \neq 1$, and the regressors are linearly independent in the population. The parameters α, β, γ, δ are all unidentified, only the composites are identified. The reflection problem therefore shows that you cannot distinguish endogenous effects from exogenous or correlated effects, although you can show whether some form of peer effects are present, namely if $\frac{\gamma+\beta\zeta}{1-\beta} \neq 0$, then either $\beta\zeta$ or γ is non-zero. Following Manski, most attention has been spent on disentangling endogenous effects from exogenous effects, while looking at correlated effects as more of a nuisance.

Another issue is endogeneity. $\mathbb{E}(y|x)$ is not an exogenous regressor for $\mathbb{E}(y|x, z)$ in (1). This endogeneity is partially due simultaneity, as peer interaction is not a one-way street. A person is influenced by her peers, while at the same time also exerting influence on her peer group. This bi-directional peer influence leads peers' outcomes to be correlated with the error term, inducing simultaneity bias. (An, 2015)

An additional issue is that of selection bias. The formation of peer groups can be endogenous. This happens if individuals who are similar in some dimension are grouped together. Self-selection may result in the overestimation of the (endogenous) peer effect. Take for example at a particular student with high-ability classmates, who gets high grades herself. It is impossible to know whether she was influenced by her peers or whether her high grades are due to some correlated effects from being assigned the same teacher in a class for advanced students. (Paloyo, 2020)

Over the years, different solutions to these identification issues have been proposed, as I discuss in sections 3.1 for the traditional approach and 3.2 for the spatial approach.

2.2 Empirical research on peer effects

Manski (1993) uses an example of a school setting to explain his three different types of peer effects. Students remain some of the most important subjects of peer effects research. Classmates in particular are found to influence students' grades (Lin, 2010; Sacerdote, 2011; Paloyo, 2020. High performing peers make other students perform better as well. Classmates also matter when it comes to deciding to enroll in higher education (Abbiati and Pratschke, 2021), choosing a study (De Giorgi et al., 2007) and even for occupation later in life (Brenøe and Zölitz, 2020).

Most of the research uses primary schools or high schools, but there is research on college students as well. In an overview of the literature Sacerdote (2011) found that the influence of roommates or course mates on grades seems to be quite moderate in most settings). On the other hand, the influence on more "social" outcomes is quite large. For instance, roommates greatly affect the probability of binge drinking or joining a student association (Sacerdote, 2011).

Financial aid or grants are usually used as regressors when looking at other outcomes, but not as dependent variables in peer effects models. For instance, Epple, Romano, and Sieg (2003) created a model in which American universities can use pricing and financial aid as tools to cultivate a student population that positively influences each other. Peer effects for the Dutch student finance system specifically have also not been researched before. An explanation of the Dutch student finance system can be found in the Appendix.

There is evidence that peers influence financial decisions. For example, Markussen and Røed (2015) show that the more people in your environment rely on social security, the more likely it is that you will make a social security claim as well. Åslund and Fredriksson (2009) find that long-term welfare dependence increases if a refugee is placed in a welfare dependent community. Peers also matter for quitting decisions, with people attempting to finish a difficult task for a much longer time if a peer is working on it as well (Rosaz et al., 2016).

3 Methods for continuous outcomes

Traditional leave-one-out linear-in-means models (LIM models) are often used to study peer effects.[2] These models include the mean value of the behaviour of the peer group as a regressor to estimate someone's behaviour. However, to calculate the mean value, the value of the person of interest is excluded (leave-one-out). In section 3.1, more details about this traditional approach are given. As most of the identification issues from the previous section were discovered in the traditional peer effects literature, the traditional modelling approach can be neatly categorised by the responses to those challenges.

Another, newer, way of estimating the different peer effects is by using spatial models. This methodology was popularised in geographical research. Instead of taking the mean for the expectations in equation (1), spatial models use a spatial weighting matrix W. In the last decade spatial models have gained some traction in peer effects research as well (Lin, 2010; Abbiati and Pratschke, 2021). Spatial models are discussed more extensively in section 3.2. The way spatial modellers have thought about the challenges from the previous section is different from the way traditional modellers have done it as I will explain in section 3.2. Despite this, the challenges remain helpful in structuring a discussion on how spatial models work.

The empirical context of this thesis is the uptake of the supplementary grant in MBO. Given that a student either receives a supplementary grant or she does not, this concerns a binary outcome variable. However, it is useful to first look at continuous outcome variables. There is more research on peer effects in continuous settings, for both the traditional and spatial approach. Besides that, many issues are easier to explain in a continuous setting, but carry over to binary outcome variables.

Furthermore, the peer effects for the uptake of the supplementary grant occur in a classroom setting. I do not have data on which students communicate or consider each other friends. I do have data on who started their studies at the same time, at the same institution, doing the same programme. These students are considered classmates. Spatial modellers call this a "non-overlapping contiguity" setting (Anselin, 2020). In such a setting it is not possible to distinguish how close people feel to each other, so "distances" between people in the same network are all equal. Everyone in the same class is considered to influence each other in the same manner. Non-overlapping means that people who are part of one peer group, are not part of another peer group.

[2] See for instance overviews in Epple and Romano (2011) and Paloyo (2020) for studies in education economics.

While these assumptions are quite unrealistic, unfortunately as long as data about real friendship networks or interactions are unavailable, it is the best we can do. A non-overlapping contiguity context has been used by spatial modellers in for example Lee (2007), Baltagi, Deng, et al. (2018) and Zhao and Qu (2021) and by traditional modellers in most studies surveyed by Paloyo (2020).

3.1 Traditional approach

In essence, traditional researchers of peer effects want to model (Yeung and Nguyen-Hoang, 2016)

$$y_{ig} = \beta \bar{y}_{-ig} + \bar{x}_{-ig}\gamma' + x_{ig}\zeta' + u_{ig} \tag{4}$$

Explaining (4) in terms of empirical application: y_{ig} is the uptake of the supplementary grant by person i, part of peer group g. This is partially determined by the average uptake in her peer group excluding her own uptake ($\bar{y}_{-ig}$). Partially it is determined by some variables x_{ig}, that say something about individual i like for example, gender, eligibility for the maximum amount of the grant or migration background. Note that I use the more conventional x for these variables, instead of Manski's z. A student is also influenced by the average characteristics of her peer group, again, excluding her own ($\bar{x}_{-ig}$). The error term is given by u_{ig}, which may have an aspect of correlated peer effects.

Of course, due to the issues discussed in section 2 correct estimation of the different peer effects is difficult. Over the years, different techniques have been proposed to adjust for the reflection problem, endogeneity, the presence of unobserved correlated effects and selection bias. The most important techniques are arguing from theory, estimating a reduced form, using instrumental variables, and using fixed effects.

3.1.1 Reflection problem and endogeneity

Traditional modelers find overcoming the reflection problem quite tricky. They spend a lot of time and effort to overcome it. The most popular solutions in the cross-sectional literature are the following: 1) theoretically arguing that there are only endogenous or only exogenous effects, 2) estimating a reduced form regression and 3) making use of instrumental variables. (Yeung and Nguyen-Hoang, 2016)

Manski (1993) already mentioned the first solution. In some situations you can theoretically argue that people are only influenced by behaviour or only influenced by the characteristics of their peer group. This solves the reflection problem, as there is no need to disentangle the different types of peer effects. For this solution a regression with only $\beta\bar{y}_{-ig}$ or only $\bar{x}_{-ig}\gamma'$ is estimated. This approach is for example employed by Gaviria and Raphael (2001) and Mora and Oreopoulos (2011).

Estimating a reduced form regression is similar in practice, as again a regression with only $\beta\bar{y}_{-ig}$ or $\bar{x}_{-ig}\gamma'$ is estimated. However, the reasoning behind it is different. Economists who use the reduced form acknowledge the reflection problem, and say that the peer effects that are found are a combination of both endogenous and exogenous peer effects. It is not really a solution to the reflection problem, as much as it is an acceptance that exogenous and endogenous peer effects cannot be disentangled. This approach has been taken by for example Ammermueller and Pischke (2009) and Ryabov (2011).

A more ambitious solution is to make use of instrumental variables (IV). The basic idea is to utilize the exogenous variation in the instruments to facilitate identification. Different IVs have been proposed over the years. Bramoullé et al. (2009) suggest making use of person i's "friends of friends", who are not person i's friends. However, this is not an option for non-overlapping peer groups, like in the case of the supplementary grant. Hinke et al. (2019) illustrate another popular set of IVs: characteristics that help determine a peer's outcome, but only influence person i's behaviour through the behaviour of the peer group. In this way the endogeneity of y_{-ig} is solved by using exogenous instruments.

Instrumental variables can also be a remedy for simultaneity bias.[3] To prevent simultaneity, the instruments have to pertain to something that is determined before the behaviour is exhibited. For example, Figlio (2007) uses the proportion of male students with a traditionally female sounding name as an instrument for the proportion of disruptive children in a classroom.

Using instrumental variables is quite popular. For instance, Yeung and Nguyen-Hoang (2016) find that 32 out of 95 studies into peer effects in the years 1980-2015 use this approach.[4] However, there are certain assumptions that need to be met when employing IV. The instruments need to meet the independence restriction. They should also only influence y_{ig} through the average peer

[3] Another solution to both the reflection problem and simultaneity bias would be to use lagged dependent variables for the peer outcomes. Someone's behaviour at time t can be influenced by peer behaviour at $t-1$, but $\bar{y}_{-ig,t-1}$ is not influenced by $y_{ig,t}$. However, that is only a viable solution with panel data.

[4] The most popular approach is lagged dependent variables (44 out of 95), but this can only be used with access to panel data.

behaviour, but not through some other channel. This is the exclusion restriction (Hinke et al., 2019). And the instruments need to be strongly correlated with $\bar{y}_{-ig}$. Yeung and Nguyen-Hoang (2016) argue that it is hard to come by studies that meet these assumptions. The doubts on whether an instrument is appropriate have resulted in traditional researchers of peer effects having to explicitly defend their instrument choices.

3.1.2 Correlated effects and selection bias

Correlated effects and selection bias can be solved with less disputed measures. Due to unobserved correlated effects, u_{ig} can have a common part for members of the same group g. Oftentimes modellers include fixed effects to take this into account (Brenøe and Zölitz, 2020; Lavy and Schlosser, 2011). Fixed effects are also a way of dealing with selection bias. Economists prefer to have random assignment into groups. However, when that is not possible, it is argued that fixed effects at an aggregate level can alleviate the problem (Paloyo, 2020). If the peer groups are formed at a lower level, like at the class level, adding cluster fixed effects at the aggregate school level removes the sorting effect (Hinke et al., 2019). Including the fixed effects α_r at the cluster level r, leads to the following equation.

$$y_{ig} = \alpha_r + \beta\bar{y}_{-ig} + \bar{x}_{-ig}\gamma' + x_{ig}\zeta' + \epsilon_{ig} \tag{5}$$

The cluster fixed effects are either added as dummy variables or they are removed from the equation by performing a within-transformation. Dummies are an option if the number of clusters is limited, and the number of individuals in each cluster is large enough. Otherwise the estimation suffers from the incidental parameter problem. The within-transformation is always possible, and solves the incidental parameter problem. This transformation uses deviations from the cluster means for each variable (Hinke et al., 2019).

3.1.3 Model specification

For the rest of this thesis, I am most interested in endongenous peer effects. This is because the uptake of the supplementary grant is the motivation behind the comparison of the traditional approach to the spatial approach. Endogenous peer effects are likely to be present, and also matter most for policy as they show how information spreads within peer groups. I therefore want a specification that at least includes endogenous peer effects $\beta\bar{y}_{-ig}$.

If you either argue that only endogenous peer effects are present, or you estimate a reduced form regression, you can use a two stage least squares (2SLS) regression.

$$\bar{y}_{-ig} = \alpha_r + \gamma\bar{x}_{-ig} + x_{ig}\tilde{\zeta}' + \tilde{\epsilon}_{-ig} \tag{6}$$

$$y_{ig} = \alpha_r + \beta\hat{\bar{y}}_{-ig} + x_{ig}\zeta' + \epsilon_{ig} \tag{7}$$

Alternatively, if you suspect exogenous peer effects are present as well, and you use instruments to overcome both the reflection problem and simultaneity bias, the regression looks as follows

$$\bar{y}_{-ig} = \alpha_r + \gamma\bar{x}_{-ig} + \bar{z}_{-ig}\theta' + x_{ig}\tilde{\zeta}' + z_{ig}\tilde{\delta}'\tilde{\epsilon}_{-ig} \tag{8}$$

$$y_{ig} = \alpha_r + \beta\hat{\bar{y}}_{-ig} + \bar{x}_{-ig}\gamma' + x_{ig}\zeta' + z_{ig}\delta' + \epsilon_{ig} \tag{9}$$

To make the equations more clear, I will relate them to the empirical application. In the first stage in (6) the average uptake of the grant is regressed on all average exogenous peer characteristics, like gender, migration background, eligibility for the maximum grant, and whether parents receive welfare benefits. Similarly, in (8), the average uptake is regressed on the same characteristics, but they are divided into X and Z. Z are the instrumental variables which are not visible to others in the peer group, like the eligibility for the maximum grant and whether a parent receives welfare benefits. X are the more visible characteristics like gender and migration background. Both can influence whether a peer applies for a supplementary grant. But it can be argued that the invisible characteristics only influence person i's decision through the behaviour change in the peer group, while the visible characteristics could potentially matter directly. In the second stages (7) and (9), these estimations of $\bar{y}_{-ig}$ are plugged in. In (7) $\bar{x}_{-ig}$ is not added directly, as again, it is assumed that it only influences y_{ig} through $\hat{\bar{y}}_{-ig}$. In (9) $\bar{x}_{-ig}$ is added, but $\bar{z}_{-ig}$ is not added.

Which specification should be chosen according to traditional researchers, depends on whether we believe exogenous peer effects are present or not. And on whether we feel more comfortable with a reduced form regression or trust the instruments to solve the reflection problem. The choice between these two IV-models is done in section 3.3, as I want the traditional and spatial models to be comparable and spatial modelers have similar concerns.

3.2 Spatial approach

Burridge et al. (2016) provides us with the most extensive spatial model to investigate peer effects, as can be seen in(10).

$$Y = \beta WY + X\zeta + WX\gamma + U, \qquad U = \lambda WU + \epsilon \tag{10}$$

Burridge et al. (2016) call this model the GNS: general nesting spatial model. The main feature that stands out in opposition to the traditional model, is the inclusion of the W-matrix. This is a matrix of size NxN that shows the relationships between every member of the data set. People that do not know each other receive a 0. Every observation on the diagonal of the matrix receives a 0 as well, as we are not interested in including the relationship between a persons behaviour with her own behaviour. People that know each other, are part of the same network and their relationships are not set at 0. Instead the distance between them is characterised by some assumption about how much the different members of the network interact. By convention W is row-normalised after setting these distances (Dubin et al., 2009). As mentioned, in the classroom setting for the empirical application we do not know which members of the same class interact. W therefore has a very particular form. Of course, it is still a block matrix with zero diagonals. Additionally within each block, every row has n_g observations with distance $\frac{1}{n_g}$, if n_g is the number of people in class g.

Spatial econometricians have proposed different responses to the issues described in section 2. Different solutions exist within the spatial approach as well as between the spatial and traditional approach. The most important solutions are zero-diagonal W-matrices, differing group sizes, information on correlated effects, instrumental variables, and fixed effects.

3.2.1 Reflection problem and endogeneity

The GNS is often regarded as only a theoretical option. For example, in the influential textbook by LeSage and Pace (2009), the GNS is mentioned, but not taken seriously for empirical research. This disregard of the GNS is sometimes caused by a loose interpretation of Manski's reflection problem, according to Burridge et al. (2016). Most of the spatial literature is, however, silent about whether the reflection problem forms an issue for estimating peer effects. It is often only implicitly assumed as a reason for why GNS should not be used.

In contrast, there have been some spatial researchers that explicitly think about the reflection problem, amongst whom are Burridge et al. (2016) and Lee, Liu, et al. (2010).[5] Lee, Liu and Lin use an idea that was first worked out by Lee (2007). Lee (2007) formally shows that estimating a GNS is possible, if you use a spatial weights matrix that is specified as an equally weighted group interaction matrix with a zero diagonal and you have enough variation in the number of group members, with groups that are not too large. The crux is to have zero diagonal elements and many different group sizes. A classroom setting with many different class sizes fits with this W-matrix. Manski's equation did not exclude person i from the expected values, while exclusion leads to a number of varying group-averages, which make identification possible (Lee, 2007). Lee (2007) also illustrated the potential of the GNS empirically, by using it to identify the three different types of peer effects in a classroom setting. Burridge et al. (2016) go even further than Lee (2007), finding that the crux to solving the reflection problem is just the zero diagonal, regardless of the group sizes.

In contrast, Lin (2010) mentions a different solution and writes that unlike the Manski model, the GNS incorporates the information contained in the spatially correlated error terms. This additional source of information is valuable for identification and evading the reflection problem. Another solution is offered in Zhao and Qu (2021), who use a technique in which a "neighbour's neighbour" provides identification, similarly to what Bramoullé et al. (2009) suggest in the traditional approach.

Another way of evading the reflection problem could be by making use of a particular estimation method. Due to the endogeneity of Y, which is partially caused by the issue of simultaneity, spatial models cannot be estimated by OLS. However, there are multiple other ways of estimating spatial models that work under endogeneity. Most prominent are maximum likelihood, generalised spatial two-stage least-square approach (GS2SLS), and Bayesian estimation. Because of the context of the supplementary grant I only use the GS2SLS approach, as I will explain in section 4.2. Kelejian and Prucha (1998) developed the GS2SLS approach, which uses WX and X as instruments of WY, to solve the endogeneity of Y. In the spatial literature, not much attention has been paid to the potential of GS2SLS to evade the reflection problem, the issue of simultaneity, nor on the independence, exclusion restriction and relevancy restriction (Gibbons and Overman, 2012).

[5] I name Lee, Liu and Lin as spatial modellers, however they are among the first authors to apply spatial techniques in a peer effects setting, and seem to be well aware of both traditional and spatial peer effects research.

3.2.2 Correlated effects and selection bias

Correlated effects are not a nuisance parameter in the GNS. Instead the GNS explicitly models a spatial relationship with parameter λ. So, you do not need fixed effects to account for the correlated effects.

Looking at selection bias, spatial modellers are likely to assume the specified network is exogenous, and as a result less likely to adjust for selection effects. Leenders (2002) and Corrado and Fingleton (2012) have argued that more attention should be paid to how the W-matrix is constructed, and that group formation is not always exogenous. The exogeneity of the W-matrix is a topic that has garnered quite some interest recently. For example, Qu and Lee (2015) and Qu, Lee, and C. Yang (2021) have developed new methods to handle self selecting into peer groups. Their methods explicitly model the W-matrix as endogenous. However, these methods are still being developed, and are not often implemented yet.

Another way of correcting for the endogeneity in W is by using fixed effects. Abbiati and Pratschke (2021) have said that fixed effects are necessary to combat selection bias in a classroom setting. Amongst others', papers by Lee (2007), Lin (2010) and Burridge et al. (2016) include group fixed effects, at the level of the peer group g, leading to (11).

$$Y_g = 1_{n_g}\alpha_g + \beta W_g Y_G + X_g\zeta + W_g X_g \gamma + U_g, \qquad U_g = \lambda W_g U_g + \epsilon_g \tag{11}$$

In which α_g are the fixed effects per group g and 1_{n_g} is a n_gx1 vector of ones. The rest of the parameters is similar to (10).

There is still an ongoing discussion in the literature on how to, and whether to include group fixed effects (GFE). Only a limited number of spatial papers make use of fixed effects. Some researchers, like the aforementioned Lee and Lin and Burridge et al. (2016), use a within-transformation. Others, like Baltagi, Deng, et al. (2018) and Zhao and Qu (2021) use dummy variables as fixed effects, which is an option if there are a limited number of groups.

Historically, most spatial researchers have not implemented fixed effects in a classroom setting. This is in large part due to spatial packages in MATLAB, R, Python and Stata not having an implementation for fixed effects if group sizes are varied.[6] This makes applying a within-transformation

[6] I choose to compare what is most commonly done in the traditional approach with what is most commonly done in the spatial approach. I therefore make use of the currently available software packages for both approaches. I discuss this further in section 8.

difficult, while including dummy variables can be problematic due to the incidental parameter problem. Another implementation issue for fixed effects occurs specifically when using GS2SLS in a classroom setting, see section 4.2.

Even when fixed effects can be included via a within-transformation, there are still some issues to worry about. The identification of the GNS worsens when group fixed effects are included via a within-transformation. Burridge et al. (2016) show that it is often a bad idea to include both WX and GFE, as this can introduce multicollinearity between the demeaned versions of X and WX (p. 231-234). The multicollinearity is caused by the combination of the way W is constructed (row-normalised contiguity matrix), the variation in group sizes, and how group fixed effects are accounted for (via demeaning).[7] Choosing whether to include WX or GFE should depend on the context. WX can give more information on specific variables, while GFE can account for more factors that influence groups. But the within-transformation does not allow you to see the size of the group fixed effects.

3.2.3 Model specification

As mentioned, GNS has mostly been seen as a theoretical possibility. Models that estimate the full range of parameters are scarce. Oftentimes, although the GNS is identified in theory, in practice it is only weakly identified (Cook et al., 2015; Burridge et al., 2016). This leads to unreliable results. Other spatial models can be used, depending on which assumptions are made about the various parameters (see figure 1)

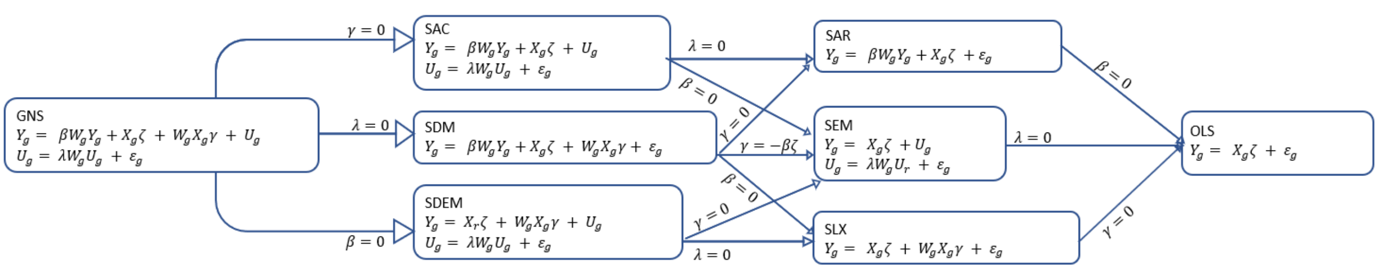

Figure 1: The relationships between different Spatial Dependence Models for cross-sectional data. Source: Adapted from Burridge et al. (2016)

How do you choose which specification is most fitting? Anselin (2017) and Elhorst (2010), have

[7] Alternatively, fixed effects at a more aggregate level can be included, which would lessen the correlation between the demeaned versions of X and WX.

both developed routines to formally test different models against each other. These routines involve LM-tests (Anselin et al., 1996). However, *"these diagnostic tests have power against incorrect alternatives (testing rejects A in favor of B, when, in fact, C is present and causes the rejection, not B), making it difficult to statistically distinguish'between these various models."* (Cook et al., 2015, p. 731).

In practice, many researchers argue from theory which peer effects they think play a role in their specific cases. Recent papers (with binary outcome variables) like Baltagi, Deng, et al. (2018) and Zhao and Qu (2021) have for example argued that exogenous effects are unlikely to be present in their research. They therefore estimate SAR models.[8] If researchers additionally suspect that errors cluster together, Cook et al. (2015) advise researchers to estimate a combined spatial autocorrelation model (SAC). SAR and SAC, although the equations are similar, can lead to quite different coefficients if unobservables are clustered.

To relate to the uptake of the supplementary grant in MBO, endogenous effects are suspected to be most prevalent and are also most relevant for policy. This is why most spatial research would opt to either use a SAR-model or a SAC-model. The accompanying SAR-model is:

$$Y_g = \beta W_g Y_g + X_g \zeta + \epsilon_g \tag{12}$$

While the SAC-model is:

$$Y_g = \beta W_g Y_g + X_g \zeta + U_g, \qquad U_g = \lambda W_g U_g + \epsilon_g \tag{13}$$

These models do not include exogenous peer effects in addition to endogenous peer effects, which would be a spatial Durbin specification (SDM). SDMs are not commonly used to estimate peer effects, SAR or SAC specifications are much more prominent.[9]

3.3 Comparison

First off, both approaches usually argue from theory whether endogenous or exogenous peer effects are most prevalent for a specific research problem. For the supplementary grant, endogenous peer effects are theoretically most interesting. In the rest of the thesis, I therefore compare a SAR-model, SAC-model and IV-model with $\bar{x}_{-ig}$ as an instrument for $\bar{y}_{-ig}$. You could also estimate

[8] Notice that in this case, correlated effects again become a nuisance instead of an explicit part of the model.

[9] A notable exception is Ajilore (2015) who uses an SDM to look the role of peers on risky sexual behaviour.

an IV-model that only uses some leave-one-out variables as an instrument as in (8). I include an estimation of this IV-model in the empirical application in section 6. However, this IV-model is less useful in showing the similarities and differences between the spatial and traditional approach, and is therefore excluded from the simulation study.

All of these models are reduced form models. This means that if exogenous peer effects are also present in practice, the estimations of $\hat{\beta}$ will be different from the true β. This holds for both spatial and traditional models. Spatial modellers often use reduced form regressions, without calling them reduced form, while traditional modellers are more aware of the possibility that exogenous peer effects can occur besides the studied endogenous peer effects.

It is important to notice that the traditional IV-model with x and $\bar{x}_{-ig}$ as instruments for $\bar{y}_{-ig}$ and the SAR-model lead to the exact same coefficients for β. Specifically, they are the identical in a classroom setting, when the spatial model is estimated with GS2SLS and the traditional model with 2SLS. Although the estimations are the same, the traditional and spatial approach differ in the way they view the reflection problem, endogeneity and selection bias and in how they have tackled these issues over the years.

The main reason that these two models are the same rests in the W-matrix. In a classroom setting, the combination of spatial weight matrix with the outcome variable WY is no different from $\bar{y}_{-i}$ in the traditional approach. W is a block diagonal matrix, with $n_g - 1$ observations that have value $\frac{1}{n_g-1}$, and zeros on the diagonal in each block for group g. WY is therefore just taking the leave-one-out average of Y.[10]

Furthermore, estimating SAR with GS2SLS, using WX and X as instruments to estimate the SAR-model, is exactly the same as using the instrumental variable approach with $\bar{x}_{-ig}$ and x_{ig} as instruments for $\bar{y}_{-ig}$. So whatever is a problem for the traditional model should also be a problem for the SAR model, and vice versa.

The SAC-model has the potential to lead to different coefficients for β than both the SAR- and IV-model. SAC explicitly puts a structure on the error term, modelling correlated peer effects. For instance, in Cook et al. (2015) the SAC-model led to different outcomes when errors were correlated at the group level, and the regressors also shared a common group part. However, when errors and regressors are exogenous, SAC should also lead to the same estimated values for β, and thereby not

[10]Other researchers have noticed this as well (o.a. Burridge et al., 2016; Hinke et al., 2019), but the link between the two literatures is often not made this explicitly.

differ from the traditional model nor from SAR.

3.3.1 Reflection problem, endogeneity and selection bias

Both approaches pay quite a lot of attention to solving the reflection problem, but there is no unified way of thinking about this issue. Spatial modellers for example never discuss the potential of GS2SLS to solve the reflection problem, while 2SLS is often mentioned as the preferred solution for traditional modellers. Traditional modellers on the other hand never mention that the reflection problem is solved by using leave-one-out strategies, but always refer to instrumental variables or lagged dependent variables as a solution (Paloyo, 2020; Yeung and Nguyen-Hoang, 2016).[11] If spatial modellers are right, and leave-one-out or leave-one-out in combination with varying group sizes is sufficient to solve the reflection problem, this is something that would be great news for traditional modellers.

Leave-one-out is, however, not enough to solve the endogeneity of Y, that is why in spatial econometrics GS2SLS, ML or Bayesian methods are employed. Instrumental variables and lagged dependent variables in the traditional approach might not be necessary to solve the reflection problem, they do have the added advantage of solving the endogeneity of Y. The endogeneity is mainly caused by simultaneity. Instruments/lagged variables are good solutions to simultaneity issues, if the instruments are exogenous variables that are determined before the outcome variable occurs. Spatial modellers do not specifically mention that endogeneity is partially caused by simultaneity, although GS2SLS is of course a solution to this issue, in the same manner as 2SLS is.

Furthermore, while both approaches frequently make use of GS2SLS or 2SLS, traditional modellers spend much more time arguing whether the assumptions for a valid instrument are met. Questions about the validity of instruments are prominent (Paloyo, 2020). Spatial modellers are prone to using GS2SLS without arguing for the exclusion restriction, independence and sufficient correlation. This is something that for instance Gibbons and Overman (2012) criticises the spatial approach for. Instead, spatial researchers often focus on other restrictions for X and the instrument matrix $H = [X \;\; WX]$, such as they should have full column rank, and the $lim_{n \to \infty} n^{-1} H_n' H_n$ should be finite and non-singular (Kelejian and Prucha, 2010). This is despite the fact, that Kelejian and Prucha (2010) also required H to be non-stochastic matrix, which is not possible if the exclusion

[11] Even in spatial modelling, there is only a minority of researchers that emphasize that leave-one-out is enough (Elhorst et al., 2019).

restriction is violated. For example Baltagi, Deng, et al. (2018) and Zhao and Qu (2021), have recently used GS2SLS without looking at whether the instruments were strong and the exclusion restriction held. Popular software packages in for example R, such as sphet and spatialreg, do not even show results from the first stage of the regressions, so tests on the validity of the instruments are not automatically performed.

Considering selection bias, both approaches agree that one way of mitigating this issue is by using fixed effects. Spatial modellers have also recently been developing methods of explicitly modelling W as endogenous, something that is not possible for traditional modellers (Qu and Lee, 2015; Qu, Lee, and C. Yang, 2021). For fixed effects, both approaches want to use either dummy variables or a within-transformation. But, there is no agreement about which level of aggregation the fixed effects should pertain to. Usually, spatial modellers mention fixed effects at the level of the peer group, like classrooms, while traditional modellers commonly use fixed effects at a more aggregate level, like schools. Spatial modellers often do not argue which level is appropriate. However, if the assumption that traditional modellers are right, and selection bias can be accounted for by including fixed effects at a higher level, this should also hold for the spatial approach. A disadvantage of the spatial approach is that there is still an ongoing discussion on the use of fixed effects, and whether including WX or fixed effects is more informative. In addition, spatial modellers are phased with implementation issues. Currently widely available software packages do not offer a within-transformation if group sizes vary. Dummy variables cannot be used if there are too many groups, due to the incidental parameter problem. As a consequence of the implementation issues, fixed effects specifications are uncommon in spatial practice.

3.3.2 Limitations for both approaches

There are some additional limitations that both approaches have in common. Firstly, both approaches generally employ linear-in-means models, but sometimes these results are not that interesting (Sacerdote, 2011). This is partially due to peer effects not being linear-in-means in real life. Students are affected more by some students than by others, and complementarities can occur. Neither approach is perfectly suited to tackle this issue. Furthermore, from a social welfare point of view, averages can be uninteresting, as we are usually more interested in groups that are disadvantaged, than in the average.

Secondly, both approaches suffer if the peer group is not correctly specified. Traditional modellers often use classes, neighbours, siblings or college roommates, without looking how strong these peer networks are. If for example classes become really large, it is unlikely that everyone interacts equally. In that case, the found peer effects are driven by the subgroup of students that actually influences each other. The W-matrix in the spatial approach allows for more flexibility, as long as there is some information about interactions in the network. Unfortunately, this is often not the case in classroom settings.

Lastly, in the last couple of years, traditional modellers have found a new issue that plagues researchers of peer effects: measurement error/exclusion bias (Angrist, 2014; Caeyers and Fafchamps, 2016). Exclusion bias occurs when fixed effects are included at the level of the selection pools. Person i cannot be her own peer, so the exclusion of i from the pool creates a negative relationship between i's characteristics and that of her peers. *"If i is above average, the average of those remaining in the pool is lower than i. After netting out the pool average via fixed effects, this implies that i's characteristics are negatively correlated with the expected value of the remaining peers in the pool. This is true irrespective of whether peers self-select each other or peers are randomly assigned." (p.3).* Exclusion bias is likely to bias estimates of peer effects downwards. Caeyers and Fafchamps (2016) offer solutions to the issue of exclusion bias, but these are not (yet) commonly implemented. Spatial modellers have not even discussed the potential of exclusion bias, possibly because in practice fixed effects specifications are still quite rare. But, if fixed effects are included in a spatial model, the same issue should be present as well.

4 Methods for binary outcomes

The previous focus was on continuous outcome variables. Using a binary outcome variable is quite different, as for example assumptions such as normality and homoskedasticity no longer hold. In this section I discuss how the traditonal and spatial approaches handle binary dependent variables.

4.1 Traditional approach

4.1.1 Logit & probit models

There are models, such as logit and probit, that explicitly take into account the binary nature of the dependent variable. This means that predicted outcomes always land between 0 and 1. Values above 1 or below zero do not occur, as could happen in a linear probability model. Logit and probit models can also be used to model peer effects (Brock and Durlauf, 2007; An, 2011; Epple and Romano, 2011). However, the question of how to incorporate IV into a logit model has yet to be answered conclusively. IV-probit models to estimate peer effects exist, and have for example been used by McVicar (2011) and recently Niu et al. (2022). There are however disadvantages to using probit models. For example group-fixed effects cannot be used in probit models, unless the groups are included as dummy variables, as a within-transformation does not lead to consistent estimates of the coefficients.

4.1.2 Linear probability models

Traditional modellers of peer effects usually opt to estimate linear probability models (LPMs) instead of logit or probit models (o.a. Markussen and Røed, 2015; An, 2015; Markussen and Røed, 2017; Brenøe and Zölitz, 2020). They often follow argumentation by Angrist and Pischke (2009) that linear methods capture effects regardless of whether the dependent variable is binary, non-negative, or continuously distributed on the real line, as long as the sample size is large enough. Average effects are especially well captured by linear models. Or researchers argue that they do not want to impose an unjustified (complicated) functional form restriction (Markussen and Røed, 2017). An (2015) mentions that any deviation from the model specification in the non-linear models could lead to dramatically biased estimates. In this thesis I also opt to estimate a linear probability model for the traditional approach.

Although LPMs can adequately estimate peer effects, they still need to be adjusted for heteroskedasticity. This is usually done via robust standard errors. Sometimes clustered standard errors are employed. Clustered standard errors allow for correlations in the outcomes of students within each school/class/start year cluster (Feld and Zölitz, 2017). However, Abadie et al. (2017) argue that you should only include clustered standard errors if you know that either sampling or assignment to treatment was clustered. The empirical application includes the entire student population in MBO-3 and MBO-4 in 2020/2021, so there is no clustered sampling. Furthermore, no assignment to treatment takes place. Therefore I prefer including only heteroskedasticity robust standard errors over clustered standard errors.

4.2 Spatial approach

4.2.1 Spatial linear probability models

To estimate a spatial linear probability model, heteroskedasticity and non-normality are important issues.[12] Most spatial modellers prefer estimating coefficients via (quasi)-maximum likelihood as explained in Lee (2004). However, (Q)ML does not work optimally with a binary outcome variable. It assumes that disturbances are normally distributed (Kelejian and Prucha, 1998) and it needs homoskedasticity (Lin and Lee, 2010). Another option is to use Bayesian estimation (LeSage, 2000). However that is computationally not feasible with a large data set, as it requires inversion of an NxN matrix in every step (Baltagi, Deng, et al., 2018). GS2SLS as proposed by Kelejian and Prucha (2010) does not need normality and they have adjusted the method to take heteroskedasticity into account. Furthermore, it is faster than the Bayesian method. Recent studies by Abbiati and Pratschke (2021), Baltagi, Deng, et al. (2018) and Zhao and Qu (2021) all use GS2SLS to estimate their spatial linear probability models with binary outcome variables. GS2SLS uses X, WX and sometimes WWX as instruments for WY.

So, GS2SLS as in Kelejian and Prucha (2010) works well under non-normality and heteroskedasticity. However, using GS2SLS with fixed effects proves to be a problem. As mentioned, there are no widely available software implementations to estimate a regression with fixed effects via a within-transformation if group sizes vary. The alternative is to use dummy variables, which should be possible as long as the number of groups is limited. Unfortunately, here another limitation occurs. The current implementations of heteroskedasticity robust GS2SLS in R use X, WX and WWX as instruments for WY, where X also includes columns for the dummy variables.[13] The issue in a classroom setting is that everyone in the same class also has the same value for the school fixed effects dummies. Premultiplying X with the block matrix W and WW leads to the same values in the columns corresponding to the dummy variables in both WX and WWX.[14] Therefore, due to classroom setting, perfect multicollinearity between certain columns in the instruments occurs.

The multicollinearity problem could be solved by only taking WX as an instrument excluding WWX. Unfortunately, this is not an option for the most current software packages, at least in

[12]For instance, the model specification tests mentioned in section 3.2 usually require homoskedasticity. Baltagi and Z. Yang (2013) developed LM tests that work under heteroskedasticity, but these are not commonly implemented.

[13]See the source code for the R package sphet at https://rdrr.io/cran/sphet/src/R/hidden_functions.R

[14]See the appendix for a numerical example of this multicollinearity.

R. Another solution is to include fixed effects that are different for individuals within each block. For example, Baltagi, Deng, et al. (2018) looks at migrant networks within cities and includes fixed effects not based on current city but on city of origin. However, if you want to use fixed effects to combat selection bias at a class or school level, that is not an option.

4.2.2 Spatial logit models

Recently, Baltagi, Deng, et al. (2018) and Zhao and Qu (2021) have estimated spatial logit models, alongside LPMs. Spatial logit models are mostly regarded as a viable, albeit quite complicated option. Complications arise from the interpretation of the spatial coefficients in the logit model. In the linear setting, it is already not straight-forward as for example deviations in someone's x-variables cause shifts in their y, which cause shifts in the y's of their peers, thereby leading to additional shifts in their own y. These direct and indirect impacts need to be calculated, and cannot be directly read from a table. The logit model complicates this even further.

An additional complication of spatial logit models is that there are different estimators available. The choice of the most appropriate one is not straightforward. Calabrese and Elkink (2014) study which type of estimator performs best in which circumstance. They compare six alternatives: the Expectation-Maximization Algorithm by McMillen (1992), The Bayesian Gibbs sampler approach by LeSage (2000), Recursive Importance Sampling by Beron and Vijverberg (2004), GMM by Pinkse and Slade (1998) and Linearised GMM by Klier and McMillen (2008). According to Calabrese and Elkink (2014) the linearised GMM approach is really accurate in large samples as long as spatial correlation is small (≤ 0.45). Furthermore, *"It is clear that the choice of estimator has significant impact on the computational requirements, with the LGMM estimator being by far the fastest."* It also seems to be the preferred approach in recent years, with again Baltagi, Deng, et al. (2018) and Zhao and Qu (2021) using it. LGMM seems most appropriate for the context of the supplementary grant, as the data set in academic year 2020/2021 spans more than 250,000 observations, which is much larger than most data sets in spatial research.

LGMM for SAR-logit works as follows. It is assumed that there is some underlying process that determines someones intention to perform a certain action. However we do not observe this intention Y_g^*.

$$Y_g^* = 1_{n_r}\alpha_r + \beta W_g Y_g^* + X_g\zeta + \epsilon_g \tag{14}$$

Instead we only see the binary outcome that results from the intention, $Y_g = 1(Y_g^* > 0)$. Following

Klier and McMillen (2008) the first step of LGMM is to estimate a standard logit model of Y_g on X_g, receiving $\hat{\zeta}$ and a probability estimate $\hat{P}_g = \frac{exp(X_g\hat{\zeta})}{1+exp(X_g\hat{\zeta})}$. Additionally, you need to calculate the residuals $\hat{\epsilon}_g = Y_g - \hat{P}_g$ and the gradients with respect to ζ and β: $G_\zeta = \hat{P}_g(1-\hat{P}_g)X_g$ and $G_\beta = \hat{P}_g(1-\hat{P}_g)WX_g$. The second step is the two-step least squares regression. First stage: regress G_ζ and G_β on the set of instruments Z, where $Z = (X, WX)$ to receive predicted values: $\hat{G}_\zeta$ and $\hat{G}_\beta$. Second stage: regress $\hat{\epsilon}_g + G_\zeta\hat{\zeta}$ on $\hat{G}_\zeta$ and $\hat{G}_\beta$. The coefficient for $\hat{G}_\zeta$ corresponds to an estimate for ζ, and similarly for $\hat{G}_\beta$ and β. LGMM simplifies estimation by linearising the estimates around the convenient starting point $\beta = 0$ and the initial standard logit estimates of ζ.

For the spatial logit model, including group fixed effects via any sort of transformation proves to be a challenge that has not even been studied in theory. Dummies can however be included, as long as only X and WX are used as instruments, excluding WWX. Estimating a SAC-logit model does not seem to be common, instead I opt to just use the fixed effects dummies to partially filter the spatially correlated errors. I therefore use LGMM to estimate a SAR-logit model, which has dummies at the school level.

4.3 Comparison

The previously discussed similarities and differences for continuous models still hold for binary outcomes. But what are the additional similarities and differences when looking at binary outcomes?

Both the spatial and the traditional approach have found ways to overcome heteroskedasticity and non-normality to estimate LPMs. The traditional approach does so by using 2SLS and heteroskedasticity robust standard errors. The spatial approach uses GS2SLS, also in combination with robust standard errors. These approaches are again indistinguishable for SAR-models and IV-models if the same instruments are chosen.

A limitation for spatial modellers stems from including fixed effects, especially in an LPM. Traditional modellers can use within-transformations to deal with fixed effects, even under heterskedasticity and non-normality. Spatial software packages to estimate LPMS using heteroskedasticity-robust GS2SLS cannot perform within-transformations when group sizes vary. Including dummy variables as fixed effects is currently not possible for the spatial LPM as current software implementations use both WX and WWX as instruments, leading to multicollinearity in a classroom setting. Fortunately, including fixed effects dummies is possible for the SAR-logit model.

Spatial modellers have a longer history and more developed literature on logit and probit models.[15] Many different spatial logit estimation approaches have been developed over the years, whereas there is no common way of implementing IV-logit. Especially LGMM to look at spatial logit models seems promising, as it works relatively well for large data sets. Spatial logit can also count on increased interest in the last ten years (Calabrese and Elkink, 2014; Lee, Li, et al., 2014). Nevertheless, both the spatial approach and the traditional approach currently favour estimating LPMs over logit/probit models.

5 Simulation study

To recap, traditional modellers and spatial modellers have been researching peer effects quite extensively over the last few decades. There are clear similarities between the approaches, and in broad strokes both modellers are faced with the same difficulties and have come up with similar solutions. However, the two approaches have not come up with a unified way of looking at these issues. Additionally, there are some issues in which spatial modellers have come up with better solutions, like a way to estimate a spatial logit model. For other issues, such as ways to include fixed effects, the traditional modellers have the upper hand.

In the coming section, I first estimate different types of traditional and spatial models to show the similarities in the most basic classroom setting with continuous outcomes. Then I complicate the setting slightly, by allowing for error terms to be correlated across aggregate groups. This is where theoretically an IV model with fixed effects and a SAC model should outperform a classic IV and SAR model. The final extension of the simulation is to specifically compare the approaches in a setting with binary outcome variables.

This simulation study serves as a preparation before evaluating the role of peer effects for the uptake of the supplementary grant. Traditional models have not been empirically compared to spatial models before, so it is important to see how their performances differ. For this purpose I look at how often significant peer effects are found in a controlled environment in which endogenous peer effects are present, the proportion of times the true value of β lies within the 95% confidence interval, as well as for example computation time. The models that seem most promising in the

[15] Traditional modellers have taken notice of the spatial literature as well, for example Boucher and Bramoullé (2020) try to bridge the gap between both methods, referring to seminal papers in both literatures. They want to propose a linear framework to deal with binary variables, that is inspired by both approaches.

simulation study are used in section 6.

5.1 Monte Carlo setup

The simulation study is inspired by the set-up in Calabrese and Elkink (2014), combined with the set-up in Cook et al. (2015), adjusted for a classroom setting. The Monte Carlo set-up includes a 1,000 replications per data generating process (DGP), with a sample size of $N = 500$ in every run.

Different DGPs are used to study the spatial and traditional approach. For each DGP two different levels of endogenous peer effects are investigated: $\beta_0 = \{0.15, 0.45\}$, while exogenous peer effects $W_g X_g$ are set to 0. Furthermore, in all DGPs a W-matrix is used. Unlike in Calabrese and Elkink (2014) and Cook et al. (2015), my W corresponds to a classroom setting. First, all 500 observations in a replication are randomly assigned to 25 groups. These groups are then randomly assigned to three different aggregate groups. This corresponds to the idea of students belonging to 25 different classes, which are a part of three different schools. Groups of size 1 are deleted from the data set, as in practice students without peers are also not taken into account. Then W is constructed, with all students that belong to the same class receiving a 1, zeros otherwise and a zero diagonal. Next, W is row-normalised.

For each DGP, the performance of the appropriate spatial and traditional models is looked at. As mentioned, I am mainly interested in how well the endogenous peer effects are estimated, in how much the models differ in their estimation and in computation time. First off, I show scaled density plots to visually illustrate the difference in the estimated values for beta in each model m: $\hat{\beta}_m$ between the models. Secondly, I look at the average $\hat{\beta}_m$ for each model m. Besides that, I evaluate how often a model rejects the null-hypothesis of no endogenous peer-effects, and compare this between the models. Furthermore, I perform t-tests with $\alpha = 0.05$ to see how often the (true) null-hypothesis of $\hat{\beta}_m = \beta_0$ is rejected. In the same vein, the coverage probability for each model is reported. Lastly, I record the average run time in seconds.

5.2 DGP-I

DGP I is the following:

$$Y_g = \beta W_g Y_g + X_g \zeta' + U_g \tag{15}$$

Following Cook et al. (2015), the outcome variable is continuous, but otherwise the set-up is similar to Calabrese and Elkink (2014). I use an intercept and one covariate X drawn from a normal distribution with expected value 2 and standard deviation 4. ζ' is set as $[4, -2]$. The error term is generated from a standard normal distribution, and is completely exogenous and homoskedastic.

The models most appropriate for this DGP are IV and SAR. SAC should also be able to consistently estimate β. The formula for SAC is:

$$Y_g = \beta W_g Y_g + X_g \zeta + U_g, \qquad U_g = \lambda W_g U_g + \epsilon_g \tag{16}$$

To the detriment of SAC, although $\hat{\lambda}$ should not be significantly different from 0, there sometimes is a loss of significance for $\hat{\beta}_{SAC}$ when λ is included in the regression (Cook et al., 2015).

Because the residuals stemming from this DGP fit the normality and homoskedasticity assumptions, (Q)MLE could be used for SAR and SAC. However, as I am most interested in using GS2SLS to look at the supplementary grant, I use GS2SLS here as well. It is expected that IV and SAR are nearly indistinguishable when GS2SLS and 2SLS are used, both in the coefficients and the corresponding standard errors. Small differences are expected as the implementation of SAR uses both WX and WWX as instruments.

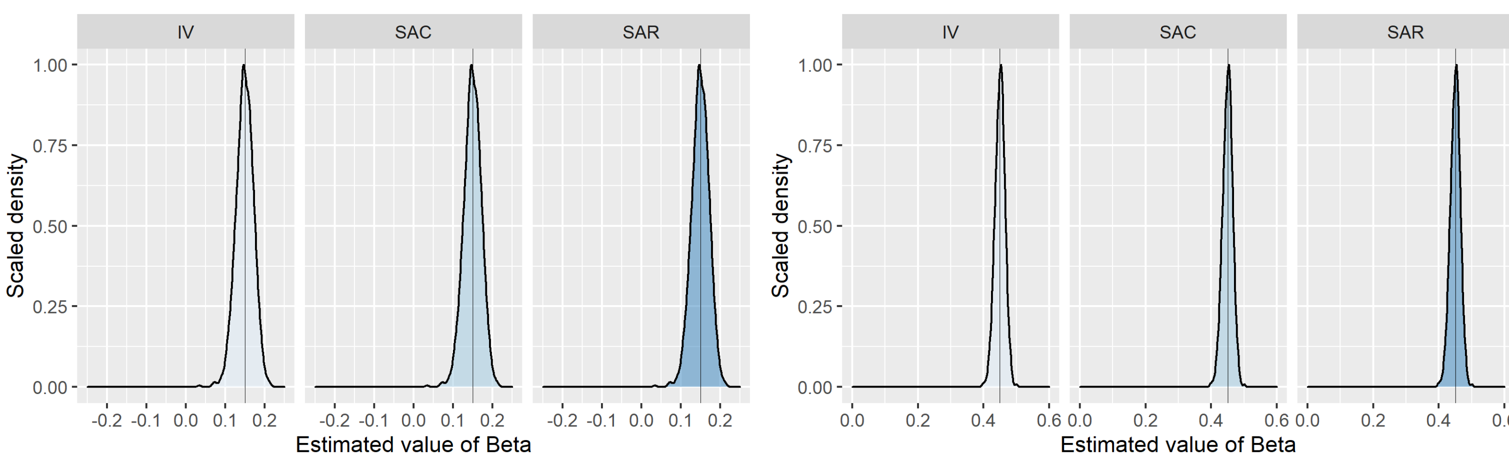

Figure 2: Scaled density for DGP-I, $\beta = 0.15$ left and $\beta = 0.45$ right

From figure 2, it seems that all models are very well able to estimate β in this set-up. The distribution shows a high peak around the true value β_0, for both levels of β_0. The peak might seem more concentrated for the larger peer effects, but is mostly due to the scaling of the figures.

Table 1: DGP-I: Exogenous innovations and continuous outcome variable

	IV	SAR	SAC	IV	SAR	SAC
		$\beta_0 = 0.15$			$\beta_0 = 0.45$	
Estimated β_m	0.1490	0.1490	0.1490	0.4497	0.4497	0.4497
Significant peer effects	998	998	998	1000	1000	1000
Reject $\hat{\beta}_m = \beta_0$	49	47	71	54	53	80
Coverage probability	95.1%	95.3%	92.9%	94.6%	94.7%	92.0%
Significant λ in SAC model			65			52
Average run time (s)	0.01	0.05	0.20	0.02	0.05	0.19

Performance of three different models for two levels of β_0 over 1000 repetitions.

Table 1 shows that indeed, all models can consistently estimate the true endogenous effects. And in over 99% of the runs, the existence of significant peer effects is captured by all models. SAC rejects the null-hypothesis of the estimated peer effect effect being equal to the true effect in 71 and 80 cases. This makes the coverage probability drop below 95%. Additionally, SAC finds significant correlated effects in 65 and 52 cases. This can be seen as evidence that indeed when only endogenous effects are present but both endogenous and correlated effects are estimated, these estimates battle for significance.

The IV and SAR model are practically indistinguishable, as was theorised in section 3.3. These models reject the null-hypothesis of the estimated effect being equal to the true effect about 5% of the time. The coverage probability is also around 95% for both β's, which is consistent with an α of 0.05. The differences of 1 and 2 in rejection of $\hat{\beta}_m = \beta_0$ for IV and SAR are probably because the implementation for SAR also includes WWX as an instrument. Even so, they are nearly indistinguishable. The difference in $\hat{\beta}_{SAR}$ and $\hat{\beta}_{IV}$ and the difference between the corresponding standard errors is never larger than 0.001 in any of the 1000 runs when the peer effects are small. In fact the maximum difference in $\hat{\beta}$ is 0.0007, when $\beta_0 = 0.15$. For the larger peer effects the maximum difference is 0.002. The traditional and spatial approach are therefore indeed estimating the same coefficients in the same manner, when SAR and IV are used.

The run time for SAC is much longer than for IV and SAR. Longer run times combined with a slightly worse performance, means that SAC is not the preferred model for this DGP. IV is quicker than SAR, while giving the same results. Using IV is therefore advised in this set-up.

5.3 DGP-II

In DGP-II the set-up is similar to DGP-I, except now the error corresponding to person i is partially correlated at the aggregate level. Allowing for correlated error terms is something that is also done in Cook et al. (2015).

$$Y_g = \beta W_g Y_g + X_g \zeta + U_g, \qquad u_{igr} = \epsilon_{igr} + \epsilon_{gr} \tag{17}$$

The first part of u_{igr} is the individual error, which differs for each individual i, who is a member of class g in school r. It is generated from a standard normal distribution. The second part is the same for each member of aggregate school r, and is also generated from a standard normal distribution.[16]

The most appropriate models in this case are IV with fixed effects (IV-fe) and SAC. IV-fe can explicitly take the correlated error terms into account, with group dummies at the aggregate level or fixed effects. As there are only three schools in the simulation, adding dummies is quite easy. Notice that SAC always looks at correlation at the class level, while the errors are actually correlated at the school level. SAC is therefore likely to perform slightly worse than IV-fe. I also include a SAR-model and IV-model, to show that these specifications are less appropriate.

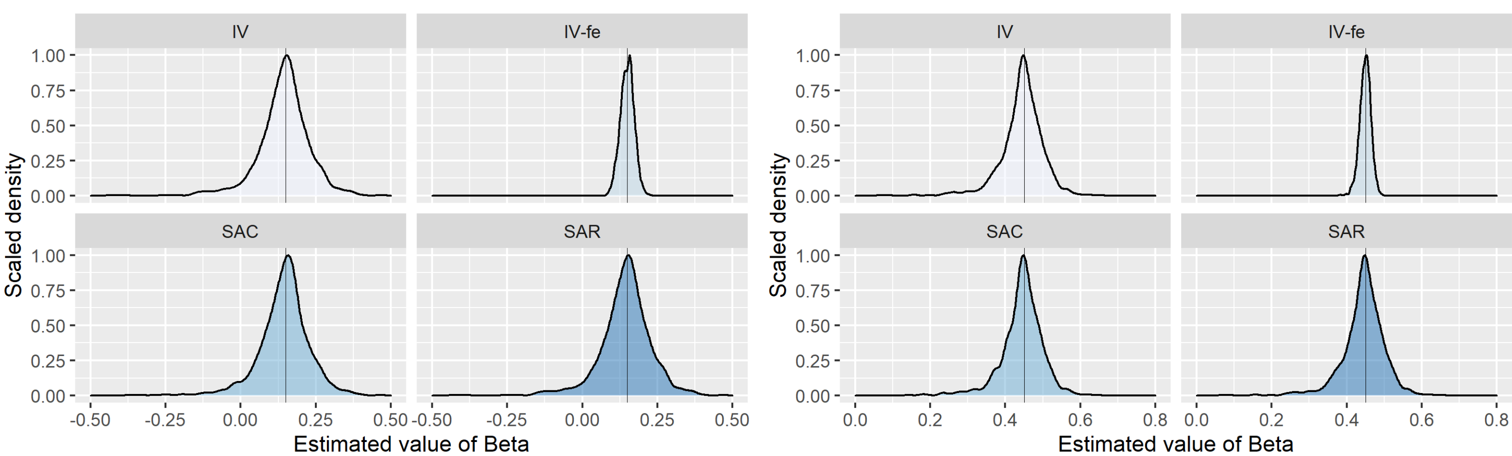

Figure 3: Scaled density for DGP-II, $\beta = 0.15$ left and $\beta = 0.45$ right

[16]This means that the standard deviation of the combined error term is twice as large as for the error in DGP-I.

Table 2: DGP-II: Partially correlated innovations and continuous outcome variable

	IV	IV-fe	SAR	SAC	IV	IV-fe	SAR	SAC
	$\beta_0 = 0.15$				$\beta_0 = 0.45$			
Estimated β_m	0.1425	0.1491	0.1425	0.1435	0.4443	0.4492	0.4443	0.4445
Significant peer effects	890	1000	890	621	998	1000	998	988
Reject $\hat{\beta}_m = \beta_0$	420	42	420	72	423	49	423	74
Coverage probability	58.0%	95.8%	58.0%	92.8%	57.7%	95.1%	57.7%	92.6%
Significant λ in SAC model				931				939
Significant school dummies		984				987		
Average run time (s)	0.01	0.01	0.04	0.17	0.01	0.01	0.04	0.17

Performance of four different models for two levels of β_0 over 1000 repetitions.

Figure 3 looks quite different from DGP-I, except for IV-fe. The mean of the distribution of β_m for all models is still quite close to the true β_0, meaning that the models only have a very minor bias. In contrast to the previous DGP, the distributions for all models except IV-fe are much more spread out. Indeed, this is also confirmed by table 2, in which we can see that IV and SAR reject $\hat{\beta}_m = \beta_0$ in about 420 of 1000 cases. SAC seems to be doing better in that regard. But at the same time, SAC is not able to find significant endogenous peer-effects in 379 cases out of a 1000 when the peer effects are small. This is a worse result than SAR/IV. So failing to reject no endogenous peer effects, but at the same time having a coverage probability of above 92% must mean that in most runs the standard error for $\hat{\beta}_{SAC}$ is high, leading to large confidence intervals.

IV-fe is one of the quickest models, and is also least likely to reject that $\hat{\beta}_{IV-fe} = \beta_0$. In the set-up of DGP-II, the traditional approach with fixed effects seems to be most effective, while again IV and SAR are indistinguishable. Unexpectedly, SAC is not performing much better than SAR/IV when it comes to finding significant peer effects, and SAC suffers from large standard errors. This is unlike the results of Cook et al. (2015), who found that the SAC model performed much better than SAR under correlated errors. However, they also used correlation between the errors and the x-variables, which is not used in the current set-up. I have chosen not to incorporate this aspect, as it would go against the requirements for good instrumental variables. If X is correlated to U and WX is used as an instrument, WX does not only influence Y through WY but also through

U. This was not a concern for Cook et al., who make use of MLE.

5.4 DGP-III

The third DGP has a binary outcome variable. As in Calabrese and Elkink (2014) it is assumed that there is an underlying process Y_g^*, while we only observe the binary outcome Y_g. My set-up differs from Calabrese and Elkink's, by allowing correlated error terms, and by making use of a W-matrix with a classroom set-up/contiguity $W - matrix$.

$$Y_g^* = \beta W_g Y_g^* + X_g \zeta + u_g, \qquad u_{igr} = \epsilon_{igr} + \epsilon_{gr} \tag{18}$$

As you can see, the underlying tendency Y_g^* is the same as DGP-II. The corresponding binary outcome is constructed as follows: $Y_g = 1$ if $Y_g^* > 0$ and 0 otherwise.

I estimate the same four models as before, but add a SAR-logit model and a SAR-logit model with fixed effects dummies. Based on the previous two DGPs, it is likely that IV-fe outperforms IV/SAR and SAC. Calabrese and Elkink (2014) find the SAR-logit model estimates consistently when endogenous peer effects are small. While at $\beta_0 = 0.8$, the SAR-logit model is likely to drastically overestimate the peer effect. Based on their findings, SAR-logit-fe should work quite well, especially when peer effects are small. SAR-logit is expected to be less accurate, as it cannot take correlated errors into account.

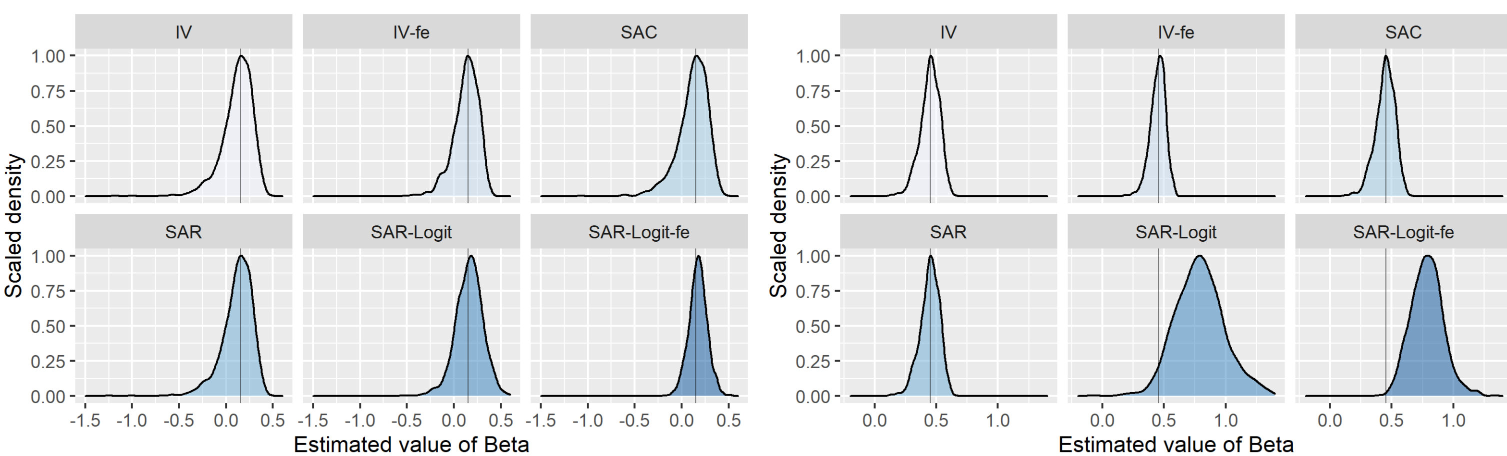

Figure 4: Scaled density for DGP-III, $\beta = 0.15$ left and $\beta = 0.45$ right

This is the first time that there are large differences between a small and a large β, as can be seen from figure 4. SAR-logit-fe seems to be the best performing model when $\beta = 0.15$, but is severely biased when $\beta = 0.45$.

Looking at the left side of table 3, all models are only able to find significant endogenous peer effects in fewer than 60% of the runs. At the same time, their coverage probability is quite high, with the exception being SAR-logit. IV-fe is least likely to find significant endogenous peer effects, while the average $\hat{\beta}_{IV-fe}$ is quite close to the true value β_0 and the coverage probability is 98%. This means that the standard errors of $\hat{\beta}_{IV-fe}$ in each individual run must be quite high.

Table 3: DGP-III:Partially correlated innovations and binary outcome variable

	IV	IV-fe	SAC	SAR	SAR-l	SAR-l-fe	IV	IV-fe	SAC	SAR	SAR-l	SAR-l-fe
			$\beta_0 = 0.15$						$\beta_0 = 0.45$			
Estimated β_m	0.1149	0.1275	0.1147	0.1135	0.1630	0.1730	0.4463	0.4490	0.4451	0.4450	0.7938	0.7863
Significant peer effects	234	197	276	237	396	399	980	991	968	981	993	1000
Reject $\hat{\beta}_m = \beta_0$	63	20	62	64	118	39	13	1	66	15	646	713
Coverage probability	93.7%	98.0%	93.8%	93.6%	88.2%	96.1%	98.7%	99.9%	93.4%	98.5%	35.4%	28.7%
Significant λ in SAC model			104						360			
Significant school dummies		533				605		450				127
Average run time (s)	0.01	0.01	0.12	0.04	0.05	0.05	0.01	0.01	0.12	0.04	0.05	0.05

Performance of six different models for two levels of β_0 over 1000 repetitions.

SAR-logit-fe has one of the highest coverage probabilities, while SAR-logit's is actually the smallest. Both logit models are, howeve,r most likely to find evidence for significant peer effects. It seems that the spatial logit models have the smallest standard errors in each individual run.

Interestingly, SAC does not perform very differently from SAR and IV. This could mean that the correlated part in the error term did not influence Y_g that much, after transforming Y_g^* into Y_g. SAC does continue to be the slowest.

For the larger endogenous peer effects on the right side of table 3, the results are quite different. Here, all models are able to find significant peer effects in more than 95% of the runs. In particular, IV-fe fits extremely well, with a coverage probability of 99.9% and significant peer effects in more than 99% of the runs. IV/SAR also only reject the null-hypothesis that the estimated peer effects are equal to the true peer effects in about 1% of the runs. This is an indication that the influence of the correlated part of the error term on y is vastly overshadowed by the magnitude of the peer effects when $\beta_0 = 0.45$. IV and IV-fe are still the fastest. SAC continues to be the slowest, and actually performs slightly worse than IV/SAR, as it rejects the true null-hypothesis about four times as often. However, the coverage probability is still 93.4%.

But the most striking difference between the two levels of β comes from the logit models. In particular the SAR-logit-fe model. While it was among the best performing models for small peer effects, it now becomes the worst with a coverage probability of only 28.7% and average estimated $\beta_m = 0.7863$. In Calabrese and Elkink (2014), overshooting the true value β_0 only happened for $\beta_0 >= 0.8$, not for smaller βs. Why do the LGMM models overshoot much more than in Calabrese and Elkink's study? It could be due to the correlated errors, which the other study did not implement. Or it could be due to the use of a contiguity W-matrix, which Calabrese and Elkink do not use. As the other models that were likely to be worse under correlated errors (SAR and IV) did not drastically decrease in performance, it is likely that most of the difference is caused by the contiguity W-matrix.

I look deeper into this possibility by means of a small simulation study in figure 5 and table 7 in the Appendix. Indeed, when I use a set-up without correlated errors and a W-matrix as in Calabrese and Elkink (2014) the results are close to what was found in their paper. However, when using the same set-up but swapping the W-matrix for a contiguity $W-$matrix, the results are much closer to what is presented in figure 4 and table 3.

5.5 Conclusion of simulations

In conclusion, SAR and IV (without fixed effects) are nearly indistinguishable in practice, as I also discussed in the comparison of the theory behind the two approaches. The small differences are because SAR uses WWX and WX as instruments, while IV only uses $\bar{x}_{-i}$. This spatial and traditional model lead to the same conclusions in almost all simulations for all DGPs, although the runtime of the IV-model is about four times smaller.

The SAC model can lead to some different results, sometimes more accurate than IV/SAR, as in DGP-II, but sometimes slightly worse, as in DGP-I and DGP-III. Furthermore it is much slower, consistently being between 12 and 20 times slower than IV. For this reason, it is not a very practical model when a data set is large. Besides that, in DGP-II and for large peer effects in DGP-III IV-fe outperformed SAC, making IV-fe the preferred choice. Note that I did not evaluate what would happen if the errors were correlated at the classroom level, instead of the school level. It could be that IV-fe would be outperformed by SAC in that case. However SAC would still be much slower for a large dataset.

The traditional approach does not have an established logit option, while the spatial modellers have SAR-logit. Under small peer effects, SAR-logit-fe was actually the preferred model, even though it also failed to see significant endogenous peer effects in about 60% of the runs. However, when β became relatively large, SAR-logit and SAR-logit-fe vastly overshot the true value of $\beta_0 = 0.45$. While the LGMM estimation of SAR-logit is known to overshoot when endogenous peer effects are large, overshooting happened much quicker in my set-up. This is mostly due to the classroom setting of the W-matrix, which is the type of W-matrix that is also present for the empirical application in the next section.

6 Empirical application

In this section I zoom in on the supplementary grant, and look at whether there are any further complications for the two approaches. I make use of IV-fe, as in the simulations studies it was quickest, consistent or only slightly biased, and often had a relatively high coverage probability. I also make use of SAR-logit-fe, which worked relatively well as long as the true peer-effects were small. Additionally, it provides a nice contrast to compare the spatial and traditional approach for a binary outcome variable.

Most endogenous peer effects found in higher education and for financial decisions (discussed in section 2) are positively correlated with the variable of interest. It is likely that this holds for the uptake of the supplementary grant as well. The more people in your environment that for instance make social security claims, the more likely it is that you will do so too (Markussen and Røed, 2015). And if someone's classmates perform well academically, or perform certain social behaviours like drinking or criminal behaviour, it becomes more likely that this person will act in the same way (Epple and Romano, 2011). The size of the endogenous peer effects in higher education vary quite a lot in different studies, ranging from moderate to large depending on the type of outcome variable (Sacerdote, 2011). All in all, I hypothesise that in the case of the uptake of the supplementary grant in MBO some positive endogenous peer effects occur, but I place no expectations on the size of the effect.

6.1 Data

As mentioned, this research focuses on MBO. Specifically, I look at MBO-3 and MBO-4. These two levels have a comparable student finance system, as the grants and student travel product are performance based (see Appendix for more info on the Dutch student finance system). They also account for the majority of all MBO students. I use cross-sectional microdata on academic year 2020/2021, provided by Statistics Netherlands (CBS). In this time period, I have 266,943 observations. 50.2% of the students are entitled to the supplementary grant.

The Netherlands Bureau for Economic Policy Analysis (CPB) has constructed a proxy to indicate entitlement to the supplementary grant, based on formulas provided by DUO (Zumbuehl and Magnée, 2022). DUO is the state body for the implementation of education, and is responsible for the student finance system. DUO makes use of these formulas to determine whether someone that requests the supplementary grant is eligible for it, and to determine the monthly amount. DUO has provided the CPB with data on which students were granted the supplementary grant. If a student applies at a certain point during the academic year, she will receive the full amount for that year retroactively.

6.2 Results

For illustration purposes, I first look at a relatively small data set. This includes every MBO-3 and MBO-4 student that is in a class with 15-25 students in the school year 2020-2021. Peer effects are often more pronounced when class sizes are smaller, as it is much more likely that students interact. This makes it interesting to first look at this subset. There are 31,966 students in the subset, belonging to 1335 different MBO schools. The SAR-logit model therefore needs to include 1334 dummy variables. In total, 16,319 students in this subset are eligible for the supplementary grant.

The x-variables used in these regressions are: total number of years a student has been in MBO education, a dummy to indicate whether a student is eligible for the supplementary grant at all. And a dummy on whether the student is eligible for the maximum monthly amount. Furthermore, I include uptake of the basic student finance grant, a dummy for whether one or both parents receive(s) welfare benefits, and two dummies on the students personal characteristics. The leave-one-out averages of all these x-variables are used as instruments in the first stage, for both IV-fe and SAR-logit-fe.

Table 4: Regression results IV-fe (1) and SAR-logit-fe.

	Second stage IV-fe (1)	SAR-logit-fe
Leave-one-out average uptake / WY	-0.220*** (0.0407)	-0.0006 (0.0007)
Total years in MBO	0.0121*** (0.0015)	-6.4233*** (0.9783)
Eligibility for supplementary grant	0.403*** (0.0048)	0.1903*** (0.0279)
Fully eligible	0.114*** (0.0061)	1.2981*** (0.0699)
Any basic grant	0.364*** (0.0041)	6.7428*** (0.2213)
Parent on welfare	0.0844*** (0.0053)	0.9189*** (0.080)
Non-Dutch background	0.0556*** (0.0049)	0.8419*** (0.0852)
Female	0.00826 (0.0046)	0.0796 (0.0841)
N	31,966	31,966
Time elapsed (s)	2.298	2954.75

The bottom seven regressors and their leave-one-out averages/WX are used as instruments for leave-one-out average uptake/WY. Standard errors in brackets. *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$

The dummy for whether a student is eligible is highly significant for both regressions. This dummy is necessary, as the regression is run on all students, even the students that are not eligible for a supplementary grant. Ineligible students will of course always have 0 as their value for uptake. I will come back to this issue after looking more closely at the results in table 4.

First off, the time elapsed is much longer for SAR-logit, just like in the simulation. However the difference is much more pronounced for this data set. IV-fe runs in about 2 seconds, while SAR-logit-fe takes about 49 minutes. The main culprit seems to be the size of the W-matrix. Even though W is a sparse matrix, it is still of size 31,966x31,966. This is much larger than the 500x500 matrices in the simulation study, and the R package uses large vectors to store accompanying information.

Looking at the results, we see a very unexpected value for the endogenous peer effects in IV-fe. The peer effects are rather large and negative: -0.220, which is significant even at the 0.1% level. This means that if if you have more classmates that take up the supplementary grant, it becomes the less likely that you take it up as well. This seems very unlikely and is not in accordance with what was hypothesised based on the literature. The SAR-logit-fe model in contrast, does not pick up any significant peer effects. But we know from the simulation study that SAR-logit-fe did not pick up significant peer effects in about 40% of individual runs, even if they were present in the DGP.

These surprising results for IV-fe made me look closer at the requirements for the instrumental variables. IV-fe's STATA software reports statistics about the first stage of the regression. As mentioned in section 3.1, the instrumental variable approach needs instruments to be strongly correlated with the endogenous variable, the instruments need to be independent and the exclusion restriction need to hold. The correlation does not seem to be an issue, as the Kleibergen-Paap rk LM test statistic is 3323.75, which means that the null-hypothesis of no correlation is rejected. The Kleibergen-Paap F-statistic is 4292.91, meaning that the null-hypothesis of weak identification is also rejected.

On the other hand, there still seems to be a problem with the validity of the instruments according to the Hansen J-test. The null-hypothesis is that the overidentifying restrictions are valid. The Hansen J-test statistic for this regression is 23.37, the corresponding p-value is 0.0000. This is a clear rejection of the null-hypothesis. These test statistics lead us to doubt the results that are presented in table 4, for both regressions, as both make use of the same instruments.

For SAR-logit-fe it was not clear that the results were influenced by flawed instruments. Gibbons and Overman (2012) also mention the lack of critical evaluation of instruments as a problem of spatial econometrics. When Klier and McMillen (2008) introduced the LGMM method, they did not mention assumptions necessary for LGMM to work. They did, however, emphasise that the instruments used were all exogenous. As the second step of LGMM is to estimate a GS2SLS regression, we can further look at Kelejian and Prucha (2010) for assumptions under which Gs2SLS works. Again the problem seems to stem from the fact that the instrument matrix H is stochastic, as WX is not exogenous. So while the exclusion restriction is necessary, also for the spatial model, it is not often investigated in practice.

Tabel 5 shows the IV-fe regression again, but now with only some leave-one-out averages of the x-variables as instruments. Particularly, average total years in MBO, migration background and gender are no longer instruments, but included as exogenous peer effects. For the SAR-logit-fe model the implementation in R does not let us choose which x-variables to use for the instruments WX. For this IV-fe regression, again the tests for correlation and weak identification show no problems (LM test statistic of 2479.72 and F test statistic of 2634.58). But this time the Hansen J-test also fails to reject the null-hypothesis, with a p-value of 0.2522, showing that these instruments are most likely a better choice. The endogenous peer effects still seem negative at first glance, -0.0706, but they are not significantly different from zero. These results also give a first indication that perhaps some exogenous peer-effects are present (especially the average total number of years in MBO education of classmates is significant at the 0.1% level). Using these exogenous peer effects as instruments in the first stage is problematic, as they are of direct influence on the uptake, not just through the average uptake.

Table 5: Regression results IV-fe (2)

Second stage IV-fe (2)	
Leave-one-out average uptake	-0.0706 (0.053)
Total years in MBO	0.0124*** (0.0015)
Eligibility for any grant	0.406*** (0.0049)
Fully eligible	0.114*** (0.0061)
Any basic grant	0.368*** (0.0042)
Parent on welfare	0.0841*** (0.0053)
Non-Dutch background	0.0556*** (0.0052)
Female	0.00716 (0.0050)
Leave-one-out average total years in MBO	-0.0270*** (0.0057)
Leave-one-out average percentage non-Dutch background	0.00582 (0.0383)
Leave-one out average percentage female	-0.0335 (0.0435)
N	31,966

Leave-one-out average percentage full eligibility, leave-one-out average percentage of use of basic grant, and leave-one-out average percentage of whether one or more parents are on welfare are used as instruments for the leave-one-out average uptake. Standard errors in brackets. *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$

As I mentioned before, the regressions in tables 4 and 5 are performed for both eligible and ineligible students. However, we are only interested in how people that are eligible for a grant are influenced by their peers. At the same time you want ineligible classmates to be able to influence the decision. Imagine for example being in a class with 15 students including 3 eligible classmates, who have all applied for the supplementary grant. This could feel very different from being in a class with 15 students including 10 eligible classmates that have all applied for a grant. Even though in both cases, 100% of your eligible classmates have applied for the grant. What you would want in terms of a SAR-model is:

$$Y_g^{eli} = \beta W_g Y_g^{all} + X_g \zeta + \epsilon_g \tag{19}$$

Unfortunately, spatial models cannot easily handle this situation. The length of vector Y_g^{eli} is different from Y_g^{all}. This is another disadvantage of the spatial approach in practice, whereas the traditional approach can easily differentiate between eligible and ineligible students in the equation. Simply calculate the leave-one-out average over all students in a class, giving a $\bar{y}_{-igr}$ which belongs to a particular eligible student, but includes her non-eligible classmates as a reference group.

Furthermore, there are disadvantages for the spatial approach that stem from using a large data set. If I want to include all students in MBO-3 and MBO-4 in school year 2020-2021, there are 266,943 observations. Previously, SAR-logit-fe took 49 minutes for a data set that was 8.4 times smaller. Even if the run time only increased linearly, this would mean that SAR-logit-fe would need to run for about 7 hours. In contrast, the IV-fe regression on the entire data sets runs in 12.977 seconds.

Regrettably, it is not even possible to run SAR-logit-fe on the entire data set. It requires saving a vector of size 15.4GB, which is far above the standard data limit for working in the microdata environment of the CBS. We already have a suspicion that the instruments are flawed, we cannot look specifically at only eligible students, and including 3839 different school dummies in the data set runs the risk of the incidental parameter problem. Combined with the expected long running times, and excessive memory requirements, paying a fee to increase the working memory to run SAR-logit-fe on the full data set would probably not be worth it.

As I can only present the results for the large data set for the traditional IV-fe approach, I opt to use a subgroup of the leave-one-out averages of the x-variables as instruments, including the others as exogenous peer effects. Specifically, the average percentages for eligibility for the full monetary amount, whether someone has a basic student finance grant and whether one or both of someone's parents receive(s) welfare benefits are used as instruments, while the leave-one-out averages of total years in MBO, gender and non-Dutch background are directly plugged into the equation. I also opt to focus solely on the uptake of eligible students for the left hand side of the equation. The IV-fe results are as follows.

Table 6: Regression results IV-fe (3)

Second stage IV-fe (3)	
Leave-one-out average uptake	-0.150*** (0.0186)
Total years in MBO	0.0165*** (0.0007)
Fully eligible	0.0927*** (0.0018)
Any basic grant	0.748*** (0.0021)
Parent on welfare	0.0685*** (0.0019)
Non-Dutch background	0.0653*** (0.0022)
Female	0.0063** (0.0023)
Leave-one-out average total years in MBO	-0.0010 (0.0020)
Leave-one-out average percentage non-Dutch background	0.0734*** (0.0135)
Leave-one out average percentage female	-0.0132 (0.0143)
N	132,048

Only the uptake decision of eligible students is used as a dependent variable in this regression. Full eligibility, use of the basic grant, and whether one or more parents are on welfare are used as instruments for the average uptake. Standard errors in brackets. *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$

Here we see that the endogenous peer effects are of a non-negligible size, negative and highly significant, which is again counter to what was hypothesised. We also see that in this regression the exogenous peer effects in the form of average years in MBO are no longer significant. Instead, the percentage of classmates that is non-Dutch positively affects the uptake of the supplementary grant, still showing some exogenous peer effects.

The Kleibergen-Paap rk LM test statistic (3774.04) and Kleibergen-Paap F statistic (4421.05) show that there again is no issue with the relevance of the instruments. The p-value for the Hansen J-test does, however, leave some doubt. It is 0.051, which means that the null-hypothesis is not rejected at the 5% level, but barely. This low p-value still casts doubts on the validity of the instruments, making me hesitant to attach too much value to the results presented in table 6.

7 Conclusion

In conclusion, two different but closely related approaches to peer effects modeling have developed over the last few decades. These approaches are plagued by the same theoretical issues, but have found different ways of discussing them and sometimes of solving them. Traditional modellers often estimate instrumental variable regressions with fixed effects, as a way of combating the reflection problem, endogeneity, correlated effects and selection bias. Spatial modellers often use generalised spatial least squares to combat endogeneity, and mention that having a zero-diagonal in the spatial weights matrix is a way of overcoming the reflection problem. They agree that fixed effects can be a solution to selection bias, but also have been working on explicitly modelling endogeneity in spatial weight matrices. Spatial modellers can also explicitly model correlated effects by enforcing a spatial structure on the error term.

Additionally, both approaches often use theory to argue which type of model specification is most appropriate. In the context of a classroom, endogenous peer effects are of particular interest. Traditional modellers are likely to call their estimations reduced form models, and argue that any peer effects that are found are a combination of both endogenous and exogenous peer effects. Spatial modellers on the other hand are more likely to argue that exogenous peer effects do not occur, or to not mention the possibility of exogenous peer effects.

I focused on a setting in which it is impossible to distinguish which classmates interact. In this case using the spatial weight matrix W is indistinguishable from taking the leave-one-out mean. Additionally, because my motivation is to research the uptake of the supplementary grant, the spatial approach employs generalized spatial two stage least squares (GS2SLS), which is nearly identical to the 2SLS used by the traditional approach. In theory, the spatial autoregressive (SAR) model and the traditional instrumental variable (IV) model should therefore lead to the exact same results, as WX and $\bar{x}_{-ig}$ are used as instruments. I showed their similarity empirically in section 5. When in addition errors are all exogenous, not only do the IV-model an SAR-model give the same results, but the spatial autocorrelation model (SAC), which incorporates spatial relationships in the error term is not so different either. SAC, however, did have a slightly lower coverage probability (92.9% versus 95.1% for IV), as it erroneously found significant correlated effects instead.

In contrast, when errors are not fully exogenous, but there is some correlation at a school level, the spatial and traditional approach are likely to pick different implementations. In theory both want to include fixed effects, however in practice this not often done by spatial modellers. Whether and how to include fixed effects is still an ongoing debate in the spatial literature. Traditional modellers use a within-transformation. Current software implementations by spatial modellers however can only use with-in transformations when group sizes are all equal. Including dummy-variables is likely to cause the incidental parameter problem. Or the dummies can cause multicollinearity when the SAR- and SAC-models use both WX and WWX as instruments in GS2SLS. This issue of multicollinearity is only present in a classroom setting, in which entire blocks of the W-matrix have the same value for the fixed effects dummies. Spatial modellers are most likely to pick a SAC-model without fixed effects when they expect correlated errors. SAC can at least account for some part of the correlation between members of the same group (albeit always on the classroom level, and not on a school level). In the second simulation study, I showed that IV with fixed effects was least likely to reject the null-hypothesis that the estimated peer effects were equal to the true peer effects. SAC was second most likely, but it also failed to find significant peer-effects in nearly 40% of the 1000 runs. SAC therefore most likely had high standard errors in each individual run. Additionally, SAC is the slowest model in all of the simulations. As expected, IV/SAR performed the worst under the second DGP, although IV is the fastest method, as it was for all DGPs

Looking specifically at a setting with a binary outcome variable, the difference between the two approaches is most striking. Binary variables means residuals are no longer normally distributed nor homoskedastic, which needs to be taken into account. Both approaches commonly estimate linear probability models, but spatial modellers additionally have an extensive literature documenting spatial logit models. I therefore chose to also estimate a SAR-logit model (with fixed effects), estimated using linearized spatial logit GMM. In the third simulation, all traditional and spatial linear probability models seem get relatively close to the true beta when averaged over 1000 runs, although the bias was larger than in the two DGPs with continuous outcomes. Furthermore, when the true value of β was 0.15, the linear models failed to find significant peer effects in more than 75% of runs. In contrast, when the true peer effects were relatively large (0.45), all linear models perform very well. Even the ones that cannot deal with the correlated errors, like SAR/IV have an average $\hat{\beta}_m$ which is close to the true β, a high coverage probability and a low likelihood of finding non-significant effects. I suspect that is caused by the fact that the magnitude of the endogenous peer effects is much larger than the correlated part of the error term. Looking at the logit models SAR-logit(-fe) is among one of the better models for the small size of β. But the logit models are the worst choice to estimate large endogenous peer effects. The logit models have a large positive bias, which also occurs in other studies when for example $\beta_0 = 0.8$. However, in my simulations, the logit models already overshot the true β when it was 0.45. I showed that this was mostly due to the classroom structure of the W-matrix in my study.

Zooming in on the uptake of the supplementary grant, the traditional approach found significantly negative peer effects in different estimation. The SAR-logit fixed effects model found no significant peer effects. Negative endogenous peer effects contradict with what was hypothesised based on the literature. The main issue for both approaches stems from finding the right instruments. There were worries about the exclusion conditions, based on for example the Hansen J-test reported by the IV-model.The SAR-logit model made use of the same instruments, and LGMM also requires exogenous instruments, but statistics on the first stage of GS2SLS are not provided. Additionally, the current implementations of the spatial approach have quite some limitations in practice. I wanted to research how everyone in a class effected the uptake decision of the subset of eligible students in that class. The spatial approach does not have an easy way of implementing this, while the traditional approach allows for that flexibility. And lastly, the run time and memory requirements of the spatial models are much more demanding than for the traditional approach, this was especially problematic for the given data set with 266,943 MBO-3 and MBO-4 students.

All in all, it seems that in the current set-up neither the spatial nor the traditional approach are suitable to find out the role of peers on the uptake of the supplementary grant in MBO. More research is necessary. Other instruments could be used or perhaps the identification of the peer effects should stem from other solutions than from the type of models. For example, identification could come from experiments or natural experiments. For instance, you could look specifically at students that switch study programmes. These students could move from a programme with few students that take up the grant to programmes in which many students do. You would of course have to compare them to other "moving" students, to see what the influence of the peer group is.

8 Discussion

As I mentioned throughout, the spatial approach has some software limitations. In particular, fixed effects are hard to implement, and R packages for heteroskedasticity robust GS2SLS do not allow you to choose your instruments. I opted to follow what most spatial modellers would do in practice. Most spatial papers do not include fixed effects. I also discussed fixed effects for varying group sizes with prof. Elhorst the Editor-in-Chief of Spatial Economics Analysis, who mentioned that currently there are no routines for varying group sizes. I could have chosen to program these packages myself, but that would have been beyond the scope of the current thesis. Suffice it to say, that the currently widely available software for spatial models would need improvement to be able to handle fixed effects in the same manner as tradtional models already can.

In the current research, all DGPs in section 5 did not include exogenous peer effects. In real life, often exogenous peer effects can occur. It would be interesting to see how the different models handle the fact that the true DGP includes $WX\gamma'$. The reduced form of all chosen specifications will suffer from a form of omitted variable bias, when exogenous peer effects are erroneously not estimated. Cook et al. (2015) allowed for exogenous peer effects in their research, but they did not look at binary outcome variables, nor at traditional IV models. It is not a priori clear if including exogenous peer effects would change the conclusions drawn in section 5.

Additionally, the simulations could have allowed for group correlation in the x-variables. However, this would also induce some correlation between the x-variables and the error term. This would mean that X is no longer exogenous, and therefore a poor choice of instrument. That would be problematic for both the traditional and the spatial approach. Having some x-variables that are correlated and another part that is exogenous would be a good solution, to still be able to estimate. Most current spatial software for heteroskedastic processes, however, does not offer to use only some of the x-variables as instruments, although it has been suggested as an option in theory.

Another note on the simulations is that I chose to have the error terms correlate at the school level, following the reasoning of traditional modellers. Perhaps if the errors were correlated at the classroom level, SAC could outperform the other spatial model and the IV-model.

Furthermore, in the simulation studies I used Monte Carlo simulations, but also looked at what happened in individual runs. As was visible in section 5, some of the approaches are consistent if you run them 1000 times, but are really lacking in individual runs. This is interesting to know, perhaps peer effects researchers looking at empirical applications could benefit from bootstrapping to get closer to the true value of β.

In the conclusion I mention that further research into the uptake of the supplementary grant is necessary. The suggestions mentioned in that section are all for cross-sectional data. Another option could be to make use of panel data. The traditional approach is much further ahead of the spatial approach concerning panel models. Lagged values of the y-variables are very common instruments to combat endogeneity issues for traditional modellers. Using panel models is difficult for spatial modellers that are interested in peer effects. Many implementations of spatial panel models use a W-matrix which is the same at each time t. Unchanging weight matrices are realistic when looking at which countries border each other, but not realistic when looking at for example friend networks or classmates. Some researchers are working on time-varying W-matrices, like Qu, Lee, and Yu (2017). Changing $y_{-ig,t-1}$ in the tradtional approach in contrast, is much easier.

References

Abadie, Alberto et al. (2017). *When should you adjust standard errors for clustering?* Tech. rep. National Bureau of Economic Research.

Abbiati, Giovanni and Jonathan Pratschke (2021). *"Like with Like" Or "Do Like"?: Modelling Peer Effects in the Classroom.*

Ajilore, Olugbenga (2015). "Identifying peer effects using spatial analysis: the role of peers on risky sexual behavior". In: *Review of Economics of the Household* 13.3, pp. 635–652.

Ammermueller, Andreas and Jörn-Steffen Pischke (2009). "Peer effects in European primary schools: Evidence from the progress in international reading literacy study". In: *Journal of Labor Economics* 27.3, pp. 315–348.

An, Weihua (2011). "Models and methods to identify peer effects". In: *The Sage handbook of social network analysis*, pp. 514–532.

— (2015). "Instrumental variables estimates of peer effects in social networks". In: *Social Science Research* 50, pp. 382–394.

Angrist, Joshua D (2014). "The perils of peer effects". In: *Labour Economics* 30, pp. 98–108.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Anselin, Luc (2017). *Spatial Regression – Specification Tests", Lecture Series.*

— (2020). "Contiguity-Based Spatial Weights". In: *GeoDa. URL https://geodacenter. github. io/-workbook/4a_contig_weights/lab4a. html# queen-contiguity (accessed 6.23. 19).*

Anselin, Luc et al. (1996). "Simple diagnostic tests for spatial dependence". In: *Regional science and urban economics* 26.1, pp. 77–104.

Åslund, Olof and Peter Fredriksson (2009). "Peer effects in welfare dependence quasi-experimental evidence". In: *Journal of human resources* 44.3, pp. 798–825.

Baltagi, Badi H, Ying Deng, and Xiangjun Ma (2018). "Network effects on labor contracts of internal migrants in China: a spatial autoregressive model". In: *Empirical Economics* 55.1, pp. 265–296.

Baltagi, Badi H and Zhenlin Yang (2013). "Heteroskedasticity and non-normality robust LM tests for spatial dependence". In: *Regional Science and Urban Economics* 43.5, pp. 725–739.

Beron, Kurt J and Wim PM Vijverberg (2004). "Probit in a spatial context: a Monte Carlo analysis". In: *Advances in spatial econometrics*. Springer, pp. 169–195.

Boucher, Vincent and Yann Bramoullé (2020). "Binary Outcomes and Linear Interactions". In: *HAL open science*.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of econometrics* 150.1, pp. 41–55.

Brenøe, Anne Ardila and Ulf Zölitz (2020). "Exposure to more female peers widens the gender gap in STEM participation". In: *Journal of Labor Economics* 38.4, pp. 1009–1054.

Brock, William A and Steven N Durlauf (2007). "Identification of binary choice models with social interactions". In: *Journal of Econometrics* 140.1, pp. 52–75.

Burridge, Peter, J Paul Elhorst, and Katarina Zigova (2016). "Group interaction in research and the use of general nesting spatial models". In: *Spatial econometrics: Qualitative and limited dependent variables*. Emerald Group Publishing Limited.

Caeyers, Bet and Marcel Fafchamps (2016). *Exclusion bias in the estimation of peer effects*. Tech. rep. National Bureau of Economic Research.

Calabrese, Raffaella and Johan A Elkink (2014). "Estimators of binary spatial autoregressive models: A Monte Carlo study". In: *Journal of Regional Science* 54.4, pp. 664–687.

Cook, Scott J, Jude C Hays, and Robert J Franzese (2015). "Model specification and spatial interdependence". In: *APSA Conference Paper*.

Corrado, Luisa and Bernard Fingleton (2012). "Where is the economics in spatial econometrics?" In: *Journal of Regional Science* 52.2, pp. 210–239.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2007). "Be as careful of the books you read as of the company you keep: Evidence on peer effects in educational choices". In: *IZA Discussion Papers*.

Dubin, Robin, AS Fotheringham, and PA Rogerson (2009). “Spatial weights”. In: *The Sage handbook of spatial analysis*, pp. 125–158.

Elhorst, J Paul (2010). “Applied spatial econometrics: raising the bar”. In: *Spatial economic analysis* 5.1, pp. 9–28.

Elhorst, J Paul et al. (2019). “Raising the bar (13)”. In: *Spatial Economic Analysis* 14.4, pp. 379–383. DOI: 10.1080/17421772.2019.1678936. URL: https://doi.org/10.1080/17421772.2019.1678936.

Epple, Dennis and Richard E Romano (2011). “Peer effects in education: A survey of the theory and evidence”. In: *Handbook of social economics*. Vol. 1. Elsevier, pp. 1053–1163.

Epple, Dennis, Richard E Romano, and Holger Sieg (2003). “Peer effects, financial aid and selection of students into colleges and universities: An empirical analysis”. In: *Journal of Applied Econometrics* 18.5, pp. 501–525.

Feld, Jan and Ulf Zölitz (2017). “Understanding peer effects: On the nature, estimation, and channels of peer effects”. In: *Journal of Labor Economics* 35.2, pp. 387–428.

Figlio, David N. (Sept. 2007). “Boys Named Sue: Disruptive Children and Their Peers”. In: *Education Finance and Policy* 2.4, pp. 376–394. URL: https://ideas.repec.org/a/tpr/edfpol/v2y2007i4p376-394.html.

Gaviria, Alejandro and Steven Raphael (2001). “School-based peer effects and juvenile behavior”. In: *Review of Economics and Statistics* 83.2, pp. 257–268.

Gibbons, Stephen and Henry G Overman (2012). “Mostly pointless spatial econometrics?” In: *Journal of regional Science* 52.2, pp. 172–191.

Hinke, Stephanie von, George Leckie, and Cheti Nicoletti (2019). “The use of instrumental variables in peer effects models”. In: *Oxford Bulletin of Economics and Statistics* 81.5, pp. 1179–1191.

Kelejian, Harry H and Ingmar R Prucha (1998). “A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”. In: *The Journal of Real Estate Finance and Economics* 17.1, pp. 99–121.

— (2010). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”. In: *Journal of econometrics* 157.1, pp. 53–67.

Klier, Thomas and Daniel P McMillen (2008). “Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples”. In: *Journal of Business & Economic Statistics* 26.4, pp. 460–471.

Lavy, Victor and Analia Schlosser (2011). "Mechanisms and impacts of gender peer effects at school". In: *American Economic Journal: Applied Economics* 3.2, pp. 1–33.

Lee, Lung-fei (2004). "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models". In: *Econometrica* 72.6, pp. 1899–1925.

— (2007). "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects". In: *Journal of Econometrics* 140.2, pp. 333–374.

Lee, Lung-fei, Ji Li, and Xu Lin (2014). "Binary choice models with social network under heterogeneous rational expectations". In: *Review of Economics and Statistics* 96.3, pp. 402–417.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010). "Specification and estimation of social interaction models with network structures". In: *The Econometrics Journal* 13.2, pp. 145–176.

Leenders, Roger Th AJ (2002). "Modeling social influence through network autocorrelation: constructing the weight matrix". In: *Social networks* 24.1, pp. 21–47.

LeSage, James P (2000). "Bayesian estimation of limited dependent variable spatial autoregressive models". In: *Geographical Analysis* 32.1, pp. 19–35.

LeSage, James P and Robert Kelley Pace (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Lin, Xu (2010). "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables". In: *Journal of Labor Economics* 28.4, pp. 825–860.

Lin, Xu and Lung-fei Lee (2010). "GMM estimation of spatial autoregressive models with unknown heteroskedasticity". In: *Journal of Econometrics* 157.1, pp. 34–52.

Manski, Charles F (1993). "Identification of endogenous social effects: The reflection problem". In: *The review of economic studies* 60.3, pp. 531–542.

Markussen, Simen and Knut Røed (2015). "Social insurance networks". In: *Journal of Human resources* 50.4, pp. 1081–1113.

— (2017). "The gender gap in entrepreneurship – The role of peer effects". In: *Journal of Economic Behavior & Organization* 134.C, pp. 356–373. DOI: 10.1016/j.jebo.2016.12.01.

McMillen, Daniel P (1992). "Probit with spatial autocorrelation". In: *Journal of Regional Science* 32.3, pp. 335–348.

McVicar, Duncan (2011). "Estimates of peer effects in adolescent smoking across twenty six European Countries". In: *Social Science Medicine* 73.8, pp. 1186–1193. ISSN: 0277-9536. DOI:

https://doi.org/10.1016/j.socscimed.2011.08.006. URL: https://www.sciencedirect.com/science/article/pii/S027795361100493X.

Mora, Toni and Philip Oreopoulos (2011). "Peer effects on high school aspirations: Evidence from a sample of close and not-so-close friends". In: *Economics of Education Review* 30.4, pp. 575–581.

Niu, Ziheng et al. (2022). "Peer effects, attention allocation and farmers' adoption of cleaner production technology: Taking green control techniques as an example". In: *Journal of Cleaner Production* 339, p. 130700. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2022.130700. URL: https://www.sciencedirect.com/science/article/pii/S0959652622003390.

Paloyo, Alfredo R (2020). "Peer effects in education: recent empirical evidence". In: *The Economics of Education*. Elsevier, pp. 291–305.

Pinkse, Joris and Margaret E Slade (1998). "Contracting in space: An application of spatial statistics to discrete-choice models". In: *Journal of Econometrics* 85.1, pp. 125–154.

Qu, Xi and Lung-fei Lee (2015). "Estimating a spatial autoregressive model with an endogenous spatial weight matrix". In: *Journal of Econometrics* 184.2, pp. 209–232. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2014.08.008. URL: https://www.sciencedirect.com/science/article/pii/S0304407614001870.

Qu, Xi, Lung-fei Lee, and Chao Yang (2021). "Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables". In: *Journal of Econometrics* 221.1, pp. 180–197. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2020.05.011. URL: https://www.sciencedirect.com/science/article/pii/S0304407620302281.

Qu, Xi, Lung-fei Lee, and Jihai Yu (2017). "QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices". In: *Journal of Econometrics* 197.2, pp. 173–201.

Rosaz, Julie, Robert Slonim, and Marie Claire Villeval (2016). "Quitting and peer effects at work". In: *Labour Economics* 39, pp. 55–67. ISSN: 0927-5371. DOI: https://doi.org/10.1016/j.labeco.2016.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0927537116000087.

Ryabov, Igor (2011). "Adolescent academic outcomes in school context: Network effects reexamined". In: *Journal of Adolescence* 34.5, pp. 915–927.

Sacerdote, Bruce (2011). "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In: ed. by Erik Hanushek, Stephen Machin, and Ludger

Woessmann. 1st ed. Vol. 3. Elsevier. Chap. 04, pp. 249–277. URL: https://EconPapers.repec.org/RePEc:eee:educhp:3-04.

Visser, D., M. Zumbuehl, and S. Konijn (2020). *Niet-gebruik van de aanvullende beurs*. Tech. rep. CPB.

Yeung, Ryan and Phuong Nguyen-Hoang (2016). "Endogenous peer effects: Fact or fiction?" In: *The Journal of Educational Research* 109.1, pp. 37–49. DOI: 10.1080/00220671.2014.918528.

Zhao, Chuanmin and Xi Qu (2021). "Social networks and internal migration in China: A spatial autoregressive model". In: *Review of Development Economics*.

Zumbuehl, M. and C. Magnée (2022). *Niet-gebruik van de aanvullende beurs in het mbo*. Tech. rep. CPB.

Appendix

Student finance in the Netherlands

The following explanation is based on Zumbuehl and Magnée (2022). An MBO student can qualify for four different forms of student finance: a basic grant, a student travel product, a loan and a supplementary grant. Whether an MBO student is entitled to student finance depends on the age, nationality, residence status of the student and the type of MBO programme. For example, students over 30 are not entitled to a student grant. Furthermore, part-time MBO students or students who do not follow a vocational training path (BOL) are not eligible for student finance. Additional rules apply for student finance for non-Dutch students.

MBO students must be over 18 to be eligible for the basic grant, a loan and a supplementary grant. Students can apply for these student finance products for the first time in the quarter after their 18th birthday. If a student applies for the supplementary grant at some point during the year, DUO (the state body for the implementation of education) retroactively pays the grant for the entire year. The duration of the student finance depends on the MBO level. Students in an MBO-3 or MBO-4 program are entitled to the basic grant and supplementary grant for the nominal duration of their studies. They are entitled to the student travel product and a loan for a maximum of 7 years. The basic grant, supplementary grant and student travel product are a performance grant. If students achieve a degree within 10 years, the grants are turned into a gift.

The amount of the basic grant and the supplementary grant depend on the student's living situation and whether the student pays tuition fees. A student who follows a BOL study and who is 18 years or older on 1 August must pay tuition fees for that academic year. A student who pays tuition will receive a higher amount than a student who does not pay tuition. A student who lives away from home also receives a higher amount than a student living at home. Additionally, the supplementary grant depends on the parental income as earned two years prior. The supplementary grant is not the same for every student, but decreases as parental income increases.

Example of multicollinearity in WX and WWX

Imagine 10 students divided over three schools and four classes, with one continuous x-variable that is drawn from a standard normal distribution. The first two students are in the same class, the second two students are as well. The next three are in a class together, as are the last three students. If entire classes are assigned to the schools, as is true in real life, the corresponding X_1 and W look like:

$$X_1 = \begin{bmatrix} 0.34 & 1 & 0 \\ 1.35 & 1 & 0 \\ -1.46 & 1 & 0 \\ -0.05 & 1 & 0 \\ -1.03 & 0 & 1 \\ -0.09 & 0 & 1 \\ -0.10 & 0 & 1 \\ -0.13 & 0 & 0 \\ -0.45 & 0 & 0 \\ 0.29 & 0 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

Here, the second and third column of X_1 are school dummies. The first four students are in school 1, the next three in school 2 and the final three students are in school 3. Calculating WX_1 and WWX_1 gives the following:

$$WX_1 = \begin{bmatrix} 1.35 & 1 & 0 \\ 0.34 & 1 & 0 \\ -0.05 & 1 & 0 \\ -1.46 & 1 & 0 \\ -0.10 & 0 & 1 \\ -0.56 & 0 & 1 \\ -0.56 & 0 & 1 \\ -0.08 & 0 & 0 \\ 0.08 & 0 & 0 \\ -0.29 & 0 & 0 \end{bmatrix}, \quad WWX_1 = \begin{bmatrix} 0.34 & 1 & 0 \\ 1.35 & 1 & 0 \\ -1.46 & 1 & 0 \\ -0.05 & 1 & 0 \\ -0.56 & 0 & 1 \\ -0.33 & 0 & 1 \\ -0.33 & 0 & 1 \\ -0.11 & 0 & 0 \\ -0.19 & 0 & 0 \\ 0.00 & 0 & 0 \end{bmatrix}$$

Clearly, the second and third column of WX_1 and WWX_1 are exactly the same. This does not happen if the ones and zeros in the second and third column of the X-matrix are instead randomly divided over people, and not the same for everyone in the same block of matrix W.

$$X_2 = \begin{bmatrix} 0.34 & 1 & 0 \\ 1.35 & 0 & 0 \\ -1.46 & 0 & 0 \\ -0.05 & 1 & 0 \\ -1.03 & 0 & 1 \\ -0.09 & 1 & 0 \\ -0.10 & 0 & 1 \\ -0.13 & 0 & 1 \\ -0.45 & 1 & 0 \\ 0.29 & 0 & 0 \end{bmatrix}, \quad WX_2 = \begin{bmatrix} 1.35 & 0 & 0 \\ 0.34 & 1 & 0 \\ -0.05 & 1 & 0 \\ -1.46 & 0 & 0 \\ -0.10 & 0.5 & 0.5 \\ -0.56 & 0 & 1 \\ -0.56 & 0.5 & 0.5 \\ -0.08 & 0.5 & 0 \\ 0.08 & 0 & 0.5 \\ -0.29 & 0.5 & 0.5 \end{bmatrix}, \quad WWX_2 = \begin{bmatrix} 0.34 & 1 & 0 \\ 1.35 & 0 & 0 \\ -1.46 & 0 & 0 \\ -0.05 & 1 & 0 \\ -0.56 & 0.25 & 0.75 \\ -0.33 & 0.5 & 0.5 \\ -0.33 & 0.25 & 0.75 \\ -0.11 & 0.25 & 0.5 \\ -0.19 & 0.5 & 0.25 \\ 0.00 & 0.25 & 0.25 \end{bmatrix}$$

Unfortunately, in the real world this would correspond to people in the same class belonging to different schools, which is impossible.

Comparison of Calabrese and Elkink's set-up with different W's

I followed Calabrese and Elkink's set-up exactly. Their R set-up is helpfully available online.[17] I compared the outcomes using their W-matrix to the results using a non-overlapping contiguity W-matrix/classroom set-up.

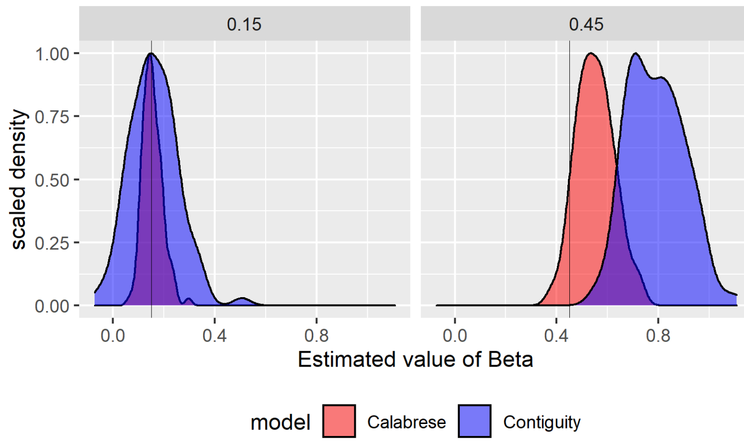

Figure 5: Densities of Calabrese and Elkink's simulations with different W-matrices

As can be seen in figure 5, the average is quite similar when the endogenous peer effects are small, but the distribution is much more concentrated for $W_{Calabrese}$. When $\beta = 0.45$ $W_{Contiguity}$ leads to a much larger positive bias. The same can be seen in table 7.

Table 7: Comparison using Calabrese and Elkink's W-matrix versus contiguity W-matrix.

	Calabrese	Contiguity	Calabrese	Contiguity
	$\beta_0 = 0.15$		$\beta_0 = 0.45$	
Estimated β_m	0.1540	0.1596	0.5507	0.7909
Significant peer effects	97	37	100	100
Reject $\hat{\beta}_m = \beta_0$	0	1	0	0
Average run time (s)	0.0501	0.0543	0.0516	0.0519

Performance of SAR-logit model with LGMM estimation for two levels of β_0 over 100 repetitions.

[17] `http://www.joselkink.net/wp-content/uploads/2010/04/bsar-mc-jrs-online-appendix.pdf`