Competition in the Public School Sector: Strategic Interaction and Policy Innovation Among US School Districts

Dissertation

zur Erlangung des wirtschaftswissenschaftlichen Doktorgrades der Wirtschaftswissenschaftlichen Fakultät der Universität Göttingen

vorgelegt von

Johannes Rincke

aus Stuttgart.

Göttingen, 2006

Videmus siquidem et ad oculum experimur, qualiter Parisius per convocacionem et congregacionem peritorum, scientificorum et prudentum Franciam irradiat et venustat, quomodo Bononia et Padwa Italiam fortificat et exornat qualiterque Praga Bohemiam illuminate et extollit, aut quomodo Vxonia totam fere Almaniam clarificat et fecundat. Profecto ad hoc summi disposicione presidii plurimarum terrarium obtinuimus prinicipatum et Regni Polonie recepimus dyadema, ut ipsum regnum claritate doctarum personarum illustremus, quarum doctrinis defectus et umbras eius possemus evellere ipsumque ceteris regionibus coequare.

STUDIUM GENERALE CRACOVIENSE A WLADISLAO REGE POLONIAE RENOVATUR.
DIE 26 MENSIS JULII 1400

We notice that Paris, by assembling scholarly, erudite and prudent men, is enlightening France, that Bologna and Padova are strengthening and adorning Italy, how Bohemia is illuminated and extolled by Prague, and that Oxford is brightening the whole of England and is making it fruitful. Truly, we have acceded the reign over the kingdom of Poland and received the crown in order to enlighten the kingdom by the glory of wise men and to eliminate by their knowledge the shadows of poorness and insufficiency and to draw level with other countries.

DOCUMENT OF THE RE-OPENING OF THE UNIVERSITY OF KRAKOW BY WŁADYSŁAW JAGIEŁŁO, KING OF POLAND.
JULY 26, 1400

Für Leńka und Jakob

Acknowledgements

This book contains work that was written between spring 2003 and winter 2005/06 while I was a Research and Teaching Fellow at the Centre for European Economic Research (ZEW) and the Department of Economics at the University of Mannheim. At the same time, I was also affiliated to the University of Göttingen as a doctoral student. Along the way, many people have helped me enormously with their advice and encouragement. The first to be mentioned are Robert Schwager and Thiess Büttner. I have learned much from them about modern theoretical and empirical approaches to local public finance and urban economics, and I am grateful to them for many hours of discussion and for their constant support.

I cannot discount the excellent training I received from many people, especially Kai A. Konrad and Jürgen Wolters at the Freie Universität Berlin. Numerous other people at ZEW and the Universities of Mannheim and Göttingen as well as participants at various conferences and workshops have assisted me with their comments and suggestions. Furthermore, I have benefitted from the advice of anonymous referees for the *Journal of Urban Economics* and *Public Choice*. A special thank goes to Jan K. Brueckner for his comments on a paper containing an earlier version of Chapter 4. The material has been published in the *Journal of Urban Economics* under the title "Competition in the public school sector: Evidence on strategic interaction among US school districts" (vol. 59, 352-369), and I would like to thank Elsevier Inc. for the permission to include it in this book. Finally, Rüdiger Göbel, Michał Kowalik, Emilia Maier, Andreas Schaich, Christoph Schottmüller, and Falko Tabbert have provided me with valuable research assistance. Although so many people have helped me, any remaining errors are my own.

Munich, June 2006 *Johannes Rincke*

Contents

Chapter 1

Introduction

1.1 Strategic interaction and policy innovation

What causes a government or administration to adopt a new policy? In particular, what is the role of intergovernmental competition in the diffusion of public sector innovations? Despite the fact that the research in political science has been very active in collecting mostly descriptive material on policy diffusion,[1] relatively little is known today about the driving forces of public sector innovation. There are basically two reasons for the unsatisfactory state of the research on the topic. Firstly, there has been little rigorous theoretical research on the incentives of decision makers in the public sector to experiment with new policies and to adopt policy innovations invented elsewhere. Secondly, the identification of the determinants of public sector innovation is complicated by numerous methodological problems. As a consequence, the number of well-crafted empirical studies is rather limited, too.

This book sheds light on one particular aspect of public sector innovation:

[1]Walker (1969) paved the way for comparative policy analysis. Other contributions dealing with US state level innovations include Mintrom (1997), Caudill, Ford, Mixon Jr., and Peng (1995), Berry and Berry (1992), Berry and Berry (1990), Savage (1985), Canon and Baum (1981), and Gray (1973). See Savage (1985) and Berry (1994) for surveys.

The strategic interaction among local governments in the adoption of policy innovations. Apart from a chapter presenting an illustrative model showing how an immanent bias against public sector innovation and in favor of traditional policies can be partly overcome by decentralization, the main contribution of this book is to provide evidence on the effects of strategic interaction among local jurisdictions on the diffusion of new programs and policies. The focus of the empirical analysis is on the US public school system in general and the local school districts in particular.

In the analysis of innovation activity both in the private and in the public sector it is important to distinguish as clearly as possible between experimentation and adoption or emulation. While experimentation is the search for innovative solutions of technical problems or problems of governance, adoption and emulation occurs if agents decide to use, apply or implement solutions which are already available. Since the conditions for making use of new technologies, concepts, and policies in most cases differ across decision makers, adoption and emulation are often risky activities in the sense that outcomes cannot be perfectly predicted. Therefore, experimentation and adoption of innovations often cannot be clearly distinguished conceptually. In fact, experimentation can often (at least partly) be substituted by emulation, avoiding the cost which is associated with performing policy experiments. Of course, this aggravates the problem of distinguishing conceptually (and empirically) between experimentation and emulation. Despite the fact that this problem can only be solved to a limited extent in this book, the analysis focusses more on the aspect of adoption and emulation, i.e. the diffusion of new policies among local jurisdictions, than on policy experimentation. Empirically, this is done by investigating adoption decisions of local school districts with regard to policies and programs which are initiated at the state level. A well known example is public school choice programs. Since 1987, many states in the US have enacted school choice laws allowing the school districts to decide whether they allow resident students to choose a school other than their neighborhood school within the school district (intra-district choice) or whether they admit non-resident students at the local public schools (inter-district choice). The school districts' adoption decisions in this kind of state-initiated programs can be understood as participation decisions. The data describing the diffusion of this kind of policy innovations used in the empirical analyses confirm the expectation that payoffs from adopting new policies are uncertain even if the policy has not been invented by the adopting jurisdiction: Some school district adopt a new policy and abandon it a few periods later. Other districts start with a low level of innovation activity at some point in time and increase the level of innovation activity later on, suggesting that experimenting with a policy invented and adopted elsewhere provides useful information.

When thinking about strategic interaction among local governments with regard to public sector innovation, there are two classes of models which should be considered. The first class addresses the flows of the factors of production and residential choices between jurisdictions as the source of interaction among local governments. The second class focusses on the political decisions of residents as voters and derives interdependencies in local governments' decisions and behavior in a principal-agent framework with asymmetric information between the local government and the residents.

The classical contribution in the first class is Tiebout (1956). Examples for the widespread application of the framework both in theoretical and empirical terms are inter-jurisdictional tax competition and the capitalization of locally provided public goods and services. With respect to the allocation of mobile capital, the underlying idea is that the corresponding flows between jurisdictions reflect differentials in the net rate of return. Since taxes on capital income affect the net return, tax rates are natural instruments for local governments with taxing authority to compete for mobile capital. In a decentralized setting, the choice of tax rates involves a fiscal externality. This externality makes the setting of tax rates interdependent across jurisdictions and leads to the well known fact of underprovision of local public goods.[2]

Models addressing Tiebout choice and capitalization effects rest on the fact that residential choices depend on the quantity and quality of locally provided services, amenities, and property prices. Each local jurisdiction is characterized by certain amenities and provides a specific bundle of local public goods. Households sort such that their utility is maximized, taking property prices into account. In this setting, as in the tax competition example, decisions in one jurisdiction will exert an externality on other jurisdictions: Changes in the bundle of public goods provided in one jurisdiction will induce relocation decisions of some households and thereby affect property values elsewhere. Since public sector innovations can be seen as a change in the way the public sector achieves the provision of goods and services, models involving Tiebout choice provide ample opportunity to modelling strategic interaction among local governments in the adoption of policy innovations. However, there is virtually no literature on Tiebout models focussing on public sector innovation.

The most seminal contribution in the second class of models is Besley and Case (1995). The paper addresses the agency problem between political representatives and voters and shows how comparative performance evaluation by voters can drive local governments into yardstick competition.[3] Applying the model to tax setting in the US states, the authors provide evidence supporting the view of yardstick competition as a discipline enhancing mechanism. The idea of comparative performance evaluation is applicable quite

---

[2]The landmarks in the theoretical tax competition literature are Wilson (1986), Zodrow and Mieszkowski (1986), and Wildasin (1989). For a survey, see Wilson (1999).

[3]The fundamental idea of yardstick competition in the public sector is borrowed from earlier studies in the industrial organization and contest theory literature. See Holmstrom (1982) for a general treatment and Nalebuff and Stiglitz (1983) for an application to labor market tournaments.

generally and lends itself to modelling the effects of yardstick competition on the incentives of governments to adopt or experiment with new policies. The model presented in Chapter 2 is a straightforward example, focussing on the implementation of political best-practice technologies.

Approaches based on Tiebout choice and yardstick competition models share the common feature that they quite naturally lead to predictions saying that decentralization promotes public sector innovation, and that the adoption of a new policy in one jurisdiction increases the probability of adoption in competing jurisdictions. Empirically, this would be reflected in adoption decisions to be positively spatially correlated. Focussing on the issue of policy experimentation, recent research has put forward the concern that learning externalities could prevent agents from experimenting at an efficient level (Strumpf, 2002, Bolton and Harris, 1999). In general, the presence of learning externalities challenges the traditional view that decentralized systems of government provide natural laboratories for public sector innovation.[4] To the extent that the diffusion of policy innovations is driven by adoption decisions with uncertain outcomes, the concerns about free-riding incentives carry over from policy experimentation to policy diffusion. As mentioned above, the problem of distinguishing between experimentation and adoption cannot be perfectly solved in this book. However, the empirical results point to positive interdependencies among adoption decisions irrespective of whether contemporaneous decisions or decisions involving time lags are considered. This can be interpreted as evidence suggesting that the effect of learning externalities on the behavior of local governments towards public sector innovation is dominated by incentives to adopt innovative policies which are (expected to be) used elsewhere.

Methodologically, the empirical parts of the book are closely related to the

[4]See Oates (1999) for a broad discussion of the issue.

vast and growing literature on strategic interaction among jurisdictions. The literature is surveyed by Brueckner (2003) and Revelli (2005). Most contributions have focussed on strategic interaction in tax rates (e.g, see Besley and Case (1995), Brett and Pinkse (1997), Hayashi and Boadway (2000), Büttner (1999, 2001), Brueckner and Saavedra (2001), Bordignon, Cerniglia, and Revelli (2003), and Egger, Pfaffermayr, and Winner (2005)) and expenditures (e.g, see Case, Hines, and Rosen (1993), Baicker (2005) and Revelli (2006)). Büttner (2003) extends this literature by providing evidence on fiscal externalities of taxing decisions of local jurisdictions. Interaction in non-fiscal policy instruments has been addressed by Brueckner (1998), dealing with growth controls in California cities, and in a study on environmental policies of the US states by Fredriksson and Millimet (2002).

1.2 The US school system as a laboratory for public sector innovation

When I started doing research on public sector innovation in decentralized systems of government, I was happy to discover that the US public school system provides a sort of laboratory for the decentralized provision of public goods and services. This laboratory offers excellent conditions for empirical research on various topics such as tax and expenditure competition, the technical and allocative efficiency of public goods provision, capitalization of quality and quantity of publicly provided goods and services in property values, and the determinants of policy experimentation and public sector innovation.

At the local level, the public school system is administrated and maintained by local school districts. As a special-purpose district, a school district is a unique body corporate and politic, and its sole responsibility is to operate the local public primary and secondary schools. Its legislative body, elected by direct popular vote, is called a school board, board of trustees, or school committee, and this body appoints a superintendent to function as the district's chief executive for carrying out day-to-day decisions and policy implementations.

The functioning of a school district can be a key influence and concern in local politics. A well run district with safe and clean schools, graduating enough students to good colleges and universities, can enhance the value of housing in its area.[5] This will directly benefit local residents and, at the same time, increase the amount of property tax revenue available to carry out the school district's operations. Conversely, a poorly-run district will adversely affect local residents. The importance of the quality of the local public schools both in terms of the students' labor market prospects and with respect to housing values results in a close monitoring of school district operations by local residents.[6] This, in particular, makes the school district level an attractive choice when testing implications and hypotheses about strategic behavior of local governments.

Another attractive fact supporting the choice of US school districts as the units of observation in empirical studies on public sector activities is the availability of exceptionally rich data. The first source to mention is the National Center for Education Statistics (NCES). It provides various data bases covering a variety of district characteristics on an annual basis and, in addition, conducts surveys on school district policies. Another valuable

[5] The capitalization of the quality of local public schools is well documented. See, for instance, Bradbury, Mayer, and Case (2001), Weimer and Wolkoff (2001), Bogart and Cromwell (2000, 1997), Colwell and Guntermann (1984), Gill (1983), Jud and Watts (1981), and Clotfelter (1975).

[6] When the General Social Survey in 1987 asked respondents to name the head of the local school system, 31% gave the correct answer.

source is the Bureau of the Census. It provides a huge variety of population and housing characteristics at the district level. Both the NCES data bases and the Census data cover all local school districts throughout the US. Finally, the Departments of Education of various states provide excellent data bases on their public elementary and secondary schools and school districts. Such state specific data sources are of tremendous importance for many empirical research projects because they often provide highly disaggregated data on test scores from standardized achievement tests. Using this data, it is possible to construct measures for the quality of education offered by the local public school sector.

To give a first impression of the characteristics of local school districts in the US, descriptive statistics for 12,556 districts in 2003/04 are given in Table 1.1.[7] Presently, a typical US school district operates about six schools, serving around 3000 students. The variation in district size is huge. The smallest district has one school serving a single student, while the largest district, Los Angeles City Unified, runs near to 700 schools with almost 750,000 students. The student-teacher ratio shows considerable variation, too, with a mean of 14.4. Public schools in the US receive funding from three different sources. On average, federal funds amount to 8 percent of all revenues, state funds cover about 50 percent, and local sources contribute around 42 percent to total revenues. Taken together, the typical school district has about 10,000 dollars of revenue per student. Again, there is huge variation in revenues per student. The numbers also imply that the average district has a budget of about 30 million dollars.

The Bureau of the Census sorts school districts into seven categories with

[7]The NCES Common Core of Data lists 14,472 local school districts for the 2003/04 school year. I have dropped districts from Alaska and all districts with missing values for some of the variables displayed. In addition, a number of non-operative districts and districts with a zero number of students or teachers have been eliminated.

Table 1.1: US local school districts in 2003/04, descriptive statistics (Nob=12,556)

	Mean	S.D.	Min	Max
Enrollment[a]	2.95	10.4	0.001	747
Number of schools	5.92	13.4	1.00	693
Student-teacher ratio	14.4	3.75	1.00	64
Total revenues per student[b]	9.86	4.38	0.668	114
Share of federal revenues	0.080	0.072	0	0.835
Share of state revenues	0.496	0.186	0	1.00
Share of local revenues	0.424	0.207	0	0.999
Large city	0.011	0.106	0	1.00
Mid size city	0.039	0.194	0	1.00
Urban fringe of large city	0.159	0.366	0	1.00
Urban fringe of mid-size city	0.094	0.291	0	1.00
Large town	0.007	0.085	0	1.00
Small town	0.114	0.318	0	1.00
Rural	0.392	0.488	0	1.00
Elementary school district	0.222	0.416	0	1.00
High school district	0.035	0.184	0	1.00
Unified school district	0.743	0.437	0	1.00

[a] In 1000 students.
[b] In 1000 \$ and 2003 prices.

regard to urbanicity, ranging from 'Large city' to 'Rural'. Only one percent of the school districts is categorized as being located in a large city, while 4 percent are located in mid-size cities. With 16 percent, 'Urban fringe of large city' is the category with the second highest frequency. About 9 percent of all districts are located in the urban fringe of mid-size cities, while roughly 50 percent belong to one of the remaining categories. 11 percent are categorized as small-town districts, while 39 percent are classified as rural.

In many states, some districts are specialized to operating only elementary or secondary schools. While a majority of 74 percent are unified school districts serving both elementary and secondary schools, 22 percent of all US school districts are elementary and 3.5 percent are high school districts.

Instead of analyzing public sector innovation at the local level, one could, in principle, also look at the level of states or provinces or even investigate adoption decisions with an international perspective. As mentioned above, there is a strong tradition of policy diffusion research beginning with Walker (1969), and most authors have used data on state policies in the US to produce empirical evidence on the determinants of public sector innovation.[8] A particular problem of these studies is the high degree of inertia in the political process which is usually present at higher level tiers of government. Technically, this is often reflected in positive adoption decisions with regard to some new policy or some newly invented model of governance being made once and for all: For each unit of observation, one can observe only a single switch towards the new policy (given that the new policy is adopted at all). This puts severe limits on the set of empirical models and techniques which might be used to investigate the determinants of adoption. At the local level, the political process seems to be much more flexible, and we regularly observe jurisdictions experimenting with new policies. Of course, 'experimenting' in the true sense means that new policies will proof inadequate in some cases. Consequently, at the local level we observe governments abandoning newly implemented policies. The resulting pattern of adoption decisions is much richer, and often the variation in observed policy choices can be exploited using a broad variety of cross-section and panel data methods.

[8]In addition to the studies cited in footnote 1, see Coughlin, Garrett, and Hernández-Murillo (2003) and Garret, Wagner, and Wheelock (2003) for recent approaches using analytical instead of descriptive techniques. Another interesting contribution is Boockman (2001), analyzing the behavior of national governments in the ratification of international conventions.

1.3 Summary of results

The remainder of the book is divided into five chapters, one providing a theoretical analysis of public sector innovation and four dealing with testing for strategic interaction in the adoption of policy innovations.

Chapter 2 presents a theoretical model of public sector innovation in a decentralized system of governments. In a setting with two jurisdictions, local governments have to choose between an old and a new policy with stochastic payoffs. There are two types of governments, benevolent and rent-seeking. For governments, it is costly to run the new policy in terms of administrative effort, and rent-seeking governments therefore have an incentive to run the old policy irrespective of the payoffs generated by both policies. There is asymmetric information between governments and voters with respect to policy payoffs: While governments are perfectly informed about policy payoffs for their own jurisdiction, voters can only observe payoffs net of transfers governments make to themselves. It is therefore difficult for the voters to distinguish a benevolent government operating under unfavorable circumstances from a rent-seeking government playing a mimicking strategy.

The model is used to illustrate that in correlated environments yardstick competition between governments of local jurisdictions may be a driving force for the diffusion of policy innovations. The intuition for this result is straightforward: Under yardstick competition it is more difficult for rent-seeking governments to maintain a sufficiently good reputation to be reelected. If it is optimal to use the new policy from a social perspective, benevolent governments will do so, and rent-seeking governments will be forced to follow in order to protect their reputation.

While other contributions dealing with public sector innovation discuss the incentives to experiment with new policies (see, for instance, Strumpf (2002), Kotsogiannis and Schwager (2006a), and Kotsogiannis and Schwager (2006b)), the model presented in Chapter 2 focusses on the incentives to implement political best-practice technologies in situations where policies maximizing the politician's utility differ from policies maximizing social welfare. This is relevant for public sector innovation since running new policies often requires higher effort. Politicians can earn rents by avoiding extra effort and by running traditional policies disproportionately often. The model shows how the resulting bias against pubic sector innovation can be alleviated by comparative performance evaluation.

The yardstick competition model of public sector innovation has a clear empirical implication. In order to learn about political best-practice technologies, citizens must collect and review information about the performance of several local governments. Since the costs will increase with the number of jurisdictions under monitoring, residents can be expected to use only a limited number of reference jurisdictions as a yardstick for policies in their own jurisdiction. Since the equilibrium with yardstick competition predicts local adoption decisions to positively depend on expected adoptions in reference jurisdictions, the model predicts positive spatial correlation in the adoption of public sector innovations across jurisdictions. Moreover, when dealing with local jurisdictions such as the US school districts, it is reasonable to think of reference or neighboring jurisdictions as being determined (at least partly) by geographical proximity. The reason is that, given that the number of potential reference jurisdictions at the local level is large and that monitoring and information costs increase with geographical distance, residents as well as policymakers will find it useful to refer to jurisdictions which are close to their own jurisdiction.[9]

[9]Local jurisdictions in large urban areas may be an exemption, since the number of reference jurisdictions is limited by the small overall number of such urban areas and information on the local public sector is often more readily available.

Chapter 3 gives an overview over empirical models with limited dependent variables incorporating spatial effects. However, the overview is not meant to be exhaustive. Rather, it aims at introducing the empirical models used in the chapters presenting empirical results on strategic interaction in public sector innovation. Four models are described and discussed: The spatial autoregressive probit, a latent variable model with time-lagged spatial effects, a probit model conditioning on neighbors' contemporaneous outcomes, and the spatial linear probability model. Advantages and disadvantages of all models are discussed, and estimation procedures are delineated. Further details and particular questions which arise in applying the models to specific estimation problems are discussed in Chapters 4 to 6.

Chapter 4 is the first out of three chapters presenting evidence on spatial effects in the adoption of policy innovations. The analysis revolves around the question how increased competition for mobile resources among jurisdictions affects the willingness of local governments to make use of new administrative concepts and regulations. In particular, I investigate how increased competition among local school districts affects the attitude of local decision makers to adopt school choice policies which, effectively, contribute to increased integration of local educational markets. The empirical analysis makes use of a significant change in the institutional environment of the public school sector in Michigan in 1996, when a voluntary inter-district choice program was established. Under this program, students were given the right to enroll at public schools outside their district of residence. Local school districts would not be forced to accept non-resident students, but districts participating in the choice program and allowing for the enrollment of transfer students would receive additional state funds depending on the number of non-resident students enrolled.

The evidence on interdependencies among the school districts' policies towards school choice is based on the districts' participation decisions in the second year of the program. The analysis exploits the cross-sectional variation in the degree of competition for non-resident students which resulted from the initial participation decisions of all districts and demonstrates that the districts' readiness to adopt the new policy was positively affected by competition of neighboring districts. The school districts' reaction to increased competition for students and resources can be described as a sort of 'striking back' behavior, i.e. the districts were more likely to compete for non-resident students if neighboring districts did so. The results suggest that the school districts did use participation in the inter-district choice program to actively engage in competition for students and resources. Facing increased risk of losing students and resources, the school districts seem to have flexibly reacted to competition from neighboring districts. The results also suggest that the impact of neighbors' decisions on the participation probability was substantial, and that competition for students contributed significantly to the increase in the share of districts allowing for inter-district transfers from 37 percent in 1996/97 to more than 70 percent in 2002/03.

Chapter 5 extends the analysis of interaction in the adoption of policy innovations among US school districts. Again, the focus is on the adoption of inter-district school choice. As in the Michigan case study, the variation in choice policies at the school district level is exploited by focussing on the districts' adoption decisions in states where participation is voluntary. However, the approach differs in several aspects from the one in Chapter 4. Firstly, the evidence on behavioral convergence in school choice policies is based on a unique data set providing information on district characteristics and policies in a large sample of districts covering five US states. Secondly, instead of modelling the interaction in adoption decisions as spatial dependence involving a time lag as in Chapter 4, the analysis allows for contemporaneous correlation in the districts' predispositions towards adoption of school choice.

Controlling for a large number of district characteristics describing local preferences and for spatial correlation in errors, the results indicate that the school districts' predispositions towards policies of open enrollment are strongly interdependent. Districts which are exposed to a composite neighbor with a strong predisposition towards adoption are significantly more likely to participate. Thus, similarly situated school districts indeed tend to affect each other in the decision whether to adopt new political technologies. The results of the chapter thus extend and confirm the evidence on positive spatial correlation in school choice policies among US school districts.

Chapter 6 is devoted to the empirical analysis of charter school policies. Charter schools are public schools operating independently from the existing school district structure. They are exempt from many state and district regulations and provide school officials and teachers with additional professional opportunities. By providing additional choice for parents and students, charter schools also tend to increase competition in the public school system.

As in the previous chapters, the focus is on district level policies. The aim of the chapter is to provide evidence on strategic interaction among local school districts in setting their charter school policies. The evidence is based on charter school policies of California school districts after the state legislature enacted a charter school law in 1992. The empirical investigation is based on two different approaches. The first approach is a finite spatial lag model based on the spatial autoregressive model proposed by Case (1992) and discussed in Section 3.3. Several extensions of the original model are proposed. Firstly, the model is applied to panel data of innovations instead of pure cross-section data. This makes it possible to model both dynamic and spatial effects and allows to account for unobserved heterogeneity. Secondly, a finite spatial lag version of the spatial autoregressive model is used. This, in contrast to the original model, allows to experiment with various spatial structures. The second approach is a spatial instrumental variables probit model as described in Section 3.5. In contrast to the finite spatial lag model, the instrumental variables probit is not applied to panel data, but to cross-sections of the California school districts.

The chapter adds to the existing literature not only by proposing empirical models which have not previously been used to estimating spatial effects in limited dependent variable models, but also by suggesting a new metric of neighborliness for local jurisdictions. The metric accounts for the small-scale mobility of workers in terms of district-to-district commuting flows. The intuition for this metric is simple: For each district, a set of reference districts is defined. It consists of all districts which are in sufficiently similar commuting distance to the average commuter workplace. For the reference districts, the degree of neighborliness is determined according to differences in median household income. The proposed metric clearly outperforms metrics relying on simple measures of geographical proximity. Most importantly, the impact of neighbors on actual district policies towards charter schools is estimated much more precisely.

The evidence provided in Chapter 6 suggests that the diffusion of charter schools as a new form of public schooling provision is heavily affected by positive interdependencies among the districts' decisions to establish and operate charter schools. For instance, based on the instrumental variables probit, a one percentage point increase in the share of neighbors operating at least one charter school is estimated to increase the original district's probability to run charter schools by 0.43 to 0.59 percentage points.

When modelling the mechanisms setting incentives for local governments to adopt public sector innovations and to implement political best-practice technologies, one can either focus on the political process or on (re-)location decisions of households and firms. Both traditional models with Tiebout choice and models focussing on the political process in a setting with asymmetric information between local governments and residents imply adoptions of policy innovations to show positive spatial correlation. The main contribution of this book is to show that the interaction among local governments with regard to public sector innovation is not only a prediction of theoretical models, but a powerful real world phenomenon.

Chapter 2

Theory: Yardstick competition and public sector innovation

2.1 Introduction

The political agency problem between voters and their elected representatives is of fundamental importance. The core of this problem is a severe informational asymmetry: Representatives are better informed than voters about the prospects of all kind of public projects as well as the costs of providing public goods and services. Furthermore, the ability of voters to monitor their representatives is limited, too. Thus investments in wasteful projects, favors to special interests and rent-seeking by politicians often cannot be distinguished from benevolent and honest political activities under unfavorable circumstances. Repeated elections are the basic means the electorate can use to sort the bad representatives from the good. But given that it is difficult for voters to assess the performance of their representatives and to distinguish between bad performance and bad luck, elections alone clearly do not work well as a discipline device for politicians. In decentralized political systems, however, voters may base their decision at the ballot box on comparative rather than absolute performance (or both). In correlated environments, inference on the quality of a jurisdiction's performance will be more precise if it is based on a performance comparison across several jurisdictions. In their strategic interaction with the electorate, representatives will anticipate the comparative performance evaluation and adjust their behavior.

In this chapter, the role of yardstick competition in the diffusion of public sector innovations is addressed from a theoretical perspective. I offer a simple model with two jurisdictions where governments face the alternative to choose between an old and a new policy with stochastic payoffs. There are two types of governments, benevolent ('good') and rent-seeking ('bad'). Running the new policy is costly in terms of additional effort needed to perform it, and rent-seeking governments therefore have an incentive to run the old policy irrespective of the payoffs generated by both policies. The model is used to illustrate that yardstick competition between governments of local jurisdictions may be a driving force for the diffusion of policy innovations in decentralized political systems. The intuition for this result is straightforward: Under yardstick competition it is more difficult for rent-seeking governments to maintain a sufficiently good reputation to be reelected. If it is optimal to use the new policy from a social perspective, benevolent governments will do so, and rent-seeking governments will be forced to follow in order to protect their reputation.

Of course, comparative performance evaluation is costly for voters in terms of collecting and processing information. Residents in any given jurisdiction can therefore be expected to use only a limited number of reference jurisdictions as a yardstick to evaluate policies in their own jurisdiction.[1] Hence, the model predicts yardstick competition between local jurisdictions' governments to make local decisions in the adoption of policy innovations depending on expected policies in a composite 'neighboring' jurisdiction. Empirically,

[1]Reference districts may be determined by the degree of neighborliness such as spatial proximity, but other criteria like the degree of similarity of social and economic conditions or factor mobility between jurisdictions may also play a role.

this would be reflected in public sector innovations being positively spatially correlated across jurisdictions.

The chapter is related to a number of theoretical papers on political agency problems. In one of the fundamental contributions in the field, Besley and Case (1995) introduced the idea of yardstick competition between jurisdictions. They show how in correlated environments the asymmetric information problem between politicians and voters can be alleviated by comparative performance evaluation across jurisdictions. Other contributions that deal with the role of yardstick competition in political agency problems include Besley and Smart (2002), who discuss the effect of yardstick competition on public good provision and wasteful spending, Belleflamme and Hindriks (2005), extending the contribution of Coate and Morris (1995) on inefficient rent-seeking to a multi-agency framework, and Wrede (2001), who provides a model with rent-seeking politicians and yardstick competition in a more richly modelled political system. This chapter extends the model of Besley and Case (1995) to allow for strategic interaction between jurisdictions with respect to policy innovations as discrete choice decisions. Kotsogiannis and Schwager (2006b) discuss the effect of outside information on the incentives to experiment with innovative public policies. They show that outside information is a two-sided edge: it is beneficial since it helps the voter to separate selfish from benevolent politicians, but it also creates an externality that reduces the incentives to engage in policy experiments. In a recent contribution, Kotsogiannis and Schwager (2006a) discuss the effect of career concerns of politicians in federal systems on their willingness to experiment with new policies. In another theoretical paper, Bolton and Harris (1999) deal with strategic interaction among players who must divide their time between safe and risky actions. Agents can learn from the current experimentation of others, giving rise to a free-rider problem. On the other hand, agents may be encouraged to experiment more if, by doing so, they can bring forward the time at which the beneficial effects from increased experimentation of all agents become available. Mukand and Rodrik (2002) provide a model of policy choice in 'follower' countries, where governments can either choose a policy which was successfully implemented in some other country or they can experiment and rely on a private signal about the appropriate policy in their own country. The effects of the informational externality created by successful leaders depends on the degree of similarity between leader and follower: Followers whose state is close to the leader benefit from mimicking the leader. In the 'near periphery' this causes countries to adopt policies which are often inappropriate to their circumstances. In contrast, countries in the 'far periphery' tend to experiment on their own since the informational externalities are too weak to outweigh the costs associated with choosing inappropriate policies. Tyran and Sausgruber (2005) report on an experimental study on policy experimentation and policy diffusion. The experimental design allows to discriminate between policy experimentation, experience, and policy emulation as factors driving the diffusion process. The main result of the study is that experimentation by itself leads to an inefficiently low innovation rate, and that policy emulation, by significantly contributing to the spread of the policy innovation, is efficiency enhancing.

The remainder of the chapter is organized as follows. In Section 2.2, the theoretical framework is presented. Section 2.3 is devoted to the discussion of the yardstick competition equilibrium and the reference equilibrium without yardstick competition. An extension of the basic model with jurisdiction-specific and common policies in presented in Section 2.4. Finally, the empirical implications of the model are derived and discussed in Section 2.5.

2.2 Incentives for public sector innovation: A theoretical model

Consider two jurisdictions $i = 1, 2$. Each jurisdiction is populated by a representative resident and has a government which performs a public policy p_i generating a payoff π_i. Governments face the alternative to choose between an old policy o and a new policy n. Assume that a special effort $e > 0$ is needed to perform the new policy.[2] Governments are either benevolent or rent-seeking. Rents take the form of transfers τ_i governments make to themselves. Bad governments' per-period utility $u_i(p_i, \tau_i)$ is $\tau_i - e$ if $p_i = n$ and τ_i if $p_i = o$.

The model has two time periods, where utility derived in the second period is discounted with the discount factor $\delta \in (0, 1)$. The timing of events is as follows. At the beginning of period 1, Nature draws three random variables $s \in \{l, h\}$ and $S_i \in \{L, H\}$, $i = 1, 2$. These variables determine the payoffs governments can produce by selecting a policy. They will be discussed shortly. In addition, Nature draws for each jurisdiction the type of an incumbent government $I_i \in \{G, B\}$, where G stands for 'good' and B for 'bad', i.e. rent-seeking. After all random variables have been drawn, incumbents in both jurisdictions simultaneously choose a pair (p_i, τ_i). Policy payoffs are realized, and the utility $v_i = \pi_i - \tau_i$ is delivered to the voters. At the end of period 1, elections take place. In each jurisdiction, the resident either reelects the incumbent or chooses a challenger who is drawn from the same distribution as incumbents. The voting is retrospective, and resident-voters care about maximizing their period-2 utility. In period 2, governments (either reelected incumbents or newly elected challengers) once again choose a policy and a transfer to themselves.

---

[2] One could see e as the effort needed to perform the new policy in excess of the effort needed to perform the old policy.

Table 2.1: Policy payoffs

Superior policy	old policy ($s = l$)		new policy ($s = h$)	
Jurisdiction-specific shock	negative ($S = L$)	positive ($S = H$)	negative ($S = L$)	positive ($S = H$)
Payoff old policy	π	$\pi + \Delta$	π	$\pi + \Delta$
Payoff new policy	$\pi - 2\Delta$	$\pi - \Delta$	$\pi + 2\Delta$	$\pi + 3\Delta$

The policy payoffs are determined as follows. Depending on the outcome for the jurisdiction-specific random variable S_i, the payoff from the old policy is either π (if $S_i = L$) or $\pi + \Delta$ (if $S_i = H$). The payoff from the new policy is jurisdiction-specific and is jointly determined by s and S_i. The outcome for s determines, jointly for both jurisdictions, whether the new policy is superior to the old policy. If $s = h$, then the new policy generates higher payoff than the old policy, and payoffs are $\pi + 2\Delta$ (if $S_i = L$) or $\pi + 3\Delta$ (if $S_i = H$). If $s = l$, the new policy generates lower payoff than the old policy, and payoffs are $\pi - 2\Delta$ (if $S_i = L$) or $\pi - \Delta$ (if $S_i = H$). Table 2.1 gives an overview of payoffs.

Let θ be the probability that the old policy is superior, i.e. that $s = l$, and γ be the probability that a good incumbent is drawn in jurisdiction i. The draws of I_1 and I_2 are independent. The degree of correlation among the shocks S_1 and S_2 is measured by the parameter σ. Denoting by $\Pr(X, Y)$ the joint probability that jurisdiction i is hit by a shock $S_i = X$, while j is hit by a shock $S_j = Y$, we have

$$\Pr(L, L) = q\sigma; \quad \Pr(L, H) = \Pr(H, L) = q(1-\sigma); \quad \Pr(H, H) = 1-q(2-\sigma).$$

Thus q is a jurisdiction's (unconditional) probability of experiencing a shock $S = L$. Since the model is not meant to deal with the case of negative correlation between S_1 and S_2, let σ be restricted by $\sigma \in [q, 1]$. With $\sigma = 1$ we have perfect correlation among shocks, whereas with $\sigma = q$ we have independence.[3]

To simplify the analysis, the transfer is restricted to $\tau \in \{0, \Delta, \bar{\tau}\}$, where $\bar{\tau}$ is the maximal transfer.[4] In addition, let $\bar{\tau} > \Delta > e$ as well as $\bar{\tau} < 3\Delta$. The former assumption means that the model allows for transfers exceeding the difference between the high and the low payoff given a realization of s and a certain policy, and that choosing the new policy and taking a transfer that equals this difference generates a positive per-period utility for incumbents. To get an intuition for the latter assumption, consider two situations in which the new policy is being implemented. In the first situation, let the new policy also be the superior policy. Note that in this situation the lowest possible per-period utility of the voter is $\pi + 2\Delta - \bar{\tau}$. Now consider a second situation where the new policy is being implemented despite the fact that the old policy is superior. Note that in this situation the highest possible per-period utility is $\pi - \Delta$. The assumption $\bar{\tau} < 3\Delta$ guarantees that, if the new policy is being used, it will always be possible for the voter to learn about the superior policy.

To complete the description of the model, the distribution of information has to be specified. All underlying distributions are common knowledge. The draws of s and I_i, $i = 1, 2$ are revealed to both jurisdictions' incumbents. Thus each incumbent knows which policy is superior, its own type and the type of the other jurisdiction's incumbent. As mentioned by Besley and Case (1995), who use the same assumptions with regard to the knowledge of incumbents' types, this is a bit too strong. However, for the main implication of the model to go through it would be sufficient to assume that incumbents know more about each other than voters do. This seems to be reasonable, given that decision makers interact with each other to some extent. It is also worth noting that the empirical implication of the model with regard to the spatial

---

[3]This density has been used by Bordignon, Cerniglia, and Revelli (2004).

[4]The assumption that the choice of transfers is discrete is not restrictive. With $\tau \in [0, \bar{\tau}]$ instead the crucial properties of the equilibria discussed in the following are the same.

distribution of policy innovations critically depends on this assumption.

Furthermore, each incumbent observes the realization of S_i in his own jurisdiction. Thus, when simultaneously choosing policies and transfers, incumbents in both jurisdictions are perfectly informed about the payoff their own policy will generate. With regard to payoffs in the other jurisdiction, incumbents know which policy is superior and the distribution of payoffs conditional on the incumbent's choice. Residents in both jurisdictions do not know neither incumbents' types nor the realizations of s and S_i, $i = 1, 2$. What residents observe is the policy in their own jurisdiction and their own utility, i.e. the policy payoff net of the transfer taken by the incumbent. If we allow for yardstick competition, residents also observe the policy and the resident's utility in the other jurisdiction.

2.3 Analysis of equilibria

2.3.1 The case of a single jurisdiction

Without yardstick competition, there is no link between both jurisdictions. The Perfect Bayesian Equilibrium described in proposition 1 is thus an equilibrium in the behavior of incumbents and resident-voters if we look at each of both jurisdictions separately. Denote the strategies of incumbents[5] by $\mu[s, S_i, I_i] = (p_i, \tau_i)$. The strategies of residents as voters are given by $\varsigma(p_i, v_i) \in [0, 1]$, denoting the probability that they reelect an incumbent who sets policy p_i and delivers utility v_i. As usually, along the equilibrium path the underlying beliefs with regard to the incumbent's type are tied down by the requirement

[5]Strictly speaking this labelling is not correct, since second period choices of governments are not described. These are trivial, however, and the chapter follows the literature in suppressing them from what is called incumbents' strategies.

that the beliefs must be computed from the incumbent's strategy via Bayes's rule.

With respect to out-of-equilibrium beliefs of voters the following simple rule is specified: Whenever it is apparent that a transfer has been taken and/or it is apparent that the inferior policy has been chosen in i, the voter in i believes that a bad incumbent holds office.[6] As in many other signalling games, different out-of-equilibrium beliefs may support other, often rather unnatural equilibria. Coate and Morris (1995) provide a discussion of the issue in a related context as well as a simple monotonicity property for out-of-equilibrium beliefs to support only 'reasonable' equilibria. For notational convenience, define the sets $C_i = \{(n, \pi + 3\Delta), (n, \pi + 2\Delta), (o, \pi + \Delta), (o, \pi)\}$ and $D_i = C_i \backslash \{(o, \pi)\}$, $i = 1, 2$. C_i is the set of all pairs (p_i, v_i) which, for some draw of s and S_i, contain as elements the superior policy and the corresponding utility of the voter given a zero transfer.

Note that good incumbents'strategies are such that they always choose the superior policy together with a zero transfer. For the sake of brevity, the formal description of their behavior is suppressed in the following propositions. Note also that the index for jurisdictions has been dropped for the remainder of this subsection.

Before presenting Proposition 1, it is useful to state the following definitions:

Definition 1 *Define* $\theta^* \equiv (1-q)/q$, $\delta^* = 1 - \Delta/\bar{\tau}$ *and suppose* $q \geq 1/2$, $\theta \geq \theta^*$ *and* $\delta \geq \delta^*$.

Note that with $q \geq 1/2$ it is guaranteed that $\theta^* \leq 1$.

---

[6]Under proposition 1, pairs (p, v) which do not occur on the equilibrium path will necessarily reveal that either a transfer has been taken or the inferior policy has been chosen, or both.

Proposition 1 *Suppose that, on the equilibrium path, beliefs are formed according to Bayes' rule and that off the equilibrium path beliefs are formed as described above. With parameters as given in Definition 1, the following strategies together with the specified beliefs constitute a Perfect Bayesian Equilibrium:*

Bad incumbents choose the old policy in all cases. The transfer equals Δ *if* $S = H$ *and* $\bar{\tau}$ *if* $S = L$*. The representative voter reelects the incumbent if the new policy is chosen and the voter's utility is either* $\pi + 3\Delta$ *or* $\pi + 2\Delta$*, and if the old policy is chosen and the voter's utility is either* $\pi + \Delta$ *or* π*. In all other cases, the incumbent is voted out of office and replaced by the challenger.*

Formally, the strategies can be stated as follows:

Bad incumbents set

$$(i) \quad \mu(h, H, B) = \mu(l, H, B) = (o, \Delta),$$

$$(ii) \quad \mu(h, L, B) = \mu(l, L, B) = (o, \bar{\tau}),$$

and the representative voter sets

$$(iii) \quad \varsigma[(p, v) \in C] = 1,$$

$$(iv) \quad \varsigma[(p, v) \notin C] = 0.$$

Proof. See the appendix to this chapter.

The intuition for this equilibrium is simple. Bad governments want to extract the highest possible transfer in the first period, but at the same time they seek reelection. With a sufficiently high discount factor, a mimicking behavior in the case of $S = H$ becomes worthwhile: The cost of a reduced transfer in the first period is outweighed by the benefit of winning reelection and being able to extract the maximum transfer in the second period. Since using the new policy is costly in terms of additional effort, bad governments prefer the old policy in all cases. For the voter, reelecting governments pretending that the old policy is superior and that payoffs are low is optimal as long as the corresponding probabilities are sufficiently high.

2.3.2 The case of two jurisdictions: Yardstick competition

For the case with yardstick competition, denote the strategies of the incumbent in i by $\mu(s, S_i, I_i; I_j) = (p_i, \tau_i)$ and the strategies of the voter in i by $\varsigma(p_i, v_i; p_j, v_j) \in [0,1]$, $i, j = 1, 2$, $i \neq j$. Before we can state the proposition describing the yardstick competition equilibrium, it is necessary to carefully specify the voters' beliefs off the equilibrium path. First of all, let us transfer the setting from the previous subsection and require that

(i) whenever, for some jurisdiction i, it is apparent that a transfer has been taken and/or it is apparent that the inferior policy has been chosen, voters in both jurisdictions believe that $I_i = B$.

In addition, we have to deal with the case that what the voter observes in his own jurisdiction is consistent with the incumbent's equilibrium strategy while the behavior of the other jurisdiction's incumbent is inconsistent with his equilibrium strategy. How should the beliefs be specified in this case? First of all note that with the strategies stated in the following proposition, for all combinations $(p_1, v_1), (p_2, v_2)$ off the equilibrium path, beliefs with respect to the type of the incumbent in at least one jurisdiction are given by (i). This being true, it is certainly the least restrictive way to complete the setting for the out-of-equilibrium beliefs by requiring that

(ii) in cases where the belief with respect to the type of incumbent in their own jurisdiction is not given by (i), voters form beliefs *as if* they performed a Bayesian updating based on the belief with regard to the type of the other jurisdiction's incumbent.

For notational convenience, define the set $E_i = \{(p_i, v_i) : p_i = n, v_i \in \{\pi+3\Delta, \pi+2\Delta, \pi+3\Delta-\bar{\tau}, \pi+\Delta, \pi+2\Delta-\bar{\tau}\}\}$, $i = 1, 2$. The crucial property of E_i is that after observing some $(p_i, v_i) \in E_i$, voters in both jurisdictions know that the new policy is superior.

Again it is useful to state some definitions before turning to the proposition:

Definition 2 *Define* $\delta^{**} = 1 - \frac{\Delta - e}{\bar{\tau}}$, $\sigma^{**} = \frac{3q-1}{2q}$ *and* $\gamma^{**} = \frac{1-q(2-\sigma+\theta(1-\sigma))}{1-q(2-\sigma+\theta(2-3\sigma))}$, *and suppose* $q > 1/2$, $\theta \geq \theta^*$, $\sigma \leq \sigma^{**}$, $\gamma \geq \gamma^{**}$ *and* $\delta \geq \delta^{**}$.

Note that $\sigma^{**} \geq q \; \forall \, q \in \left[\frac{1}{2}, 1\right]$, such that for all such q there exists some $\sigma \in \left[\frac{1}{2}, 1\right]$ for which $q \leq \sigma \leq \sigma^{**}$. Note furthermore that the denominator in γ^{**} is positive for all $\theta \in [\theta^*, 1]$ given that $q \in \left(\frac{1}{2}, 1\right]$ and $\sigma \in [q, \sigma^{**}]$, and that $\gamma^{**} \leq 1 \; \forall \, \sigma \geq 1/2$.

Proposition 2 *Suppose that, on the equilibrium path, beliefs are formed according to Bayes' rule and that off the equilibrium path beliefs are formed as described above. With parameters as given in Definition 2, the following strategies together with the specified beliefs constitute a Perfect Bayesian Equilibrium:*

Bad incumbents choose the new policy only if the following three conditions hold: The new policy is superior ($s = h$)*, the jurisdiction-specific shock is positive* ($S_i = H$)*, and the other jurisdiction's incumbent is good. In all other cases, the old policy is chosen. The transfer equals* $\bar{\tau}$ *if* $S_i = L$ *and* Δ *if* $S_i = H$*. The representative voter reelects the incumbent if the new policy is chosen and the voter's utility is either* $\pi + 3\Delta$ *or* $\pi + 2\Delta$*, and if the old policy is chosen and the voter's utility is either* $\pi + \Delta$ *or* π*, provided that the outcome in the other jurisdiction is not an element of* E*. In all other cases, the incumbent is voted out of office and replaced by the challenger.*

Formally, the strategies can be stated as follows:

Bad incumbents set

$$(i) \quad \mu(h, H, B; B) = (o, \Delta),$$

$$(ii) \quad \mu(h, H, B; G) = (n, \Delta),$$

$$(iii) \quad \mu(h, L, B; B) = \mu(h, L, B; G) = \mu(l, L, B; B) = \mu(l, L, B; G) = (o, \bar{\tau}),$$

$$(iv) \quad \mu(l, H, B; B) = \mu(l, H, B; G) = (o, \Delta),$$

and the representative voters sets

$$(v) \quad \varsigma[(p_i, v_i) \in \{(n, \pi + 3\Delta), (n, \pi + 2\Delta)\}; \cdot] = 1,$$

$$(vi) \quad \varsigma[(p_i, v_i) \in \{(o, \pi + \Delta), (o, \pi)\}; (p_j, v_j) \in E_j] = 0,$$

$$(vii) \quad \varsigma[(p_i, v_i) \in \{(o, \pi + \Delta), (o, \pi)\}; (p_j, v_j) \notin E_j] = 1,$$

$$(viii) \quad \varsigma[(p_i, v_i) \notin C_i; \cdot] = 0.$$

Proof. See the appendix to this chapter.

The crucial point in comparison with the behavior of incumbents described in Proposition 1 is part *(ii)* of bad incumbents' strategy. By now being able to compare policies in two jurisdictions, voters have additional information to base their reelection decision on. If the new policy is superior, a bad incumbent knowing that the government in the other jurisdiction is good is now forced to choose the new policy to gain reelection: The choice of a good government in the other jurisdiction will reveal that the new policy is superior, and running the old policy would therefore result in the government being removed from office. The voter thus benefits from the additional information by getting the superior policy more often. The selection properties, i.e. the probabilities of finding out a bad incumbent, are the same in both equilibria.

The condition on the discount factor δ is stronger for the yardstick equilibrium than for the equilibrium without comparative performance evaluation. The reason is that, if the other government is good and the new policy must be used for a successful mimicking, the additional effort e has to be invested to perform the public policy. Note, however, that the role of e is only to make the new policy, other things being equal, less attractive for bad governments than the old policy. Thus e might be arbitrarily small, and $\lim_{e\to 0}\delta^{**} = \delta^*$. If the new policy is superior, payoffs are high and the other government is bad, or if the old policy is superior and payoffs are high, bad governments imitate good governments in equilibrium by choosing (o, Δ) and delivering $v = \pi$. Consequently, a sufficiently high γ is necessary to make reelection the optimal choice for the voter after observing (o, π) in both jurisdictions.

The most interesting parameter is σ. The yardstick equilibrium is supported by values of $\sigma \in [q, \sigma^{**}]$, i.e. the positive correlation between S_1 and S_2 must be sufficiently low. The intuition is that it must be worthwhile for the voter to reelect an incumbent in a situation where (a) the same policy has been chosen in both jurisdictions, (b) the voter's utility in the other jurisdiction equals the maximum payoff for the policy chosen, and (c) the voter's own utility equals the maximum payoff for the policy chosen minus Δ. As a counterexample, consider the case with perfect correlation $[\sigma = 1]$ between S_1 and S_2 for some $q \in \left(\frac{1}{2}, 1\right)$. In this situation, it cannot be optimal for the voter in i to reelect an incumbent choosing $p_i = n$ and delivering, say, $v_i = \pi + 2\Delta$ given that $v_j = \pi + 3\Delta$ since it is apparent that a transfer has been taken.

2.4 Extension: Yardstick competition with common and jurisdiction-specific policies

In the following, a straightforward extension of the theoretical analysis will be presented. It covers asymmetric jurisdictions and a multi-dimensional policy space.

Consider the model from section 2.2 with the extension that we now have two tasks $k = A, B$. Suppose that A is a common task which has to be fulfilled in both jurisdictions and let B be a task which is specific for jurisdiction 2 and which is not a policy task in jurisdiction 1. Clearly, with regard to the specific task, yardstick competition between jurisdictions is impossible. However, yardstick competition may takes place with respect to the common task.

The government in $i = 1$ has to make the same choice as before. In $i = 2$ the government now has to choose two policies p_{2A} and p_{2B}. For each task, the choice is between an old and a new policy. Assume that task-specific state variables $s_k \in \{l, h\}$, $k = A, B$ are independently drawn from the same distribution. θ is now the probability that $s_k = l$ is drawn for task $k = A, B$. If $s_k = h$ [$s_k = l$], the new [old] policy is superior for task k. As in the model with a one-dimensional policy space, policy payoffs are either low or high depending on the draw for $S_i \in \{L, H\}$. Note that policy payoffs are jurisdiction-specific, but not task-specific. Together with the policies p_{2A} and p_{2B}, task-specific transfers τ_{2A}, $\tau_{2B} \in \{0, \Delta, \bar{\tau}\}$ must be chosen. The per-period utility of the government in $i = 2$ from performance in task k now is

$$u_{2k}(p_{2k}, \tau_{2k}) = \begin{cases} \tau_{2k} - e & \text{if} \quad p_{2k} = n \\ \tau_{2k} & \text{if} \quad p_{2k} = o \end{cases} \quad .$$

In the situation without yardstick competition little is changed compared to subsection 2.3.1. In $i = 1$, proposition 1 holds. For $i = 2$, it is easy to show that an equilibrium exists where

(i) bad governments choose (o, Δ) for both tasks if $S_1 = H$;

(ii) bad governments choose $(o, \bar{\tau})$ for both tasks if $S_1 = L$;

(iii) the voter opts for the challenger if it is apparent that a transfer has been taken or it is apparent that the inferior policy has been chosen for some task $k = A, B$, or both, and reelects the incumbent otherwise.

For the situation with yardstick competition, let the strategies of incumbents in $i = 1$ be denoted by $\mu(s_A, S_i, I_i; I_j) = (p_{1A}, \tau_{1A})$, while strategies of incumbents in $i = 2$ are denoted as $\mu(s_A, s_B, S_i, I_i; I_j) = (\mu_A, \mu_B)$, where $\mu_k = (s_k, S_i, I_i; I_j) = (p_{2k}, \tau_{2k})$, $k = A, B$. Assume out-of-equilibrium beliefs defined by analogy to the model with yardstick competition in a single common policy.

Definition 3 *Define* $\tilde{\sigma} = 1/(1+\theta^2)$, $\tilde{\gamma} = \frac{1-q(2-\sigma+\theta^2(1-\sigma))}{1-q[2-\sigma+\theta^2(1-\sigma)+\theta(1-\sigma(1+\theta))]}$, $\tilde{\theta} = [1-q(2-\sigma)]/[q(1-\sigma)]$ *and* $\tilde{\tilde{\theta}} = [(1-q)/(3q-1)]^{\frac{1}{2}}$ *and suppose*

(a) $q \geq 1/2$,

(b) $\theta \geq \max\{\tilde{\theta}, \tilde{\tilde{\theta}}\}$,

(c) $\tilde{\sigma} \leq \sigma \leq \sigma^{**}$,

(d) $\delta \geq \delta^{**}$,

(e) $\gamma \geq \tilde{\gamma}$ *if* $1 - q[2-\sigma+\theta^2(1-\sigma)+\theta(1-\sigma(1+\theta))] > 0$.

Before turning to the proposition, note that $\sigma \leq \sigma^{**}$ implies $\tilde{\theta} \leq 1$. Furthermore, $\theta \geq \tilde{\tilde{\theta}}$ guarantees that for all $q \in \left[\frac{1}{2}, 1\right]$ there exists some σ such that $1/(1+\theta) \leq \sigma \leq \sigma^{**}$ holds. With $\sigma \geq 1/(1+\theta)$ it is guaranteed that $\tilde{\gamma} \leq 1$. In addition, with $\theta \geq \tilde{\tilde{\theta}}$ we can be sure that for all $q \in \left[\frac{1}{2}, 1\right]$ there exists some $\sigma \in \left[\frac{1}{2}, 1\right]$ such that $\tilde{\sigma} \leq \sigma \leq \sigma^{**}$ holds.

Proposition 3 *Suppose that, on the equilibrium path, beliefs are formed according to Bayes' rule and that off the equilibrium path beliefs are formed as described above. With parameters as given in Definition 3, the following strategies together with the specified beliefs constitute a Perfect Bayesian Equilibrium:*

Bad incumbents choose the old policy for the jurisdiction-specific task in all cases. The new policy is chosen for the common task if the following three conditions hold: The new policy is superior for the common task ($s_A = h$), the jurisdiction-specific shock is positive ($S_2 = H$), and the other jurisdiction's incumbent is good. In all other cases, the old policy is chosen for the common task. The transfer equals $\bar{\tau}$ (for all tasks) if $S_i = L$ and Δ (for all tasks) if $S_i = H$. Voters in $i = 1$ ($i = 2$) vote for the challenger whenever (for some task $k = A, B$) it is apparent that a transfer has been taken or if it is apparent that the inferior policy has been chosen in i, or both. In all other cases, the incumbent is reelected.

Formally, the strategies of bad incumbents can be stated as follows:

Bad incumbents in $i = 1$ follow their strategies described in proposition 2. Bad incumbents in $i = 2$ set

$$(i) \quad \mu_k(h, H, B; B) = (o, \Delta), \quad k = A, B$$

$$(ii) \quad \mu_A(h, H, B; G) = (n, \Delta),$$

$$\mu_B(h, H, B; G) = (o, \Delta),$$

$$(iii) \quad \mu_k(l, H, B; B) = \mu_k(l, H, B; G) = (o, \Delta), \quad k = A, B$$

$$(iv) \quad \mu_k(h, L, B; B) = \mu_k(h, L, B; G) =$$

$$\mu_k(l, L, B; B) = \mu_k(l, L, B; G) = (o, \bar{\tau}), \quad k = A, B.$$

Proof. See the appendix to this chapter.

From proposition 3 we see that the role of yardstick competition as a discipline device for rent-seeking politicians carries over from the simple to a richer model with various political tasks of governments and varying degrees of yardstick competition across tasks. However, in a model with only a subset of tasks being subject to yardstick competition the pattern of government performance becomes richer. Even benevolent governments now switch from innovative to traditional policies across tasks. Moreover, rent-seeking governments can stick to traditional policies and avoid innovations for jurisdiction-specific tasks where comparative performance evaluation by voters is impossible. This gives rent-seeking governments additional possibilities to masquerade as acting in the interest of the voter. For common tasks the same discipline-enhancing effect of comparative performance evaluations by voters is at work: if a new policy is shown to be superior by successfully innovating good governments in the neighborhood, bad governments seeking reelection are forced to use the innovation, too.

2.5 Implications

The model presented above illustrates how yardstick competition between governments of local jurisdictions can drive the diffusion of public sector innovations among jurisdictions. Although the model is highly stylized, it captures some crucial aspects of decentralized policy making and the adoption of innovative policies. Most importantly, politicians are better informed about the prospects of new policies than the citizens. At the same time, rent-seeking politicians tend to avoid the additional effort needed to implement new policies. By comparing the performance of local governments across jurisdictions, citizens learn about political best-practice technologies, and office-motivated governments are forced to implement successful new policies more often compared to a situation without comparative performance evaluation across jurisdictions.

However, the increased probability for optimal policies being implemented comes at a cost. In order to learn about political best-practice technologies, citizens must collect and review information about the performance of several local governments. Since the costs will increase with the number of jurisdictions under monitoring, residents can be expected to use only a limited number of reference jurisdictions as a yardstick for policies in their own jurisdiction. Since the equilibrium with yardstick competition predicts local adoption decisions to positively depend on expected adoptions in reference jurisdictions, an empirical implication of the model is positive correlation in the adoption of public sector innovations across jurisdictions belonging to the same reference group.

The empirical implications of the extended model are the following: Firstly, the share of tasks where innovative policies are used should on average be higher with yardstick competition than without. Hence, governments should implement innovative policies more often if policy competition is stronger. Secondly, the share of tasks where governments use innovative policies should show spatial correlation. Finally, the model suggests that the degree of spatial correlation in policy innovations varies across policies with the intensity of policy competition.

Appendix to Chapter 2: Proofs of propositions

Proof of proposition 1:

We begin by checking for profitable deviations of the voter. Since good incumbents maximize the voter's utility and bad incumbents according to the proposed strategy never play in a way such that $(p, v) \in D$, reelecting the incumbent after observing $(p, v) \in D$ and not reelecting after observing $(p, v) \notin C$ is optimal for the voter. Let $\Pr(I = G|p, v)$ denote the voter's beliefs. After observing (o, π), the probability of a good government holding office is $\Pr(I = G|o, \pi) = \frac{\gamma\theta q}{\gamma\theta q+(1-\gamma)(1-q)}$. This is at least as high as γ, the probability of drawing a good challenger, if $\theta \geq \theta^*$.

Now we have to check for profitable deviations of bad incumbents. For a bad incumbent experiencing $S = L$ playing $(o, \bar{\tau})$ dominates any other action: Reelection can only be gained by taking a zero transfer in the first period, which cannot be optimal given that the utility from taking the maximum transfer in the second period is $\delta\bar{\tau}$. Since choosing the new policy is costly, it can also not be optimal to play $p = n$ and take some transfer $\tau \in \{0, \Delta, \bar{\tau}\}$. If $S = H$, a bad government receives utility $\Delta + \delta\bar{\tau}$ by following the proposed strategy. As before, playing $p = n$ instead cannot be optimal since $e > 0$. Finally, given that $\delta \geq \delta^*$, a deviation to $(o, \bar{\tau})$ is not profitable.

Proof of proposition 2:

Once again we begin by checking for profitable deviations of the voter. Let $\Pr(I_i = G|p_i, v_i; p_j, v_j)$ denote beliefs of voters. As before, if the voter observes $(n, \pi+3\Delta)$ in his own jurisdiction, he believes a good incumbent to hold office with probability one. For $(p_i, v_i) = (n, \pi+2\Delta)$, we have to consider three cases. Firstly, $\Pr(I_i = G|n, \pi + 2\Delta; n, \pi + 3\Delta) = \frac{\gamma q(1-\sigma)}{\gamma q(1-\sigma)+(1-\gamma)(1-q(2-\sigma))}$. This is at least as high as γ since $\sigma \leq \sigma^{**}$. Secondly, $\Pr(I_i = G|n, \pi + 2\Delta; n, \pi + 2\Delta) = \frac{\gamma\sigma+(1-\gamma)(1-\sigma)}{\gamma\sigma+2(1-\gamma)(1-\sigma)}$, which is at least as high as γ given that $\sigma \geq 1/2$. Thirdly, after observing $(n, \pi + 2\Delta; (p_j, v_j) \notin \{(n, \pi + 3\Delta), (n, \pi + 2\Delta)\})$, reelecting the incumbent is optimal since by applying Bayes' Rule and – if necessary – the out-of-equilibrium beliefs, in all cases we find that the probability of a good incumbent holding office is one.

Part (vi) of the voter's strategy reflects out-of-equilibrium beliefs. Since the observation of each of the elements in E_j reveals that the new policy is superior, it is optimal not to reelect a government in $i \neq j$ choosing $p_i = o$.

With regard to part (vii) of the voters strategy, first note that bad governments according to the proposed strategies never play $(o, \pi + \Delta)$. As long as $(p_j, v_j) \notin E_j$, it is therefore optimal to reelect the incumbent in $i \neq j$ after observing $(p_i, v_i) = (o, \pi+\Delta)$. Furthermore, note that $\Pr(I_i = G|o, \pi; o, \pi+\Delta) = \Pr(I_i = G|n, \pi + 2\Delta; n, \pi + 3\Delta)$. Thus it is optimal to reelect the incumbent in i after observing $(o, \pi; o, \pi + \Delta)$ given that $\sigma \leq \sigma^{**}$. If voters observe (o, π) in both jurisdictions, the probability of a good government holding office is $\frac{\gamma^2\theta q\sigma+\gamma(1-\gamma)\theta q(1-\sigma)}{\gamma^2\theta q\sigma+2\gamma(1-\gamma)\theta q(1-\sigma)+(1-\gamma)^2(1-q(2-\sigma))}$. This is at least as high as γ given that $\gamma \geq \gamma^{**}$. If $(o, \pi; o, \pi - \bar{\tau})$ is observed, the probability for the voter in i of having a good government is $\frac{\gamma\theta\sigma}{\gamma\theta\sigma+(1-\gamma)(1-\sigma)}$. This is at least as high as γ for all $\theta \in [\theta^*, 1]$ given that $\sigma \geq q$. Finally, applying Bayes' Rule together with the out-of-equilibrium beliefs specified above shows that for all remaining $(p_j, v_j) \notin E_j$ it is optimal to reelect the incumbent after observing (o, π) in i.

Part $(viii)$ of the voters strategy is optimal since by observing $(p_i, v_i) \notin C_i$ it is revealed that either a transfer has been taken or the inferior policy has been chosen in i, or both.

Now we check for profitable deviations of bad incumbents. Given that bad governments facing a bad government in the neighboring jurisdiction never implement the new policy, yardstick competition does not provide the voter with additional information in this case compared to the situation with only a single jurisdiction, and policy choices and transfers identical to those described in proposition 1 together with the specified beliefs and the proposed voting rule constitute an equilibrium in the presence of yardstick competition. If the other government is good, however, it is not longer optimal for a bad incumbent in i to choose (o, Δ) if $s = h$ and $S_i = H$ since the superior new policy will be used in j, leading the voter in i to vote for the challenger. Instead, it is now optimal for i's incumbent to set (n, Δ), to deliver $v_i = \pi + 2\Delta$ to the voter and gain reelection: Since $\delta \geq \delta^{**}$, the utility $\Delta - e + \delta\bar{\tau}$ is at least as high as the one from the best alternative, $(o, \bar{\tau})$. If $s = l$ and $S_i = H$, (o, Δ) remains the optimal choice since given the proposed voting rule the probability of reelection is one. If $S_i = L$ and the other government is good, the same reasoning as in the proof of proposition 1 applies.

Proof of proposition 3:

I will be brief since the proof is similar to the one for proposition 2. We start with the voter in $i = 2$. Let $\Pr[I_2 = G|(p_{2A}, v_{2A}), (p_{2B}, v_{2B}); (p_{1A}, v_{1A})]$ denote his beliefs. Consider the following five cases on the equilibrium path:

(I) $\Pr[I_2 = G|(n, \pi + 2\Delta), (o, \pi); (n, \pi + 3\Delta)] = \frac{\gamma\theta q(1-\sigma)}{\gamma\theta q(1-\sigma)+(1-\gamma)(1-q(2-\sigma))}$. This is at least as high as γ given that $\theta \geq \tilde{\theta}$.
(II) $\Pr[I_2 = G|(n, \pi + 2\Delta), (o, \pi); (n, \pi + 2\Delta)] = \frac{\gamma\theta\sigma+(1-\gamma)\theta(1-\sigma)}{\gamma\theta\sigma+(1-\gamma)(1+\theta)(1-\sigma)}$. This is at least as high as γ given that $\theta \geq \frac{\gamma(1-\sigma)}{\gamma\sigma+(1-\gamma)(1-\sigma)}$, which is implied by $\theta \geq \tilde{\theta}$.
(III) $\Pr[I_2 = G|(o, \pi), (o, \pi); (o, \pi + \Delta)] = \frac{\gamma\theta q(1-\sigma)}{\gamma\theta q(1-\sigma)+(1-\gamma)(1-q(2-\sigma))}$. See (I).
(IV) $\Pr[I_2 = G|(o, \pi), (o, \pi); (o, \pi)] = \frac{\gamma^2\theta^2 q\sigma+\gamma(1-\gamma)\theta^2 q(1-\sigma)}{\gamma^2\theta^2 q\sigma+\gamma(1-\gamma)\theta(1+\theta)q(1-\sigma)+(1-\gamma)^2(1-q(2-\sigma))}$. This is at least as high as γ if the denominator in $\tilde{\gamma}$ is smaller than or equal to zero given that $\sigma \geq 1/(1 + \theta)$, which is implied by $\sigma \geq \tilde{\sigma}$. If the denominator in $\tilde{\gamma}$ is strictly positive we have $\Pr[I_2 = G|(o, \pi), (o, \pi); (o, \pi)] \geq \gamma$ since $\gamma \geq \tilde{\gamma}$.
(V) $\Pr[I_2 = G|(o, \pi), (o, \pi); (o, \pi - \bar{\tau})] = \frac{\gamma\theta^2\sigma}{\gamma\theta^2\sigma+(1-\gamma)(1-\sigma)}$. This is at least as

high as γ given that $\sigma \geq \tilde{\sigma}$.

In all remaining cases the optimality of the voting rule is immediately obvious from Bayes' Rule or given by the out-of-equilibrium beliefs. By repeating the same procedure for the voter in $i = 1$ it is easy to show that the proposed voting rule is optimal given that the conditions specified above hold.

For the incumbent in $i = 1$ the situation is basically unchanged compared to the situation with an identical neighborhood jurisdiction. He must be willing to invest e in order to gain reelection. Since $\delta \geq \delta^{**}$ this is the case. The incumbent in $i = 2$ must prefer the new policy for task A if $s_A = h$, $S_2 = H$ and $I_1 = G$. This requires $\delta \geq 1 - (\Delta - \frac{1}{2}e)/\tilde{\tau}$, which is implied by $\delta \geq \delta^{**}$.

Chapter 3

Spatial effects in limited dependent variable models

3.1 Introduction

Compared to the vast theoretical and applied literature on spatial effects in linear models,[1] the number of contributions on limited dependent variable models with spatial effects is rather limited.[2] One reason is that the estimation of the standard model, an autoregressive spatial probit, is computationally involved. Standard econometric software packages do not allow to estimate the spatial probit, and constructing own procedures requires considerable programming effort. Secondly, estimation of the spatial probit relies on maximum likelihood (ML) techniques, whereas the literature on linear spatial models shows a distinct preference for instrumental variables (IV) and general method of moments (GMM) approaches. Both IV and GMM methods have

[1] Anselin (1988) and the volumes "New directions in spatial econometrics", edited by L. Anselin and R.J.G.M. Florax in 1995 and "Advances in spatial econometrics, edited by L. Anselin, R.J.G.M. Florax and S.J. Rey in 2004 may serve as introductory texts.

[2] See Beron and Vijverberg (2004), Pinkse, Shen, and Slade (2003), Hautsch and Klotz (2003), LeSage (2000), Pinkse and Slade (1998), Dubin (1995), McMillen (1995, 1992), and Case (1992).

certain advantages compared to ML, in particular in terms of the robustness of parameter estimates with regard to spatial error dependence.

However, processes with discrete outcomes are an important real world phenomenon, and it is crucial to develop and improve on existing analytical techniques suitable for analyzing spatial effects in processes involving discrete choice. In particular, the analysis of spatial effects in public policies with discrete outcomes requires appropriate empirical techniques. In the following, a short introduction to various empirical approaches is given. However, the discussion is not meant to be exhaustive. In particular, the chapter does not cover Bayesian approaches to the estimation of limited dependent variable models. The interested reader is referred to LeSage (2000). Rather, the purpose of the chapter is to give the reader an overview of the empirical models which are used in Chapters 4 to 6.

3.2 The spatial linear probability model

The spatial linear probability model (LPM) is certainly the simplest framework that can be used to estimate a spatial limited dependent variable model. The structural equation is

$$y_i = \phi y_{-i} + x_i \beta + u_i, \tag{3.1}$$

where y_i are outcomes, y_{-i} is the outcome of a composite neighbor of i, x_i is a $K+1$ row vector of control variables where the first element is unity, ϕ and β are (vectors of) coefficients to be estimated, and u_i is an error term. It is standard in applications incorporating spatial effects to write the outcome of the composite neighbor as a linear combination,

$$y_{-i} = \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, y_j, \tag{3.2}$$

where the w's are weights defined according to some metric of neighborliness among the units of observation such that $\sum_{j \neq i} w_{ij} = 1$. The normalization of the sum of weights is usually done in spatial models to ensure that the potential overall impact of the composite neighbor is the same for all units of observation. Without this normalization, it would in most applications be difficult to give a clear interpretation to the coefficient ϕ.

The limitations of the model are obvious.[3] Nevertheless, as a simple check for results derived from more elaborate models, the LPM is a very useful tool that should not be dismissed.

Since the LPM is a linear model, the standard approaches to deal with the problems of endogeneity of y_{-i} and spatial correlation in the errors u_i can be used in estimation.[4] For instance, one can apply 2SLS, where instruments are derived as described in Section 3.5.

3.3 The spatial autoregressive model

Consider a limited dependent variable model where a latent variable y_i^* determines actual outcomes y_i according to

$$y_i = 1[y_i^* > 0], \quad i = 1, \ldots, N, \tag{3.3}$$

where $1[\cdot]$ is the indicator function. A structural spatial auto-regressive (SAR) model for the latent variable is specified as

$$y_i^* = \phi\, y_{-i}^* + x_i \beta + u_i, \tag{3.4}$$

where y_{-i}^* is the latent variable of a composite neighbor of i, x_i is again a $K+1$ row vector of control variables with the first element being unity, and

[3]See Wooldridge (2002), p. 454-57 for an extensive discussion of the LPM.

[4]These are discussed in more detail in the following sections.

u_i is an i.i.d. error distributed symmetrically about zero with variance σ^2. To guarantee stability assume that $|\phi| < 1$. In this model, a positive ϕ would mean that the latent variables are positively interdependent.

A general discussion of the autoregressive model is given in McMillen (1992) and Beron and Vijverberg (2004). As mentioned above, the usual way of parameter estimation based on Eq. (3.3) and Eq. (3.4) is to run a spatial probit on a reduced form of the latent variable equation using ML methods. The reduced form can be written in matrix form as

$$y^* = (I - \phi W)^{-1}(X\beta + u), \tag{3.5}$$

where W is the N-dimensional square matrix of weights. In general, the ML approach to this estimation problem is computationally involved and difficult to handle. There are two main problems with ML estimation based on Eq. (3.5): Firstly, the spatial structure of the model induces heteroskedasticity and therefore causes inconsistency of the parameter estimates. Secondly, the spatially correlated covariance structure does not allow to simplify the multivariate distribution to a product of univariate distributions. Consequently, ML estimation involves an N-dimensional integration problem.

A useful simplification of the model has been proposed by Case (1992). Her approach accounts for the heteroskedastic error structure, but neglects the presence of non-zero elements in the variance-covariance matrix. In its original formulation, Case's model requires W to be a block-diagonal matrix of weights. This is equivalent to a metric of neighborliness where the units of observation are divided into groups. All units of observation belonging to the same group are defined as 'neighbors' to each other. Suppose that i belongs to group $m(i)$ and that $n_{m(i)}$ is the number of units in that group. The composite neighbor's latent variable is then given as the average of the latent variables of units belonging to the same group as i,

$$y^*_{-i} = \frac{1}{n_{m(i)} - 1} \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{ij}\, y^*_j, \tag{3.6}$$

where d_{ij} is an indicator taking value 1 if $m(j) = m(i)$ and zero otherwise. As Case (1992) has shown, with the resulting block-diagonal weight matrix the individual equations of the reduced form can be rewritten as

$$y^*_i = \varrho_{m(i)} x_i \beta + n_{m(i)} \vartheta_{m(i)} \bar{x}_{m(i)} \beta + \varrho_{m(i)} u_i + n_{m(i)} \vartheta_{m(i)} \bar{u}_{m(i)}, \tag{3.7}$$

where $\bar{x}_{m(i)}$ is the vector of mean characteristics for districts in $m(i)$, $\bar{u}_{m(i)}$ is the mean of errors in $m(i)$, $\varrho_{m(i)} = (n_{m(i)} - 1)/(\phi + n_{m(i)} - 1)$, and $\vartheta_{m(i)} = \phi/[(1-\phi)(\phi + n_{m(i)} - 1)]$. Denote the error of the reduced form as v_i such that

$$v_i = \varrho_{m(i)} u_i + n_{m(i)} \vartheta_{m(i)} \bar{u}_{m(i)}. \tag{3.8}$$

The variance covariance matrix of the system is

$$\Omega = E(vv') = (I - \phi W)^{-1} (I - \phi W)^{-1'} \sigma^2. \tag{3.9}$$

Eq. (3.9) reveals that v is heteroskedastic by construction. A variance normalizing transformation to restore homoskedasticity is to pre-multiply the system by the inverse of the square root of diag(Ω). It is straightforward to show that the transformation matrix D is a diagonal matrix, with the i^{th} element on the main diagonal given as

$$D(i,i) = [(\varrho_{m(i)} + \vartheta_{m(i)})^2 + (n_{m(i)} - 1)\vartheta^2_{m(i)}]^{-\frac{1}{2}}. \tag{3.10}$$

Based on the transformed system, the parameters can be estimated by standard ML methods.

As in any model aiming at the identification of spatial dependence, a particular problem is to separate spatial correlation in the variables of interest from spatial error correlation. To account for the potential presence of spatial correlation in u, the model can be extended to incorporate shocks driven by a spatial autoregressive process such as

$$u = \rho W u + \epsilon, \tag{3.11}$$

where it is assumed that $|\rho| < 1$ and that ϵ is an i.i.d. error. In matrix notation, the composite error of the reduced form is then given by

$$v = (I - \phi W)^{-1}(I - \rho W)^{-1}\epsilon. \tag{3.12}$$

The variance normalizing transformation is constructed along the same lines as described before. The elements of the transformation matrix are functions of ϕ and ρ, and the estimation can again rely on standard ML methods.

It has been questioned in the literature whether ϕ and ρ can be separately identified given that they play very similar roles in the extended model.[5] To avoid this identification problem, a spatial moving average instead of the spatial autoregressive process can be used to model spatial error dependence:

$$u = \rho W\epsilon + \epsilon = (I + \rho W)\epsilon. \tag{3.13}$$

The difference between Eq. (3.11) and Eq. (3.13) becomes clear once the autoregressive process is iterated. After M iterations, it can be written as

$$u = (I + \rho W + \rho^2 W^2 + \cdots)\epsilon + (\rho W)^{M+1}u. \tag{3.14}$$

Eq. (3.14) illustrates that with a general weight matrix, in contrast to the spatial moving average process, the autoregressive process links the error of any given unit of observation to the errors of *all* other units of observation in the system. Note, however, that with the particular weight matrix proposed by Case (1992), the autoregressive and the moving average error process are again very similar. The reason is that with weights defining groups of neighbors without any relation between units belonging to different groups, weights

[5] See, for instance, Anselin and Florax (1995).

matrices of higher order are again block diagonal matrices such as W. A useful way to facilitate the identification of ϕ and ρ in Case' model is to use a weight matrix different from W when modelling spatial error dependence. For instance, the spatial error process could be written as

$$u = \rho W_2 \epsilon + \epsilon, \tag{3.15}$$

where the elements of W_2 are defined by some metric of neighborliness different from the one underlying W. Since W_2 will in general not be a block-diagonal matrix, it is slightly more complicated to derive the estimation equation. Using a reasonable approximation for $(I - \phi W)^{-1} u$, the error of the reduced form can be written as

$$v = (I + \phi W + \phi^2 W^2 + \phi^3 W^3)(I + \rho W_2)\epsilon. \tag{3.16}$$

Starting from Eq. (3.16), the variance normalizing transformation can be derived for general weight matrices W and W_2.

An advantage of the spatial autoregressive model compared to the models discussed in sections 3.4 to 3.2 is that it can be applied in situations with incomplete information on y. For instance, if y is observed only for a subset of all units of observation, nothing precludes the estimation of ϕ based on the available information. The reason is simple: Specifying the right hand side of the reduced form requires information only on a unit of observation's own and neighbors' characteristics. Knowledge on neighbors' outcomes is not required. In many applications, information on the characteristics x are readily available for all units of observation. The spatial autoregressive model can then be applied even if information on y is available only for a sample of all units in the population.

3.4 Spatial effects with time lags

Irrespective of the fact that the function which determines the index y_i^* (usually called the index function) is commonly taken to be linear in the parameters, discrete choice models such as logit or probit are nonlinear models by definition. This is because the index is unobservable and the likelihood function is determined by making use of a distributional assumption on the errors. It should have become clear by now that, compared to linear models, the identification of spatial dependence in discrete choice models is complicated by a number of issues. Of course, the question arises how consistent estimation of spatial effects in probit and logit models can be achieved more easily. One way to circumvent some of the problems mentioned so far is to condition on neighbors' observed outcomes in previous periods instead of neighbors' contemporaneous latent variables. A linear specification for the latent variable with a one-period time lag can be written as

$$y_{it}^* = \phi y_{-i,\,t-1} + x_{it}\beta + u_{it}, \tag{3.17}$$

where $y_{-i,\,t-1}$ is now a linear combination of neighbors' lagged outcomes of a form analogous to the one displayed in Eq. (3.2).

In some applications, it may be necessary to include $y_{i,\,t-1}$ among the characteristics x_{it} to account for inertia in the process driving y. If $y_{i,\,t-1}$ is included, the conditional probability for the outcome $y_{it} = 1$ is

$$\Pr\left(y_{it} = 1 \,|\, \{y_{j,\,t-1}\}_{j=1}^{N}, x_{it}\right) = \Pr\left(y_{it}^* > 0 \,|\, \{y_{j,\,t-1}\}_{j=1}^{N}, x_{it}\right). \tag{3.18}$$

With an appropriate assumption on the distribution of u, the composite parameters ϕ/σ and β/σ are identified and average partial effects can be estimated using standard ML techniques.

A number of issues must be addressed with respect to the model displayed in Eq. (3.18). First of all, the question arises why the interaction among the units of observation should take the specific form assumed here. In general, the assumption that the interaction involves a time lag may be more or less reasonable, depending on the particular application. For instance, when y describes the outcome of a complex political process, the notion that the reaction to other decision makers' choices is lagged by one period may often be quite realistic.

Obviously, the model with neighbors' lagged decisions appearing on the right hand side avoids the simultaneity issue that requires the construction of a reduced form in the case of the spatial autoregressive model.[6] However, with serial correlation in errors, $y_{i,t-1}$ may be correlated with u, causing inconsistency of the parameter estimates. Consequently, if Eq. (3.18) has $y_{i,t-1}$ as a right hand side variable, it is crucial to assume that u is serially uncorrelated. Unfortunately, the absence of serial correlation cannot be tested in cases where only a single cross-section of observations is available. Instrumenting the potentially endogenous lagged decision would allow to circumvent the problem. However, apart from the fact that the standard approach of instrumenting in a limited dependent variable framework does not cover the case of a binary endogenous explanatory variable, it is extremely difficult to find valid instruments for a lagged dependent variable.

It is also worth noting that the model with neighbors' lagged decisions does not put any restriction on the size of the interaction coefficient. However, as mentioned above, in most probit or logit models the parameter estimates and the estimate for the variance of the error are not separately identified anyway. Case (1992) discusses the issue in a related context.

When spatial effects are modelled to occur with time lags, it is difficult to properly account for the potential presence of spatial error correlation. Some-

[6] Studies with continuous dependent variables relying on time lags to identify spatial interactions include Hayashi and Boadway (2000) and Fredriksson and Millimet (2002).

times it is reasonable to include a full series of dummy variables for certain regions, i.e. subgroups of the units of observation, to alleviate the problem of correlated shocks affecting the outcomes. In any case, one should test how the inclusion of region-specific dummy variables affects the correlation among residuals in a model without interaction among the units of observation. This can be done by running a test for the presence of spatial error correlation in a baseline model and a model including region-specific dummy variables. Suppose that the residuals follow a spatial auto-regressive process as in Eq. (3.11). Since y_{it}^* is a latent variable, the ordinary residuals, $\hat{u}_{it} = y_{it}^* - x_{it}\hat{\beta}$, are not observed. Nevertheless, the null hypothesis of spatial independence ($\rho = 0$) can be tested based on generalized residuals using the Lagrange-multiplier (LM) test for probit models proposed by Pinkse and Slade (1998).

3.5 Conditioning on neighbors' contemporaneous outcomes: The instrumental variables probit

Both the autoregressive model and the model with time-lagged spatial effects have attractive features. The instrumental variables probit (IV probit) combines two of them, one from each model. Firstly, from the spatial autoregressive model the feature of contemporaneous interaction is adopted. Secondly, as in the model with time-lagged spatial effects, the model has a linear combination of neighbors' outcomes instead of neighbors' latent variables on the right hand side of the structural equation, which is written as

$$y_i^* = \phi y_{-i} + x_i\beta + u_i. \tag{3.19}$$

Again, y_{-i} is a linear combination of neighbors' outcomes of the usual form. Note that the y's on the right hand side of Eq. (3.19) are functions of y_i^*

through their dependence on y_i:

$$\begin{aligned} y_l &= 1[\phi \sum_{\substack{j=1 \\ j\neq l}}^{N} w_{lj} y_j + x_l\beta + u_l > 0] \\ &= 1[\phi(w_{l1}y_1 + \cdots w_{li}y_i + \cdots w_{lN}y_N) + x_l\beta + u_l > 0]. \end{aligned} \tag{3.20}$$

When estimating ϕ and β, the endogeneity of y_{-i} has to be taken into account.[7] Several approaches to deal with the endogeneity of explanatory variables in discrete choice models have been proposed. They all make use of a reduced form equation for the endogenous regressor y_{-i},

$$y_{-i} = x_i\delta_1 + z_i\delta_2 + \nu_i. \tag{3.21}$$

Here, z_i are instruments for y_{-i}. The most straightforward estimation approach is to use conditional maximum likelihood (CML) as discussed in Wooldridge (2002), p. 475-77. It allows estimation of ϕ and β in a simple one-stage procedure. The approach is straightforward to implement and can be used to calculate average partial effects. Other approaches rely on two-step procedures. Newey (1987) has proposed a particularly useful two-step procedure which is similar in spirit to the two-stage least squares (2SLS) approach for linear models with endogenous regressors.[8] A natural choice for the instruments is to use spatially transformed values of the exogenous explanatory variables. That is, the N-dimensional square matrix W is multiplied with the N-dimensional vectors of characteristics, X_k, to produce vectors of instruments, $Z_k = WX_k$.[9] Of course, as with all instrumental variables approaches,

[7]Beron and Vijverberg (2004) claim that a model with y_{-i} on the right hand side is infeasible. However, if y_{-i} is properly treated as an endogenous explanatory variable, nothing precludes estimation of the parameters.

[8]See Rivers and Vuong (1988) for an alternative two-step approach for probit models with endogenous explanatory variables.

[9]Note that characteristics with missing values for some units of observation cannot be used to construct instruments.

the identification of the parameters critically depends on the quality of the instruments: The instruments must be partially correlated with y_{-i} once the other exogenous variables have been netted out, and they must be uncorrelated with u, i.e. z must be exogenous in Eq. (3.21).

Apart from the assumptions with regard to the instruments, both the CML and the two-step approach of Newey rely on fairly strong assumptions. In particular, u_i and ν_i are assumed to be jointly normal. Hence, y_{-i} given (x_i, z_i) should also be normally distributed. Irrespective of how y_{-i} is scaled, the values this variable can take will always be bounded below by zero and above by some positive number. Therefore, the assumption of joint normality of u_i and ν_i may not be perfectly met. To check the robustness of the results, it is often useful to estimate a spatial linear probability model by 2SLS where no distributional assumption with regard to the errors is needed. In addition, the LPM offers a simple way to check the results for robustness with respect to spatial error dependence. As Kelejian and Prucha (1998) have shown for the standard case of a continuous dependent variable, instrumental variables estimation generates consistent estimates of the coefficients even in the presence of spatial error dependence.

Chapter 4

School choice in Michigan: Competition as a driving force for public sector innovation

4.1 Introduction

An ongoing debate revolves around the effects of competition on public school performance. Given that public schools cannot engage in price competition, the question which is typically asked is how a marginal increase in the degree of competition affects the academic achievement of students and the efficiency of public schooling. This is justified by the fact that student achievement and the efficiency of service provision are the key variables students, parents and taxpayers are interested in. However, looking at the correlation between output variables and competition measures does not reveal much about how the public school sector is reacting to the forces of competition. Consider an example with metropolitan areas, some of which have stronger competition among public schools in the sense that they are divided into more school districts.[1] With more school districts, it is easier for households to sort themselves according to preferences for public schooling, property and amenities. In general, to the extent that stronger sorting will increase the match qual-

[1] This is how Hoxby (1994) measures competition among public schools.

ity between students and schools, competition will benefit average student achievement. Hence, even if the behavior of public schools and school districts as schooling producers given their student body is unaffected by the degree of competition, stronger competition may nevertheless lead to better outcomes. For thinking about and assessing reforms towards increased parental choice, a better understanding of the channels through which competition affects the relevant outcomes of public schooling is crucial. Most importantly, we need more insight into the behavior of public schools facing increased competition from other public and private schools.

This chapter examines the effect of increased competition in the public school sector on the behavior of local school districts. In particular, I ask whether public schools really compete for students and resources. For answering this question, I make use of a significant change in the institutional environment of the public school sector in Michigan in 1996, when a voluntary inter-district choice program was established. Under this program, students were given the right to enroll at public schools outside their district of residence. Local school districts would not be forced to accept non-resident students, but districts participating in the choice program and allowing for the enrollment of transfer students would receive additional state funds depending on the number of non-resident students enrolled. The analysis focusses on the districts' participation decisions in the second year of the program. It exploits the variation in the degree of competition for non-resident students which resulted from the initial participation decisions of all districts, and asks how competition affected the districts' readiness to experiment with the new policy. The results suggest that the school districts did use participation in the inter-district choice program to actively engage in competition for students and resources. Facing increased risk of losing students and resources, the school districts seem to have flexibly reacted to competition from neighboring districts. They did this by 'striking back', i.e. they were more likely to compete for non-resident students if neighboring districts did so. The results also suggest that the impact of neighbors' decisions on the participation probability was substantial, and that competition for students contributed significantly to the share of districts allowing for inter-district transfers increasing from 37 percent in 1996/97 to more than 70 percent in 2002/03.

The analysis presented in this chapter adds to the existing literature in explicitly addressing the reactions of schooling producers to increased competition. There is only a limited number of studies that have done so before. Vedder and Hall (2000) and Hoxby (1994) find evidence that public schools react to greater competition from private schools by paying higher teacher salaries. Evidence suggesting that competition enhances the work effort of teachers is presented by Rapp (2000). Finally, Hoxby (2000) shows that increased competition among public schools reduces per-pupil spending and makes school districts allocate resources away from other inputs towards reducing the student-teacher ratio.

In contrast to the limited number of papers on the effects of competition on schooling producers' behavior and policies, an extensive literature has addressed the effects of competition on the performance of public schools and on student achievement. While the theoretical predictions are unclear,[2] the empirical findings suggest that the overall effects of competition are positive.[3] Hoxby (2000, 2003) shows that competition working through residential choices as well as competition provided by recent choice reforms such as vouchers and charter schools positively affects student achievement and school pro-

[2]Hoxby (1999), for instance, offers a principal-agent model of the productivity of schooling producers, showing that a system with property tax finance and Tiebout choice among many jurisdictions reduces rent taken by producers. Epple and Romano (1998) show that low-ability students may be adversely affected by school choice due to peer-group effects.

[3]See Belfield and Levin (2002) for a literature survey. The authors report on a large number of contributions, with a sizable majority showing beneficial effects of competition.

ductivity. Sandström and Bergström (2005) find that school results in Swedish public schools have improved due to competition from independent schools. Couch and Shugart II (1993) find competition by private schools to positively affect public school performance.[4] Dee (1998) and Hoxby (1994) also address the question whether the achievement of students in public schools is improved when the proportion of students attending private schools is higher. Both studies conclude that the competition from private schools has a significant positive effect on the quality of public schools. Grosskopf, Hayes, Taylor, and Weber (2001) find that stronger competition in terms of a lower market share of the biggest schooling producers reduces allocative inefficiencies in some communities.

The chapter proceeds as follows. In the next section, the Michigan inter-district school choice program and potential factors affecting policy preferences of districts are described. Section 4.3 deals with the estimation approach, and the data and the estimation results are presented in Section 4.4.

4.2 Inter-district school choice in Michigan

Inter-district school choice allows students to attend a public school in a school district other than the district of residence. In the US, some states have enacted mandatory choice laws. In these states, school districts are, under certain conditions, obliged to enroll non-resident students at local public schools. In contrast to this, participation of school districts in Michigan and several other states is voluntary. While school districts cannot prevent resident students from choosing a school outside the district, they can prevent the transfer of non-resident students to local schools. For each school year, a school district

[4] The robustness of the results of Couch and Shugart II (1993) has been questioned by Newmark (1995).

in Michigan has to determine whether or not it will accept applications for enrollment by non-resident students. That is, the districts' decisions whether or not to participate in the program are not made once-and-for-all, but can be adjusted for each commencing school year. Moreover, the number of positions available may be specified. Michigan's inter-district public school choice program has been launched by a state law enacted in 1996.[5] Under the new law, districts were free to enroll any applicant in the district's schools provided that the student's home district belongs to the same regional educational service agency. In Michigan, regional educational service agencies are called Intermediate School Districts (ISDs). ISDs originally were created to provide local school districts with services and programs too expensive or too extensive to be offered by districts individually. In 1997, Michigan had 554 school districts and 57 ISDs.

It is important to notice that in its school finance scheme Michigan had shifted from a system relying primarily on local property taxes to a scheme with a per-student state guarantee financed essentially by an increase in the state's sale taxes rate in 1994. The reform markedly increased the degree of centralization in Michigan's school finance scheme. In 1992/93, local revenues on average contributed 65 cents to each dollar of total revenues, while 31 cents came from state funds. By 1997/98, the share of local revenues had dropped to 28 percent and the proportion of state revenues had soared to 68 percent. The minimum school foundation allowance going from the state to local school districts was $ 4,200 per student for the school year 1994/95 and had increased to $ 5,124 per student for the school year 1997/98.[6] Districts losing students under the school choice regime would thus immediately suffer a significant decrease in revenues. At the same time, the school choice program offered

[5]For details see Michigan Compiled Laws, Section 388.1705 (Act 300, 1996).

[6]For details on the school finance reform in Michigan, see Michigan Department of Treasury (2002).

districts the chance to attract students from elsewhere and thereby to raise their revenues.

School districts attracting students from other districts do not only receive additional state funds, but also face an increase in costs. In general, districts will consider to enroll non-resident students only if per-student state aid is at least as high as the costs to serve an additional student. For many school districts, this condition will not hold. This, however, does not preclude that there are strong incentives for many school districts to make use of inter-district school choice. Suppose, for instance, that the districts have perfectly adjusted their capacity to the number of resident students, but that some districts, given their physical capital such as school buildings and the number of teachers employed, operate somewhat below their capacity limit in terms of student enrollment. For these school districts, luring non-resident students to local schools makes good sense, since the cost of serving some additional students will be modest. The competition for students and state funds is a zero-sum game, so if some districts stand to gain students and funds, others stand to lose. Since adjusting capacity is costly or, for small changes in enrollment, even impossible, districts that lose students now face incentives to participate in inter-district choice in order to fully exploit their capacity. Again, other districts will be affected, and so forth. The bottom line is that even if the initial situation is one with only a minority of districts benefitting from inter-district transfers, student mobility across district borders will ultimately provide many districts with incentives to compete for non-resident students. It should be noted, however, that the fiscal incentive to attract or retain students is mitigated by the capitalization effects of inter-district school choice and the corresponding effect on property tax revenues. As Reback (2005) has shown for inter-district choice in Minnesota, residential property values appreciate in districts where students are able to transfer to preferred schools outside the district and decline in districts that accept transfer students.

The school districts' attitude toward school choice might also be affected by factors other than capacity and revenues. In the following, the variables included as controls in the empirical specifications are briefly discussed. First of all, the districts' preferences may vary with size. Bigger districts are more closed, so open enrollment may be a less relevant policy for them. Smaller district may also show a stronger tendency to welcome transfer students due to economies of scale, and they may be more flexible in adjusting to the opportunities provided by the new state law. Another important factor influencing the propensity to participate in open enrollment might be the quality of local public schools in terms of educational achievement. Districts with better schools are more likely to be able to attract non-resident students and should therefore be more inclined towards school choice than districts with worse schools. On the other hand, districts with worse schools will see more students leaving, and the incentives to compensate for the loss in revenues may be stronger. With regard to some characteristics, a district's position relative to its immediate neighbors may be relevant. The point is that, due to transportation to more distant schools being either unavailable or prohibitively costly, school districts will be able to attract students only from nearby districts. A district's relative attractiveness for potential transfer students and the average characteristics of non-resident students whose application is anticipated will therefore depend on a district's characteristics relative to its neighbors. To capture this, I construct two additional control variables. The first one describes a district's relative position with respect to the share of minority students. This variable is conveniently defined as the difference between the district's own share of minority students and the mean of this share for all contiguous districts within the ISD,[7] weighted by district population. The relative position

[7]Recall that participating districts did only accept applications from students residing within the same ISD.

with regard to the districts' median housing value is constructed in the same way. This variable is meant to capture the potential capitalization effects of inter-district school choice. Districts with property values well above those of neighboring districts may be particularly hesitant to participate given the potential of inter-district choice to diminish property value differentials which reflect differences in school quality.

Based on the preceding discussion, I include as control variables in the empirical specification enrollment as a measure for the districts' size; the student-teacher ratio, measuring the capacity for enrollment of transfer students; total revenues per student as a measure for fiscal stress; the average percentage of 7^{th} graders performing satisfactorily in reading and math; the difference between own and neighbors' share of minority students; and the difference between own and neighbors' median housing value.

4.3 Estimation approach

As mentioned above, the Michigan open enrollment law requires a school district in each year to announce whether in the following school year it will admit non-resident students at local schools. This is a discrete choice decision problem which is captured in a latent variable model allowing for time-lagged spatial effects. The model has been discussed in Section 3.4 and is therefore only briefly recapitulated here.

Suppose that district i's predisposition towards the adoption of open enrollment in period t is a function of lagged adoption decisions of other districts $\{y_{j,t-1}\}_{j\neq i}^{N}$, i's lagged own decision $y_{i,t-1}$ and a vector of exogenous characteristics where the first element is unity. A convenient specification for the latent variable is

$$
\begin{aligned}
y_{it}^* &= z_{it}\delta + u_{it} \\
&= \phi \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij}\, y_{j,\,t-1} + \lambda\, y_{i,\,t-1} + x_{it}\beta + u_{it},
\end{aligned}
\tag{4.1}
$$

The inclusion of $y_{i,\,t-1}$ accounts for inertia in the policy process by which the districts' participation decision is determined. The parameter of primary interest in this specification is ϕ. A non-zero value of ϕ would imply that the attitude towards the adoption of open enrollment in any given district depends on lagged adoption decisions in other districts.

There are a number of issues making the model with time-lagged spatial effects an attractive choice in the context of this chapter. First of all, the school districts' participation decision is the result of a complex political process, and it seems reasonable to account for a certain time lag when the districts adjust their behavior to decisions in neighboring districts. The structural equation (4.1) is a particularly attractive choice in applications where lagged decisions are good predictors of actual predispositions. As we will see, this is the case here.

As mentioned in Chapter 3, an important issue that has to be taken into account in estimating spatial effects is the potential presence of spatial error correlation. Consider an example with two regions A and B, where districts in A are more inclined towards school choice than districts in B due to some common time-invariant unobserved characteristic. With the diffusion rate being higher in A than in B, the specification in Eq. (4.1) would erroneously attribute the effect of the unobserved characteristic to neighbors' past decisions. Hence, accounting for the potential presence of spatially correlated components in the errors is crucial.

The econometric approach presented in this chapter takes account of spatial error correlation in a simple but straightforward way. The approach takes the example of unobserved region-specific effects literally and extends the structural equation by a number of dummy variables for regions. Of course, for the approach to make sense these regions have to be defined in a meaningful way. With regard to the Michigan school districts, the natural way to proceed is to define regions according to Intermediate School Districts as regional educational service agencies. ISDs are higher level authorities in the federal educational system of Michigan, and the vertical impact of ISD policies on local school districts may well lead to spatial correlation in the school districts' behavior towards open enrollment. Suppose, for instance, that ISDs engage in policy coordination among affiliated districts,[8] or that ISD officials have certain preferences towards inter-district school choice and try to affect policies at the local level accordingly. The dummy variables will take account of any such region-specific effect on district policies and thereby help to remove spatial correlation from the errors. A test for the presence of spatial error correlation in a model without interaction among districts is used to make sure that the approach works. Suppose that the residuals follow a spatial auto-regressive process,

$$u_{it} = \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, u_{j,t} + \epsilon_{it}, \tag{4.2}$$

where $\epsilon \sim N(0, I_N)$. Since y^*_{it} is a latent variable, the ordinary residuals are not observed. The null hypothesis of spatial independence ($\rho = 0$) is therefore tested based on generalized residuals using the Lagrange-multiplier (LM) test for probit models proposed by Pinkse and Slade (1998). Running the test on generalized residuals from a standard probit and a probit including ISD dummies will reveal to which extent the inclusion of ISD dummies removes

[8]Note that under Michigan law ISDs could run their own ISD-wide school choice programs. Local school districts in these ISDs would then be exempt from the provisions of the statewide program. See Michigan Compiled Laws, Section 388.1705b (effective since June 1997).

spatially correlated components from the residuals.

Since the structural equation has $y_{i,t-1}$ as an explanatory variable, a potential problem could also arise from serial correlation in u. Unfortunately, the absence of serial correlation cannot be tested based on a single cross-section of observations. Instrumenting the potentially endogenous lagged decision would allow to circumvent this problem. Apart from the fact that the standard approach of instrumenting in a limited dependent variable framework does not cover the case of a binary endogenous explanatory variable, it is extremely difficult to find valid instruments for a lagged dependent variable. In the light of these difficulties, I use a different approach to check whether the findings on district interactions are affected by serial correlation in the residuals. Taking advantage of data on the districts' school choice policies from 1999 to 2002, I estimate a probit for the 2002 cross-section with lagged own decisions from the previous three years and a full set of ISD-dummies as explanatory variables. With three lags included, the model is likely to account for any unobserved heterogeneity which might cause the errors in Eq. (4.1) to be serially correlated. Comparing the results from the original model and the model with additional lags included will then provide evidence on the robustness of the interaction coefficient.

A further issue is the choice of the weights w_{ij}. In general, it is difficult to define appropriate weights since no general criterion for discriminating between alternative definitions is available. In the present case, however, things should be less complicated than in many other applications. Note firstly that students in 1997 could only transfer to districts within the same ISD. Hence, policies of districts outside the ISD should not have affected participation decisions. Consequently, the weights should select as neighbors only districts within the ISD. Secondly, given the costs associated with being enrolled at a school which is far away from a place of residence, competition for students should in most cases take place among adjacent districts. Hence, the row-standardized spatial weights are defined as

$$w_{ij} = \frac{d_{ij}}{\sum\limits_{\substack{j=1 \\ j \neq i}}^{N} d_{ij}} \quad \text{if } i \neq j \quad \text{and} \quad w_{ij} = 0 \quad \text{if } i = j,$$

where d_{ij} is an indicator taking value 1 if j belongs to the same ISD and is adjacent to i, and zero otherwise.

I also experimented with a number of other metrics where, for instance, I adjusted the weights according to the size of neighbors measured by enrollment or population. The results were very similar to those obtained with the weights described here.

A final point in the discussion of the estimation approach relates to the possibility that the districts' response to lagged decisions of neighbors systematically differs between certain groups of districts. A natural asymmetry to explore is the one between adopters (districts which adopted school choice in $t-1$) and non-adopters. More specifically, I also estimate a model where neighbors' policies are interacted with $y_{i,t-1}$,

$$\begin{aligned} y_{it}^{*} &= \phi_1 \, y_{i,t-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} \, y_{j,t-1} \\ &+ \phi_2 \, (1 - y_{i,t-1}) \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} \, y_{j,t-1} + \lambda \, y_{i,t-1} + x_{it}\beta + u_{it}. \end{aligned} \tag{4.3}$$

For districts which have adopted open enrollment in $t-1$ the second term on the right hand side of Eq. (4.3) equals zero. Hence, ϕ_1 measures the extent to which neighbors' lagged decisions affect current policies of first-period adopters. If open enrollment has not been adopted in $t-1$, the first term equals zero, and ϕ_2 measures the neighborhood influence on current policies of those districts which did not adopt open enrollment previously. A difference between ϕ_1 and ϕ_2 would indicate that it depends on lagged own decisions how districts are affected by policies in neighboring districts. Of course, other asymmetries in responses can be analyzed in a similar way.

Table 4.1: District participation 1996-1997 – descriptive statistics

	Full sample		Reduced sample	
	Mean	S.D.	Mean	S.D.
Participation in 1997	0.453	0.498	0.494	0.501
Participation in 1996	0.369	0.483	0.391	0.489
Nob	521		338	

4.4 Data, estimation and results

The empirical analysis is based on data on 521 Unified School Districts in Michigan.[9] The analysis focuses on the behavior of school districts in the first two years of the Michigan open enrollment program. Since indicators for lagged decisions are included in all specifications, the spatial interaction among districts is identified using the cross-section of districts from the second year of the program, 1997. Table 4.1 provides descriptive statistics on district participation in 1996 and 1997 both for the sample of 521 districts and a reduced sample of 338 districts which is used for estimations including ISD dummies.[10]

In the first year, 192 out of 521 districts in the sample allowed for enrollment of non-resident students. In 1997, 60 districts joined and 16 districts left the program. With 236 open enrollment districts, the participation rate in 1997 was 45.3 percent. Table 4.2 provides descriptive statistics on district

[9] A minority of 30 Michigan school districts runs only elementary schools and is excluded from the sample. Furthermore, the concept of neighborliness used in this study is not applicable to two Unified School Districts which are islands. Detroit City School District served almost 165,000 students in 1997 (about 10 percent of all students in Michigan) and is excluded as an influential observation.

[10] The reduction of the sample to only 338 districts in estimations including ISD dummies is for technical reasons and has nothing to do with the exclusion of districts for the reasons discussed in footnote 9.

Table 4.2: School choice in 1997 - descriptive statistics (Nob=521)

District characteristics	Mean	S.D.	Min	Max
% neighbors participating in 1996[a]	0.359	0.392	0	1.00
Enrollment[b]	2.80	3.30	0.077	26.1
Student teacher ratio	14.9	1.91	7.29	19.7
Revenues per student[c]	7.21	1.26	5.35	14.7
% math/reading satisfactory grade 7[d]	0.557	0.130	0.063	0.896
Difference between own and neighbors' median housing value[e]	0.010	0.278	-1.41	1.98
Difference between own and neighbors' % minority students	-0.038	0.184	-0.896	0.752

[a] Neighbors defined as contiguous districts in own ISD.
[b] in 1000 students.
[c] In 1000 $.
[d] Average percentages of students with satisfactory proficiency in math and reading.
[e] In 100,000 $.

characteristics for 1997.[11]

The data on district participation are from Arsen, Plank, and Sykes (1999), p. 33. Data on enrollment, minority students, teachers, revenues, and math/reading proficiency are for the 1997/98 school year and have been obtained from the K-12 database of the Michigan Department of Education, Center for Educational Performance and Information (CEPI). The data were accessible via the web page of the Department at the time of writing. Data on the median value for specified owner-occupied housing units in 1999 are from the School District Demographic System of the National Center for Education Statistics (NCES).

The first step in the analysis is to run two baseline regressions where the potential impact of neighbors' lagged policies is ignored. The baseline regressions are meant as a first test whether the approach of estimating a discrete choice model for the adoption of open enrollment policies with the given set

[11] The descriptive statistics for the reduced sample are very similar to those displayed in Table 4.2 and are omitted.

Table 4.3: School choice in 1997 – baseline probit estimations

Specification includes ISD-dummies	No		Yes	
	Coeff.	dP/dX	Coeff.	dP/dX
Participation in 1996	2.36 ***	0.718	2.74 ***	0.635
	(0.164)		(0.309)	
Enrollment	-0.014	-0.003	0.026	0.004
	(0.026)		(0.041)	
Student teacher ratio	-0.182 ***	-0.038	-0.330 ***	-0.053
	(0.045)		(0.086)	
Revenues per student	-0.191 ***	-0.040	-0.354 ***	-0.057
	(0.066)		(0.124)	
% math/reading satisfactory	1.34 **	0.278	3.06 ***	0.489
	(0.613)		(1.04)	
Difference between own and neighbors' median housing value	-0.724 **	-0.150	-1.34 **	-0.215
	(0.350)		(0.529)	
Difference between own and neighbors' % minority students	0.388	0.081	0.397	0.063
	(0.487)		(0.624)	
Nob	521		338	
Log-likelihood	-196.52		-97.57	
Percent correctly predicted	85.6		87.6	
p-value for test for spatial error correlation[a,b]	0.00		0.608	

Standard errors in parentheses.
[a] Neighbors defined as contiguous districts in own ISD.
[b] Based on 10,000 replications.
** 5% significance level.
*** Idem., 1%.

of control variables is meaningful at all. Furthermore, the baseline regressions will help us to address the issue of spatially correlated unobserved effects. Table 4.3 reports the results.

The first two columns display coefficients and average partial effects of a simple probit specification with participation decisions in 1997 as the dependent variable. Five out of the seven explanatory variables show coefficients significant at least at the 10 percent level, and the model correctly predicts more than 85 percent of all decisions. As expected, the participation decision in 1996 is a strong predictor for participation in 1997. Furthermore, the coefficients of the student-teacher ratio and the revenue variable are significant and show the expected sign. The results also suggest that districts with higher proficiency levels are more and that districts with housing values above those in neighboring districts are less inclined towards adopting school choice. Taken together, the estimation results for the baseline model suggest that a number of important school district characteristics affecting the adoption of open enrollment as a district policy have been identified.

The second set of results in Table 4.3 is from a probit of participation decisions in 1997 on the same set of explanatory variables as before and, in addition, a full set of ISD dummies. Note that the estimation with ISD dummies is based on a sample of only 338 observations. The reason is that in 22 out of 57 ISDs all affiliated local school districts either adopted open enrollment in 1997, or they all opted out of the program. With dummy variables for ISDs, these observations have to be removed from the sample in order to avoid complete separation. With the reduced sample and the additional regressors, the general picture is the same as before, although the average partial effects of the student teacher ratio, revenues per student, proficiency level and housing values now are somewhat more pronounced. This is only partly attributable to the inclusion of the dummy variables. It also reflects the fact that, after dropping observations from ISDs with no variation in the dependent variable, the link between the explanatory variables and participation decisions is more clearly visible.

As mentioned in the previous section, controlling for region-specific effects is a straightforward way of removing spatially correlated components from the residuals. To see how this approach works in the present setting, a LM test for spatial error correlation has been performed for both specifications. As suggested by Pinkse and Slade (1998), a bootstrapping method is used to derive p-values for the null of zero spatial error correlation. This accounts for the fact that the LM test statistic does not have a limiting χ^2-distribution but, instead, depends on the matrix of spatial weights. The results are reported in the last row in Table 4.3. Based on 10,000 replications, the result for the model without ISD dummies clearly suggests the rejection of the null of zero spatial correlation in errors. In contrast to this, once region-specific dummies are included, the p-value of the test reveals that there is no significant amount of spatial error correlation. Therefore, when moving to specifications with spatial interaction in adoption decisions, there is little reason to suspect that the results will be significantly affected by spatially correlated unobserved effects once the impact of ISDs on district policies is accounted for.

We now turn to estimates of the spatial probit model, where the focus is on identifying the impact of neighbors' lagged decisions on actual participation decisions. A first set of results with the dependent variable being the districts' decisions in 1997 is displayed in Table 4.4. Note that, as suggested by the results of the tests reported above, a full set of ISD dummies is included to remove spatially correlated components from the errors. The main result from Table 4.4 is that there is positive neighborhood influence in the adoption of open enrollment policies and that the impact of lagged adoption decisions of neighbors on current policies is substantial. A one percentage point increase in the share of neighbors participating in the first year of the school choice program increases the current probability of adoption by about 0.2 percent, implying that a district with a share of participating neighbors one standard deviation above that of an otherwise identical reference district is about 7.8 percent more likely to accept transfer students. A likelihood ratio test on equality of the log-likelihood of the spatial probit and the baseline probit with ISD dummies gives a test statistic of 5.48. This is significantly different from zero at the 2 percent level, implying that not accounting for the impact of neighbors' lagged decisions removes a significant amount of information from the system.

Districts which already adopted open enrollment in 1996 are about 64

Table 4.4: School choice in 1997 – spatial probit

	Coefficient	dP/dX
% neighbors participating in 1996[a]	1.29 **	0.200
	(0.560)	
Participation in 1996	2.83 ***	0.636
	(0.323)	
Enrollment	0.023	0.004
	(0.043)	
Student teacher ratio	-.363 ***	-0.056
	(0.090)	
Revenues per student	-.365 ***	-0.057
	(0.126)	
% math/reading satisfactory	3.32 ***	0.516
	(1.07)	
Difference between own and neighbors' median housing value	-1.50 *** (0.542)	-0.233
Difference between own and neighbors' % minority students	0.385 (0.642)	0.060
Nob	338	
Log-likelihood	-94.83	
Percent correctly predicted	87.3	

Standard errors in parentheses.
Regression includes a full set of ISD dummies.
[a] Neighbors defined as contiguous districts in own ISD.
** 5% significance level.
*** Idem., 1%.

percent more likely to allow for the transfer of non-resident students than districts which did not participate in the first year of the choice program. In addition, a number of other explanatory variables affect the districts' choices. An additional student per teacher decreases the probability of adoption by 5.6 percent, while $ 1,000 of additional revenues per student decrease the participation probability by 5.7 percent. These findings support the presumption that crowded schools and the relative abundance of available resources deter school districts from actively competing for non-resident students. Furthermore, ceteris paribus, a higher share of students scoring satisfactorily in reading and math increases the probability of adoption. A one percentage point increase in the share of proficient students is associated with an increase in the participation probability of 0.52 percent. This may reflect the fact that better local schools, holding fixed all other factors, make a district a more attractive choice for potential transfer students. Introducing inter-district school choice is therefore more likely to be a successful policy in districts with a high share of proficient students. Finally, the adoption probability decreases with the difference between a district's median housing value and the average median housing value in neighboring districts. On average, an increase in this difference by $ 1,000 makes participation in school choice about 0.2 percent less likely. This points to the fact that households which have invested in their children's education by purchasing (relatively) expensive housing will most likely suffer a loss in the value of their houses if their school district participates in inter-district school choice.

As mentioned in the previous section, a problem with the results presented so far is the potential presence of serial correlation in the errors. If u_{it} is correlated with $u_{i,t-1}$, the lagged dependent variable in Eq. (4.1) is endogenous and the estimates of all coefficients are bound to be inconsistent. Why should the errors be serially correlated? The most plausible source of serial correlation in the errors is certainly the presence of an unobserved time-invariant effect in the equation describing the districts' predisposition towards open enrolment. A simple way to address this problem is to include a sufficient number of lags of the dependent variable as additional regressors. By conditioning on the observed outcomes from previous periods, we are likely to account for any important unobserved factors which drive the districts' participation decisions. This way of dealing with serial correlation in the errors driven by unobserved effects is similar in spirit to the approach of Blundell, Griffith, and Van Reenen (1995), who analyze technological innovations of firms and propose to utilize the pre-sample history of innovation activity to account for unobserved effects. Of course, it is not possible to include a sufficient number of lags with data on

Table 4.5: District participation 1999-2002 – descriptive statistics

	Full sample		Reduced sample	
	Mean	S.D.	Mean	S.D.
Participation in 2002	0.704	0.457	0.638	0.481
Participation in 2001	0.645	0.479	0.578	0.495
Participation in 2000	0.610	0.488	0.509	0.501
Participation in 1999	0.478	0.500	0.404	0.491
Nob	521		334	

Source: Michigan Department of Education (see footnote 12).

the districts' participation decisions only from the first two years of the school choice program. I therefore use information on district participation in the Michigan school choice program from 1999 to 2002 provided by the Michigan Department of Education.[12] A problem with the data is that they do not allow to distinguish between school districts which did not participate in the program and districts which did participate and experienced a zero demand from non-resident students to be enrolled at the district's schools.[13] In what is reported below, districts with zero incoming transfer students are treated as not participating in the school choice program. Note that less than 2.7 percent of all districts with a strictly positive number of incoming transfer students report to receive less than two full-time equivalent students. Hence, it is highly unlikely that the picture of participation decisions emerging from the data by treating districts with zero transfers as non-participating is seriously biased.

Table 4.5 displays descriptive statistics for district participation from 1999 to 2002 both for the full sample and for a reduced sample of 334 districts.[14] Note that the share of participating districts lags behind in the reduced sam-

[12]The data have been provided by Dr. Arthur Vrettas at the Michigan Department of Education, who was in charge of the Schools of Choice Program at the time of writing.

[13]It is for this reason that I do not use this data in the main analysis.

[14]As before, the sample has be to reduced to avoid complete separation in an estimation including ISD dummies.

Table 4.6: School choice in 2002 – probit with three lags of dependent variable

	Coefficient	dP/dX
% neighbors participating in 2001[a]	1.10 **	0.207
	(0.499)	
Participation in 2001	1.67 ***	0.349
	(0.301)	
Participation in 2000	0.767 **	0.156
	(0.359)	
Participation in 1999	0.076	0.014
	(0.328)	
Nob	334	
Log-likelihood	-113.53	
Percent correctly predicted	85.0	

Standard errors in parentheses.
Regression includes a full set of ISD dummies.
[a] Neighbors defined as contiguous districts in own ISD.
** 5% significance level.
*** Idem., 1%.

ple. This is because most of the excluded districts are from ISDs with participation of all affiliated districts. Results of a probit of district participation in 2002 on neighbors' lagged decisions, three lags of the dependent variable and a full set of ISD dummies are shown in Table 4.6.[15] The average partial effect of neighbors' lagged decisions is about 0.21 compared to 0.2 in the spatial probit for the 1997 cross section. The impact of participation in the previous year is now 0.35, which is significantly lower than with only one lag included. The coefficient of the participation decision lagged by two periods is also significant. All other things being equal, a district which did participate two years ago is still about 16 percent more likely to allow for inter-district transfers. However, the positive effect of lagged adoption decisions is dying out with more distant lags. If we are willing to believe that the inclusion of three lags of the dependent variable is likely to remove any significant amount

[15]Including additional controls as in the regressions reported before did not significantly alter the estimates of coefficients and average partial effects.

Table 4.7: School choice in 1997 – probit with asymmetric responses

	Coefficient	dP/dX
Neighbors' impact on previous year adopters[a]	0.764	0.118
	(0.797)	
Neighbors' impact on previous year non-adopters[a]	1.52**	0.235
	(0.618)	
Nob	338	
Log-likelihood	-94.41	
p-value for Wald test on equality of coefficients	0.357	

Standard errors in parentheses.
Additional regressors: see Table 4.4.
[a] Neighbors defined as contiguous districts in own ISD.
** 5% significance level.

of serial correlation which might potentially be present in the errors, then the results displayed in Table 4.6 suggest that the positive and significant estimate for the interaction coefficient does not suffer from endogeneity bias.

Since the estimate of the spatial interaction coefficient does not seem to be affected by serially correlated errors, we now return to results derived from the districts' participation decisions in the second year of the school choice program. Table 4.7 presents the estimates for interaction coefficients from a probit allowing for asymmetric responses between previous year adopters and previous year non-adopters.[16] The hypothesis that previous-year adopters' choice of current policies is not affected by lagged decisions of neighbors cannot be rejected at the 10 percent significance level. At the same time, the interaction coefficient for previous year non-adopters is significant at the 5 percent level. The partial effect indicates that previous-year non-adopters are 0.24 percent more likely to participate in open enrollment if the share of previous-year adopters among neighbors is increased by one percentage point. The results suggest that school districts which did not participate in the first

[16]The estimates of coefficients and average partial effects for the control variables are very similar to those obtained before and are omitted.

year of Michigan's inter-district school choice program were 'pulled' to participation in the second year by previous-year adopters in their geographical environment, while first year adopters seem to have been unaffected by the behavior of other districts. This finding is consistent with the view that participation in the choice program was propagated by its effects on the degree of competition for students and state funds. Note, however, that the difference in the estimates for the interaction coefficients is not significant at conventional levels.

In general, one might think of forces other than competition for non-resident students making the districts' participation decisions interdependent. For instance, some districts may have abstained from participation in the first year of the program's existence in order to learn from the experience with the new policy made by other districts. The positive impact of neighbors' lagged decisions could thus also be explained by the notion that the quality and the amount of information available to districts positively depends on the number of experiments in nearby districts.[17] The Michigan school choice law in its original formulation offers a simple way to discriminate between the competition hypothesis and alternative explanations for interdependencies among the districts' decisions. As mentioned earlier, the law allowed for inter-district transfers only within ISDs. Therefore, school districts whose school choice policies were motivated by concerns about losing students by inter-district transfers should have reacted to lagged adoption decisions of neighboring districts within the ISD, but they should have been indifferent towards decisions of neighbors outside their ISD. In contrast to this, the difference between neighbors within and outside the ISD should have played no role if interdependencies among the districts' choices were driven by information spillovers. To

[17] As mentioned in Chapter 2, there is a small body of theoretical literature on strategic behavior of jurisdictions in policy experimenting and policy innovation. See, e.g., Strumpf (2002) and Kotsogiannis and Schwager (2006a).

Table 4.8: Effect of lagged decisions of neighbors within and outside ISD

	Coefficient	dP/dX
Lagged decisions of neighbors within ISD	1.25**	0.192
	(0.554)	
Lagged decisions of neighbors outside ISD	-0.465	-0.072
	(0.392)	
Nob	338	
Log-likelihood	-94.12	
p-value for Wald test on equality of coefficients	0.010	

Standard errors in parentheses.
Additional regressors: see Table 4.4.
** 5% significance level.

discriminate between the alternative explanations, I estimate a model with the share of participating neighbors within the ISD and the corresponding share of adjacent districts outside the ISD included as separate regressors. The results are shown in Table 4.8. While the estimates for the interaction coefficient and the corresponding average partial effect for neighbors within the ISD are very similar to those presented in Table 4.4, the interaction coefficient for adjacent districts outside the ISD is insignificant. Hence, the average district was only influenced by adoptions of school choice which did affect the choices available to students residing in the district. This finding clearly supports the view that competition for students was the driving force for the districts' participation decisions being positively interdependent.

Chapter 5

Strategic interaction in school choice policies: Additional evidence

5.1 Introduction

Extending the analysis of interaction in the adoption of policy innovations, this chapter provides evidence on behavioral convergence in school choice policies among local school districts. As in the Michigan case study, the variation in choice policies at the district level is exploited by focussing on the adoption of inter-district public school choice in states where the school districts' participation in this sort of school choice is voluntary. The evidence is derived from a unique data set providing information on school district characteristics and policies in a large sample of districts. The estimation approach is based on the spatial autoregressive probit model proposed by Case (1992) and outlined in Section 3.3. Recall that the spatial structure in Case's model is required to fulfill certain regularity conditions. In particular, the matrix of spatial weights is assumed to be block-diagonal. To accommodate to this particular feature of Case's model, a simple and somewhat stylized measure for geographical proximity is used to describe reference districts. More specifically, county borders are used to define groups of districts which are 'neighbors' to one another. Thus, for any given district, a composite reference district is defined from the set of all local school districts belonging to the same county.

The remainder of the paper proceeds as follows. In the following section, the empirical approach is briefly recapitulated, and some extensions are discussed. Section 5.3 describes the data, and the estimation results are presented in Section 5.4.

5.2 Estimation approach

A special feature of the analysis provided in this chapter is that the evidence on spatial interactions in local school districts' decisions to adopt choice policies is based on data describing policies in a sample of local school districts (and not the population of districts). As mentioned in Section 3.3, the spatial autoregressive model is an attractive choice for the analysis of such data. The model does not require the knowledge of policies, y, for all districts in the population (i.e., all districts in the states under consideration) and can therefore be estimated based on sample data on y without any difficulty.

Given the mere number of almost 15,000 school districts in the U.S., it seems reasonable to assume that decision makers at the district level tend to perceive the situation in nearby districts as particularly informative with regard to the prospects of new policies. In the present context it is therefore less restrictive than in many other applications to use some geographical structure to build a block-diagonal weight matrix with each block defining a geographically defined group of 'neighbors'. For most US states, the natural geographical structure to define the blocks in W are the county borders. This is because in many states, counties are regional educational service agencies.[1]

[1] As discussed in Chapter 4, Michigan with its Intermediate School Districts as regional educational service agencies is an exception.

In those states, county boards of education provide important services to local school districts and serve as mediators between the state board of education and local school boards. However, given that a matrix of weights defined according to county borders will be similar to other weight matrices based on geographical proximity, it is a reasonable approach to use weights based on county affiliation even in states where counties do not have any responsibilities for public schools and local school districts.

For the reasons given above, the empirical investigation of school choice policies of US school districts presented in this chapter will rely on the spatial probit model proposed by Case (1992). Since the model has been discussed in general terms in Section 3, I will be very brief in outlining the econometric approach.

Consider the latent variable model from Section 3.3. Let the latent variable y_i^* describe the predisposition of some school district to experiment with a new policy. Actual adoptions y_i are determined as

$$y_i = 1[y_i^* > 0] \quad i = 1, \ldots, N. \tag{5.1}$$

The structural equation for the predisposition towards adoption of the new policy is specified as

$$y_i^* = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, y_j^* + x_i\beta + u_i. \tag{5.2}$$

Recall that the number of districts belonging to county m is denoted as n_m. With the equations sorted by county, a block-diagonal row-standardized weight matrix W is constructed by setting $w_{ij} = 1/(n_{m(i)} - 1)$ if $m(j) = m(i)$ and $w_{ij} = 0$ otherwise.

The issue of spatial error dependence has also been discussed in Section 3.3. To facilitate the identification of the interaction coefficient, ϕ, and the coefficient from the spatial error process, ρ, I use a spatial moving average together with a weight matrix W^c which is based on contiguity indicators to model u. The error process can thus be written as

$$u_i = \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}^c \, \epsilon_j + \epsilon_i, \tag{5.3}$$

where the weights are derived from a matrix of contiguity indicators, $\tilde{W}$, with $\tilde{w}_{ij} = 1$ if i and j share a common border and $\tilde{w}_{ij} = 0$ otherwise. The contiguity based weights are computed as

$$w_{ij}^c = \tilde{w}_{ij} \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} \tilde{w}_{ij} \right)^{-1} \quad \text{if } i \neq j \quad \text{and} \quad w_{ij}^c = 0 \quad \text{if } i = j. \tag{5.4}$$

5.3 Data

The empirical analysis of inter-district choice policies at the district level is based on a sample of school districts in Arkansas, California, Idaho, Massachusetts and Ohio. All five selected states share the common feature that they established inter-district choice programs between 1989 and 1993 and that districts were given discretionary power to decide whether they would admit nonresident students at local schools.[2] Thus, with regard to the selected states, inter-district open enrollment can be considered a policy innovation implemented at the school district level.

The information on the school districts' policies towards open enrollment is from the Schools and Staffing Survey (SASS) 1993/94, providing data on a large sample of local school districts.[3] The Schools and Staffing Survey

[2]Choice programs started in Arkansas and Ohio in school year 1989/90, in Massachusetts and Idaho in 1991/92, and in California in 1993.

[3]To access the data, refer to National Center for Education Statistics (1998). For technical information, see National Center for Education Statistics (1996).

is administered by the National Center for Education Statistics (NCES). In the 1993/94 wave, the survey (among numerous other topics) asked districts whether they had 'a choice program in which students can enroll in another school or district outside their attendance area without justification based on individual special needs'. Districts which affirmed were then asked whether the program allowed for enrollment of students from other districts. In the empirical analysis, the answer to this last question is used to determine which districts did participate in inter-district open enrollment in the 1993/94 school year. Hence, the dependent variable is defined exactly as in the previous chapter and captures the districts' willingness to serve non-resident students by accepting inter-district transfers.

In the following, the control variables are briefly discussed. They are similar to those used in the Michigan case study presented in Chapter 4. In all five states, fiscal incentives for participation were set by rewarding districts with additional funds for the admittance of transfer students. Participating districts could thus hope to raise additional revenues by attracting transfer students. The model accounts for the fiscal position of districts by including revenues per student among the control variables. Other explanatory variables in the empirical specification are district enrollment, the student-teacher ratio, the share of minority students,[4] the median household income, and four dummy variables, one for districts in large or mid-size central cities, one for suburban school districts, one for elementary school districts and one for high school districts. In addition, state dummy variables are included. They account for all kinds of state-specific influences on the predisposition towards open enrollment, such as differences in school choice laws, state-specific fiscal incentives for districts promoting participation in open enrollment programs or the length of time the program was in place at the time of data collection.

[4]The share of minority students is defined as one minus the share of white non-Hispanic students.

Table 5.1: Number of school districts/observations and participation rate, by state

				Participation rate[a]	
	# districts	Nob SASS	Nob sample	SASS	Sample
Arkansas	315	119	84	0.372	0.446
California	1000	192	152	0.294	0.446
Idaho	113	74	40	0.678	0.628
Massachusetts	259	115	102	0.165	0.229
Ohio	611	149	136	0.534	0.524
All	2298	649	514	0.373	0.458

[a] Weighted by inverse of sampling probabilities.

For the five selected states, the SASS provides information on open enrollment policies in 649 local school districts. For the empirical analysis, the sample was reduced to 514 districts.[5] Table 5.1 displays the number of school districts/observations and the participation rate by state. In the first column, the total number of local school districts in 1993/94 is provided. Column two and three show the number of districts sampled in the SASS and the number of districts used in estimations. The last two columns display the participation rates, both for the SASS sample and the sample used in estimations. Note that the participation rates are computed as weighted averages of indicators for participation, where the weights are defined as the inverse of the sampling probabilities. The districts covered in the SASS are selected on the basis of a

[5] 17 districts had to be excluded from the sample since they represent a whole county, i.e. they have no neighbors. Another 14 districts had missing values for explanatory variables. In a next step, districts with less than 800 students were removed from the sample. The reason for doing so is the presumption that the political behavior of a very small district will resemble that of an average school more closely than that of a larger district. The threshold of 800 students was determined by increasing the minimum number of students by increments of 100 (starting from zero) until each of the remaining districts had at least two schools. Finally, in order to identify influential observations, a linear probability model was estimated using the remaining 539 observations. Based on the approach proposed by Krasker, Kuh, and Welsch (1983), 25 observations were removed. This left 514 school districts for the analysis.

complex survey design, with considerable variation in sampling probabilities. School districts in large central urban areas, for instance, are sampled with a high probability to account for the fact that the overall number of such districts is relatively small. In contrast to this, districts belonging to the huge number of rural districts are sampled with a relatively low probability. Using the inverse of the sampling probabilities as weights corrects for distortions which result from the survey design. Descriptive statistics accounting for the variation in the sampling probability should not systematically differ from the respective statistics based on a random sample of school districts.

A further point worth to be mentioned with regard to Table 5.1 are the differences in participation rates between the SASS sample and the sample used in estimations. One reason for this difference to be significantly negative for Arkansas, California and Massachusetts is the exclusion of districts with less than 800 students. Both the participation rate of these districts and their sampling probability in the SASS is below average. Since low sampling probabilities translate into high weights, the positive effect of dropping very small districts from the sample on the average participation rates is clearly visible.

As shown in Section 3.3, the identification of the parameters of the spatial probit model from its reduced form requires information on mean district characteristics by county. In Case (1992), survey data are used to estimate the mean characteristics. I use a more robust approach and derive mean characteristics from the population of districts. To do so, data sources providing information on all US school districts are used. The data are from the Public Education Finance Data of the Bureau of the Census (revenues and district type indicator), the School District Demographic System of the NCES (median household income[6]) and the Common Core of Data of the NCES (enrollment,

[6]Data on median household income is from 1999.

Table 5.2: Descriptive statistics (Nob=514)

	Mean[a]	S.D.[a]	Min	Max
Number of neighbors[b]	17.7	18.4	1.00	80.0
Central city	0.105	0.307	0	1.00
Suburb	0.302	0.459	0	1.00
Elementary school district	0.091	0.287	0	1.00
High school district	0.032	0.177	0	1.00
Enrollment[c]	5.36	8.25	0.803	127
Student-teacher ratio	19.7	3.54	11.9	29.5
% minority students	0.206	0.252	0.00	0.960
Revenues per student[d]	5.00	1.19	2.97	11.5
Median household income[d]	39.7	13.2	17.1	98.2

[a] Weighted by inverse of sampling probabilities.
[b] Number of districts belonging to the same county.
[c] In 1000 students.
[d] In 1000 $ and 1993 prices.

number of minority students, number of teachers, and urbanicity indicator). Table 5.2 provides descriptive statistics for the sample used in estimation.

5.4 Estimation and results

All estimations based on the SASS data have to account for the distortions induced by the sample design. With regard to parameter estimation, the effect of the survey design on the composition of the sample is accounted for by including the inverse of the sampling probabilities as a weights in the likelihood function. In addition, the weights have to be taken into account in the computation of the coefficients' asymptotic variance-covariance matrix. The best way to proceed is to use a Huber-White formula for probit models. Suppose we want to compute robust standard errors for a K-dimensional vector of coefficients, γ. Let the log-likelihood function be

$$L(\gamma) = \sum_{i=1}^{N} L_i(\gamma) = \sum_{i=1}^{N} \alpha_i [y_i \ln G(\gamma, z_i) + (1 - y_i) \ln(1 - G(\gamma, z_i))], \tag{5.5}$$

where $G(\cdot)$ is the cdf of the standard normal distribution with density $g(\cdot)$ and α_i is the weight of i. In general, a robust asymptotic variance-covariance matrix looks like

$$\widehat{Avar}(\hat{\gamma}) = H^{-1} \sum_{i=1}^{N} s_i {s_i}' \; H^{-1}, \tag{5.6}$$

where H is the Hessian evaluated at $\hat{\gamma}$ and s_i is the score vector,

$$s_i = \begin{pmatrix} \frac{\partial L_i(\gamma, z_i)}{\partial \gamma_1} \\ \frac{\partial L_i(\gamma, z_i)}{\partial \gamma_2} \\ \vdots \\ \frac{\partial L_i(\gamma, z_i)}{\partial \gamma_K} \end{pmatrix}. \tag{5.7}$$

For a model with a linear index the score vector has a particular simple form,

$$s_i = \frac{\alpha_i g(z_i\gamma){z_i}'[y_i - G(z_i\gamma)]}{G(z_i\gamma)[1 - G(z_i\gamma)]} \quad . \tag{5.8}$$

Of course, the score vector for the reduced form of the spatial autoregressive probit with its non-linear index function would look a bit more involved.

The first step in the empirical analysis is a simple baseline regression where the potential impact the predispositions of neighbors may have on the attitude towards open enrollment in any given district is ignored. Again, the baseline regression is meant as a first, albeit crude test whether the estimation approach can be expected to give meaningful results. Table 5.3 reports the results of a weighted maximum likelihood estimation of a standard probit framework. The latent variable model is specified as

$$y_i^* = x_i\beta + u_i, \tag{5.9}$$

where x_i includes a constant and the control variables discussed above. The first column displays the parameter estimates for the baseline specification model, the second column gives robust standard errors, and the third column provides the average partial effects, i.e. the sample averages of estimated

Table 5.3: Baseline Probit

	Coefficient	S.E.	Slope[a]
Central city	0.007	0.236	0.003
Suburb	-0.290	0.194	-0.105
Elementary school district	0.404	0.294	0.145
High school district	-0.456	0.407	-0.164
Enrollment[b]	0.010	0.008	0.004
Student-teacher ratio	-0.107**	0.047	-0.039
% minority students	-0.951**	0.412	-0.343
Revenues per student[c]	-0.269***	0.098	-0.097
Median household income[c]	-0.001	0.007	-0.000
Log-likelihood		-323.27	
Percent correctly predicted		66.1	

Estimation equation includes state dummies as additional regressors.
Standard errors are Huber-White accounting for sampling weights.
[a] Weighted average of estimated changes in probabilities.
[b] In 1000 students.
[c] In 1000 $.
** 5% significance level.
*** Idem., 1%.

changes in the probability of adoption associated with a change in the explanatory variable. A quick inspection of the results shows that districts with crowded schools, as we presumed, seem to be less willing to open up their local schools for non-resident students. Furthermore, districts with a higher share of minority students are less likely to participate. In addition, districts with lower revenues per student are more inclined towards open enrollment than high revenue districts.

The significance of a number of district characteristics together with the fact that the model correctly predicts almost two thirds of all adoption decisions suggests that all explanatory variables together provide a strong signal for the predisposition of school districts to participate in open enrollment programs. Thus we can hope that the spatial autoregressive model, where we rely on neighbors' mean characteristics in order to identify the impact of neighbors predispositions towards adoption, is capable to provide significant results on

Table 5.4: Spatial autoregressive probit

	SAR model		SARMA model	
	Coefficient	Slope[a]	Coefficient	Slope[a]
Neighbors' predisposition, ϕ	0.440 **	-	0.504 ***	-
	(0.174)		(0.192)	
Central city	-0.058	-0.021	-0.041	-0.015
	(0.235)		(0.207)	
Suburb	-0.221	-0.082	-0.186	-0.071
	(0.176)		(0.175)	
Elementary school district	0.377	0.139	0.343	0.130
	(0.263)		(0.257)	
High school district	-0.699*	-0.258	-0.739*	-0.280
	(0.418)		(0.396)	
Enrollment[b]	0.008	0.003	0.007	0.003
	(0.008)		(0.008)	
Student-teacher ratio	-0.105***	-0.039	-0.093**	-0.035
	(0.041)		(0.043)	
% minority students	-0.689**	-0.254	-0.617*	-0.234
	(0.349)		(0.338)	
Revenues per student[c]	-0.234***	-0.086	-0.199**	-0.075
	(0.084)		(0.096)	
Median household income[c]	-0.000	-0.000	-0.001	-0.000
	(0.005)		(0.005)	
Spatial correlation in errors, ρ	-	-	-0.667**	-
			(0.315)	
Log-likelihood	-319.97		-319.35	
Percent correctly predicted	65.4		65.8	

Estimation equation includes state dummies as additional regressors.
Huber-White standard errors accounting for sampling weights in parentheses.
[a] Weighted average of estimated individual changes in probabilities.
[b] In 1000 students.
[c] In 1000 $.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

potential interdependencies among districts.

Results for the spatial autoregressive probit are presented in Table 5.4. The first two columns display coefficients and average partial effects for the SAR model. Recall that in this model, the errors are assumed to be i.i.d.

The results for the SAR model suggest that predispositions towards adoption of inter-district open enrollment are positively interdependent among school districts. The positive and significant coefficient for the spatial lag, ϕ, measuring the impact of composite neighbors' attitudes on the predisposition towards adoption, indicates that open enrollment policies of local school districts are significantly affected by the anticipated behavior of neighboring districts. Apart from the predisposition of neighbors, a number of district characteristics affect the discrete choice decision whether to participate in open enrollment. High school districts are about 26 percent less likely to allow for inter-district transfers relative to unified school districts. As expected, districts with crowded schools are less inclined towards admitting transfer students. An additional student per teacher lowers the probability that open enrollment policies are adopted by 0.039. At the same time, a one percentage point increase in a district's share of minority students, with all other things being equal, makes the district 0.25 percent less likely to adopt open enrollment. This may reflect the fact that districts with a higher share of minority students will, on average, expect to be less successful in attracting students from elsewhere. Thus, it may not be worthwhile for these districts to adjust their policies towards open enrollment regulations. Furthermore, higher revenues per student make districts less willing to participate in inter-district school choice. $ 1000 of additional revenues per student make the average district 8.6 percent less likely to admit non-resident students.

The next step in the analysis is to check whether the results of the SAR model are biased by spatial error correlation. This is done by estimating a spatial autoregressive moving average (SARMA) probit, where the structural model is given by Eq. (5.2) together with Eq. (5.3). The output for the SARMA model is displayed as the second set of results in Table 5.4. The first and most important thing to note is that allowing for spatial error dependence does not break the link between neighbors' predispositions. On the contrary, the link becomes even stronger: The estimate for ϕ increases, and it is significant at the 1 percent level. At the same time, some evidence is found for the presence of negative spatial error correlation. Note that accounting for spatial error dependence does only marginally increase the log-likelihood of the model. At the same time, the ability of the model to correctly predict the school districts' decisions is slightly strengthened. We can conclude that the positive spatial correlation in adoption decisions is not driven by spatial error correlation.

The results for the control variables are similar to those derived from the SAR model and need not be discussed again in full detail. However, it is worth to note that the average partial effects of the student-teacher ratio and revenues per student are well in line with the corresponding results presented in Chapter 4 (see Table 4.4).

One could question the significance of the spatial effect in modelling the school districts' behavior towards school choice with reference to the share of correctly predicted decisions as a measure for the goodness of fit. Although the percent correctly predicted is very similar for all models, the baseline model outperforms the spatial probit models in terms of predicting the districts' choices. Therefore, as an additional check for the significance of the spatial effect, I computed the corresponding likelihood ratio test statistics. For the null hypothesis that the log-likelihood values are equal for the baseline and the SAR model, the statistic is 6.61. With one degree of freedom, this is significantly different from zero at the 2 percent level. For the test comparing the baseline model with the SARMA model, the statistic is 7.84. With two degrees of freedom, this is significant at the 2 percent level, too. The results of the likelihood ratio tests reveal that not accounting for the neighborhood influence removes a significant amount of information from the system.

Table 5.4 does not provide average partial effects for neighbors' predispo-

Table 5.5: Non-linear least squares

	Coefficient	S.E.
Adoptions of neighbors, ϕ	0.339 ***	0.125
Central city	-0.040	0.089
Suburb	-0.098**	0.051
Elementary school district	0.214 **	0.085
High school district	-0.157	0.183
Enrollment[a]	0.003	0.005
Student-teacher ratio	-0.050***	0.012
% minority students	-0.406***	0.137
Revenues per student[b]	-0.096***	0.026
Median household income[b]	-0.002	0.002

Estimation equation includes state dummies as additional regressors.
[a] In 1000 students.
[b] In 1000 $.
** Idem., 5%.
*** Idem., 1%.

sitions. The reason has been discussed in Section 3.3: ϕ is the coefficient of a latent variable with unknown scale, and it is therefore impossible to evaluate the average partial effect with regard to y^*.

To get an impression of the strength of the neighborhood influence, it is useful to estimate a spatial linear probability model. Recall from Section 3.2 that this model can be written as

$$y_i = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, y_j + x_i \beta + u_i. \tag{5.10}$$

Since the information on y is incomplete, a reduced form is constructed along the same lines as described for the spatial autoregressive probit in Section 3.3. The reduced form is then estimated by non-linear least squares. Of course, given that the school districts' policies are discrete responses, the linear probability model cannot be fully appropriate. Furthermore, the estimation ignores the issue of spatial error dependence. Bearing the limitations of the underlying model in mind, the results presented in Table 5.5 should rather be viewed to suggest an order of magnitude of the interaction effect than to provide robust evidence on it.

The estimate of the interaction coefficient suggests that in their open enrollment policies, the school districts are heavily influenced by their neighbors' actual policies. A one percentage point increase in the share of neighboring districts allowing for inter-district transfers is estimated to make participation 0.34 percent more likely. Again, this is broadly in line with the magnitude of the interaction effect derived from the model with time-lagged spatial effects and the data on the Michigan Open Enrollment Program in Chapter 4 (see Table 4.4).

Chapter 6

Spatial effects in charter school policies: Evidence from California school districts

6.1 Introduction

The policy innovation under consideration in this chapter is charter schools established under the authority of local school districts. Charter schools are public schools operating independently from the existing school district structure. They are exempt from many state and district regulations and provide school officials and teachers with additional professional opportunities. By providing additional choice for parents and students, charter schools also tend to increase competition in the public school system.

The aim of the chapter is to provide evidence on interdependencies among school districts' policies towards charter schools. The evidence is based on charter school policies of California school districts after the state legislature enacted a charter school law in 1992. The analysis is based on two different empirical approaches. The first approach is a finite spatial lag model based on the spatial autoregressive model proposed by Case (1992) and discussed in Section 3.3. The original model is extended in two directions. Firstly, it is applied to panel data of innovations instead of pure cross-section data.

This allows to account for unobserved heterogeneity in a model of innovation activity with both dynamic and spatial effects. Secondly, a finite spatial lag version of the model is used. This, in contrast to the original model, allows to experiment with various spatial structures. Secondly, a cross-sectional IV probit model as described in Section 3.5 is estimated for each year of data.

The spatial dependence in the districts' predisposition towards charter schools is modelled using a new metric of neighborliness which accounts for the small-scale mobility of workers in terms of district-to-district commuting flows. The intuition for this metric is simple: For each district, a set of reference districts is defined. It consists of all districts which are in sufficiently similar commuting distance to the average commuter workplace. For the reference districts, the degree of neighborliness is determined according to differences in median household income. The proposed metric clearly outperforms metrics relying on simple measures of geographical proximity. Most importantly, the impact of neighbors on actual district policies towards charter schools is estimated much more precisely.

The chapter proceeds as follows. Section 6.2 describes charter schools as a school district policy innovation in California. Section 6.3 describes how data on district-to-district commuting flows are used to construct a spatial weighting scheme for the California school districts. Section 6.4 sets out the finite spatial lag probit model as the first empirical approach and presents the results derived from applying this model. The second empirical approach, the IV probit, and the corresponding results are presented in Section 6.5.

6.2 Background: Charter schools in California

A charter school is a publicly funded school that, in accordance with an enabling state statute, has been granted a charter exempting it from selected state or local rules and regulations. It is typically governed by a group or organization under a contract or charter with the state or the local school district. California has been the second state to enact a charter school law in 1992. The law specifies that local school districts may sign a contract with any one or more persons in order to establish a charter school. Although charter schools are part of the public school system, they operate independently from the existing school district structure. They are exempt from most state and district regulations and benefit from substantial autonomy with regard to the curriculum, teaching methods, hiring decisions, and other spending decisions. Since state and district money in California basically follows students, the funding of charter schools is similar to those of traditional public schools.[1]

As explicitly mentioned in the law, it was the intent of the California Legislature to 'encourage the use of different and innovative teaching methods', to 'create new professional opportunities for teachers, including the opportunity to be responsible for the learning program at the schoolsite', to 'provide parents and pupils with expanded choices in the types of educational opportunities that are available within the public school system', and to 'provide vigorous competition within the public school system to stimulate continual improvements in all public schools'.[2]

In California, charter schools can also be authorized by county offices of education and the Department of Education. However, the vast majority of charter schools has been established by local school districts. From 1993 to

---

[1] For details on the California school finance scheme, see Hoxby (2001) and references given there.

[2] The California Charter Schools Act is available at http://www.cde.ca.gov/sp.

2002, among the almost 1,000 school districts in California, 219 signed at least one charter contract. By the end of 2002, the total number of charters under the authority of local school districts was 490.

6.3 A metric of neighborliness based on commuting flows

In this chapter, various weighting schemes are used to define a composite neighbor for each school district. Together with weighting schemes similar to those used in previous chapters, a new metric of neighborliness of local jurisdictions based on commuting flows is proposed and applied. Although a technical description of all schemes is given in the chapter appendix, it seems useful to present the underlying idea and intuition of the commuting-based weighting scheme already at this point.

Given the mere number of almost 1,000 school districts in California, it seems reasonable to assume that in general decision makers at the district level are able to track conditions for policymaking and actual decisions only in a small fraction of all districts. In addition, one can expect the degree to which a given district is considered a reference district by local decision makers to sharply decline with geographical distance. One reason for this to hold is that households, once we take the location of jobs as given, can choose their residence only within commuting distance to the workplace. Thus, for the typical household, the quality of public schools matters only as far as districts within commuting distance to the workplace are concerned. As long as (re-)location decisions of households are a primary concern of local decision makers, we should therefore expect to find interactions in local school policies among districts which, from the point of view of the average household, belong to the same local educational market. Furthermore, the average household will consider to move to another district only if social conditions, amenities and housing prices are sufficiently similar to those at the actual place of residence.

There are, of course, various ways how to determine which districts belong to the same educational market. However, since commuting distances are crucial for location decisions of households, it seems reasonable to define local educational markets with respect to observed commuting flows. In the following, I give the intuition of a simple measure of geographical proximity which is derived from district-to-district workflow data. The commuting-based distance measure can be used to derive spatial weights which, implicitly, define local educational markets. Although the computation of the distance measure described below is computationally demanding for a large number of local jurisdictions, the intuition is straightforward.

The first step in deriving a commuting-based distance measure for a pair of districts is to compute, for both districts, the share of residents commuting to each of the remaining districts. The distance δ is then computed as the sum of differences in those shares in absolute values,

$$\delta_{ij} = \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \left| \frac{n_{ik}}{\sum_{l \neq i,j} n_{il}} - \frac{n_{jk}}{\sum_{l \neq i,j} n_{jl}} \right|, \tag{6.1}$$

where n_{ik} is the number of commuters from i working in k. Note that by construction, $\delta \in [0, 2]$. By applying a suitable distance-decay function to the commuting-based distance measures, it is easy to transform distances into spatial weights. Intuitively speaking, spatial weights based on the commuting-based distance measure state a high degree of neighborliness if, for the average commuter, both districts are in similar commuting distance to the average commuter workplace.

To illustrate the derivation of commuting-based weights, consider an arbitrary example with four districts named $A, \ldots, D$. Suppose the matrix of inter-district commuting flows looks like

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 300 & 0 & 200 \\ 800 & 0 & 200 & 0 \\ 200 & 200 & 0 & 200 \\ 400 & 100 & 100 & 0 \end{bmatrix} \end{array}, \tag{6.2}$$

where in row $i = A, \ldots, D$ the number of commuters from i to all other districts is displayed. The matrix of distances derived from the commuting flows is

$$\Delta_{ij} = \begin{bmatrix} 0 & 2 & 0.2 & 1 \\ 2 & 0 & 1 & 0 \\ 0.2 & 1 & 0 & 0.6 \\ 1 & 0 & 0.6 & 0 \end{bmatrix}. \tag{6.3}$$

A convenient choice for the distance-decay function is $exp(-\delta)$.[3] After row-standardization, the resulting matrix of weights is

$$W = \begin{bmatrix} 0 & 0.102 & 0.619 & 0.278 \\ 0.090 & 0 & 0.245 & 0.665 \\ 0.472 & 0.212 & 0 & 0.316 \\ 0.192 & 0.522 & 0.286 & 0 \end{bmatrix}. \tag{6.4}$$

Note how the similar patterns of commuting flows originating from districts A and C translate into high weights w_{AC} and w_{CA}. The same holds for commuting from B and D. In contrast to this, commuting flows originating from A and B are very dissimilar. Accordingly, the weights w_{AB} and w_{BA} are relatively small.

In many applications, the pattern of commuting flows of a typical jurisdictions will resemble the patterns of only a limited number of other jurisdictions, while the patterns of many other jurisdictions will be very dissimilar. Since the distances derived from commuting flows are bounded from above by 2,

[3] The distance-decay function is only applied to off-diagonal elements, while elements on the diagonal are set to zero.

it may be useful to truncate the resulting weights by setting 'small' weights to zero. An extreme example would be a matrix of truncated weights selecting only the 'nearest neighbor'. In our example, after row-standardizing the weights matrix selecting nearest neighbors would be

$$W^{tr} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{6.5}$$

With regard to the weights derived from inter-district commuting in California, a less extreme truncation is applied. The median number of non-zero weights per row in the matrix of truncated weights is 22.

As alternative criteria to define a set of neighbors for each district, I use geographic contiguity (neighbors are those districts sharing a border) and affiliation to counties (neighbors are those districts belonging to the same county). These are necessarily crude measures of geographical proximity, but they are easy to obtain and may serve as a point of reference for the more flexible metric based on commuting flows.

To account for the fact that only districts with relatively similar social and economic conditions will be considered as alternative places of residence, I also include in the computation of weights a measure for the similarity between districts in terms of household income. This is done by substituting all non-zero weights based on commuting patterns, contiguity or county borders by $exp(-|MHI_i - MHI_j|)$, where MHI is the median household income in 10,000 \$. Of course, with respect to the commuting based weights, the substitution has to be done after truncating 'small' weights. For districts with similar (dissimilar) median household income, the resulting weights will be relatively large (small).

6.4 Applying a finite spatial lag probit for panel data

6.4.1 Estimation approach

While looking at the diffusion of some new political technology among local jurisdictions, we often observe a small group with a high level of activity while the majority of jurisdictions do not adopt the new policy. Often, the difference in adoption behavior will not be solely attributable to differences in the jurisdictions' observable characteristics. Unobserved heterogeneity among jurisdictions is likely to play an important role among the factors driving the observed innovation activity. Therefore, whenever possible, panel data should be used to estimate empirical models of innovation activity in the public sector.

However, even in applications without spatial dependence, the estimation of limited dependent variable models with unobserved effects is complicated by a number of methodological issues. The standard model is the fixed effects logit.[4] The log likelihood of the model is written conditional on $n_i = \sum_{t=1}^{T} y_{it}$, and units of observation with $n_i = 0$ or $n_i = T$ cannot be informative for estimating the parameters. Therefore, the fixed effects logit should not be used to estimate the determinants of innovation activity when $y_{it} = 1$ is a rare event. The situation becomes even more involved when the structural equation incorporates (functions of) lags of the dependent variable. In these cases, the strict exogeneity assumption on the explanatory variables cannot be met, and the standard approaches to deal with unobserved heterogeneity cannot be applied. Recently, techniques have been proposed to deal with unobserved effects in discrete choice panel data models when the strict exogeneity assumption does not hold. Wooldridge (1997) has proposed a transformation which eliminates a multiplicative unobserved effect and provides moment conditions which can be exploited in GMM estimation. The technique is applicable quite

[4] See Wooldridge (2002), p. 490-92 for a discussion.

generally and may be used to estimate count data as well as discrete choice models with unobserved effects. Montalvo (1997) contains a similar approach. Apart from problems to apply the proposed techniques to small samples and in cases with explanatory variables moving slowly over time, it is an open question how GMM approaches with quasi-differencing could be used to estimate models incorporating spatial effects.

In the light of these difficulties, I use an alternative approach to account for unobserved effects proposed by Blundell, Griffith, and Van Reenen (1995). In their study on technological innovations, the authors use the pre-sample history of the variable of interest to control for permanent unobservable differences across firms. Adapting this argument, I argue that there are two main sources of unobserved heterogeneity among local school districts. Firstly, residents and school district officials may differ with respect to their preferences for a new policy in a way that cannot be inferred from observable district characteristics. Secondly, school districts may have different knowledge stocks with regard to running innovative policies. The idea of the approach is to attempt to measure the unobserved heterogeneity by the level of innovation activity that has been observed during some period of time before the unit of observation enters the sample. Technically, this requires defining a 'pre-sample' period and including measures for the innovation activity during this period as proxies for a permanent unobserved effect in the latent variabel equation. Of course, since observations from the pre-sample period cannot be used in estimation, in defining this period one has to compromise between the quality of the approximation of the unobserved effect and the effective sample size. In the estimations presented below, the period 1993-1997 is defined as the pre-sample period, and the period 1998-2002 is used in estimation.

Since only 8 percent of all California school districts take action in the establishment of charter schools between 1993 and 1997, relying on the number of authorized schools would clearly not suffice to obtain a reasonable approximation of the unobserved effect. To alleviate this problem, I make use of information about two related innovative district policies. The first one is the policy towards magnet schools. A magnet program is any program or school within a school designed to attract students away from their school of residence. A magnet program is established and operates on the basis of a particular curriculum theme and/or a particular instructional mode or structure, and may or may not be intended for achieving racial balance. Secondly, I utilize information about district policies towards independent study programs. Having such a program means that individualized education plans are designed to meet the need of individual students. As part of this plan, the student may enter into an agreement with the district to complete specific assignments under the supervision of a teacher. As proxies for the unobserved heterogeneity, I include the number of authorized pre-sample charters, a dummy indicating the existence of a magnet program during the pre-sample period, a similar dummy for an independent study program, and a dummy indicating a zero value in all three proxies describing the pre-sample innovation activity. The last dummy is included to account for the fact that all three proxies are bounded below by zero.

In the following, the empirical model is described more formally. In each school year, any given school district faces the discrete choice decision problem whether to establish one or more charter schools. The model is supposed to explain the observed choices. In setting up the model, Eq. (3.4) is modified by adding subscripts for time periods and proxy variables for a time-constant unobserved effect. The structural equation is written as

$$y^*_{it} = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y^*_{jt} + x_{it}\beta + c_i\gamma + u_{it}, \quad i = 1, \ldots, N \quad t = 1, \ldots, T, \tag{6.6}$$

where c_i is the (1×4) vector of proxies and γ is the corresponding vector of coefficients. In order to separate spatial correlation in the latent variables from spatial error correlation, the u's are allowed to be spatially correlated according to

$$u_{it} = \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} \epsilon_{jt} + \epsilon_{it}, \tag{6.7}$$

where ϵ_{it} is assumed to be homoscedastic and serially uncorrelated. Note that this form of spatial error dependence assumes zero correlation among errors of neighbors of second and higher order.

For $|\phi| < 1$ and appropriately defined weights, repeated substitution in a stacked version of Eq. (6.6) provides us with

$$y_t^* = (I + \phi W + \phi^2 W^2 + \cdots)(x_t \beta + c\gamma + u_t). \tag{6.8}$$

The difficulties in estimating the reduced form of the spatial autoregressive model have been discussed in Section 3.3. A simple way to generalize the approach proposed by Case (1992) is to approximate the true reduced form equation by the finite spatial lag model

$$y_t^* = D\,(x_t \beta + c\gamma + u_t), \tag{6.9}$$

where

$$D \equiv I + \phi W + \phi^2 W^2 + \phi^3 W^3. \tag{6.10}$$

The main advantage of the finite spatial lag model is that estimation does not require any restriction on the W matrix.[5] As described in Section 3.3, the variance-normalizing transformation proposed by Case (1992) is used before estimation to restore homoscedasticity.

The vector of district characteristics includes the school district's size (enrollment), its ethnic composition (share of minority students), the student-teacher ratio, the share of students eligible for free or reduced price lunch, an

[5] To check for any bias that might be induced by the approximation, I used a block diagonal W matrix based on the districts' affiliation to counties and estimated both the true reduced form and the finite spatial lag model. The results were almost identical.

indicator for districts located in central cities, and indicators for elementary and high schools districts.

6.4.2 Data

Table 6.1 presents descriptive statistics on the establishment of charter schools between 1998 and 2002 and on the covariates used in the empirical analysis. Data on California charter schools come from the California Department of Education.[6] Data on the location of districts relative to urban areas are from the National Center for Education Statistics (NCES, Common Core of Data (CCD), Local Education Agency (School District) Universe Survey Data). Data on district types, enrollment, teachers, minority students, and students eligible for free or reduced price lunch are provided by the California Department of Education (California Basic Educational Data System (CBEDS), Public School Enrollment and Staffing and CalWORKS Data Files). Data on alternative education programs used to approximate unobserved effects are from the CBEDS School Information Form (SIF), Sections G.

For the construction of spatial weights I used place-to-place commuting data from the Bureau of Transportation Statistics (CTPP 2000 Part 3 journey-to-work (JTW) tables) and data on district median household income from the NCES School District Demographics System.

6.4.3 Results

Table 6.2 presents results for the finite spatial lag probit explaining the establishment of charter schools by California school districts in the years 1998-2002. The weight matrix used in the underlying regressions defines a com-

[6]See the Charter School Locator, online available at http://www.cde.ca.gov/ds/si/cs.

Table 6.1: California school districts - descriptive statistics

	Mean	S.D.	Min	Max
Indicator for at least one additional charter	0.048	0.214	0	1.00
Enrollment[a]	6.16	25.5	0.007	747
Student teacher ratio	19.5	3.19	4.10	50.6
% minority students	0.450	0.281	0	1.00
% students free lunch	0.431	0.266	0	1.00
Central city	0.127	0.333	0	1.00
Suburb	0.470	0.499	0	1.00
Elementary school district	0.591	0.492	0	1.00
High school district	0.090	0.287	0	1.00
Fixed effect (FE) proxies				
FE 1: Number of charters	0.133	0.639	0	10.0
FE 2: Magnet program	0.102	0.303	0	1.00
FE 3: Independent study program	0.573	0.495	0	1.00
FE 4: Zero innovation activity	0.399	0.490	0	1.00

Balanced panel of 941 California school districts, 1998-2002. The FE proxies are based on the pre-sample period 1993-1997.
[a] In 1000 students.

posite neighbor for each district based on the districts' affiliation to counties. The results displayed are based on a balanced panel of 941 districts.[7] The table reports coefficients, standard errors and average partial effects (slopes) for two estimations, one where the spatial weights are determined without any consideration of differences in median household income, and one where those differences are taken into account. Estimation results are shown for the spatial lag, an indicator for at least one additional charter school in the previous three periods, and the proxy variables for the fixed effect derived from innovation activity in the years 1993-1997. The results for the remaining covariates are omitted. The coefficient of the spatial lag, ϕ, represents the impact of the composite neighbor's predisposition towards the establishment of additional charter schools on the district's own predisposition. The estimated ϕ's show a positive sign in both specifications. However, as long as the definition of

[7] I had to exclude some districts due to changes in school district boundaries.

Table 6.2: Spatial probit, composite neighbor based on counties

	Coefficient	S.E.	Slope[a]
Weighting scheme A: Weights do not account for income differences			
Spatial lag, ϕ	0.103	0.147	-
Charter in previous three years	0.471 ***	0.119	0.038
FE 1:	0.135 **	0.060	0.011
FE 2:	0.169	0.107	0.014
FE 3:	-0.157	0.232	-0.013
FE 4:	-0.433*	0.235	-0.035
Weighting scheme B: Weights do account for income differences			
Spatial lag, ϕ	0.234 *	0.121	-
Charter in previous three years	0.467 ***	0.118	0.038
FE 1:	0.132 **	0.059	0.011
FE 2:	0.157	0.108	0.013
FE 3:	-0.176	0.229	-0.014
FE 4:	-0.455*	0.234	-0.037

Probit on balanced panel of 941 districts (1998 to 2002).
Standard errors (S.E.) are robust for serial correlation in scores.
Estimations account for spatial error correlation.
Other covariates include district type dummies, city and suburb dummies, a Los Angeles City Unified dummy, enrollment, student-teacher ratio, % minority students, % students free lunch, and year effects.
Log of likelihood: -765.86 (weighting scheme A), -764.90 (weighting scheme B).

[a] Average of estimated individual changes in probabilities.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

composite neighbors does not account for differences in median household income, the null of zero spatial dependence cannot be rejected at conventional levels of significance. If those differences are used to differentiate the weights among reference districts, the coefficient of the spatial lag is weakly significant. The coefficients of the indicator for lagged establishments of charter schools are positive and highly significant. The average partial effects indicate that districts which signed a charter in the previous three years are 3.8 percent more likely to authorize additional schools. In addition, two of the four proxy variables for the fixed effect are significant at least at the 10 percent level. As expected, districts which were early movers in establishing charter schools are more likely to sign charters for additional schools. At the same time, districts which did not use any of the included innovative policies during the pre-sample period are significantly less likely to set up charter schools. The results for the FE proxies can be interpreted in two ways. Firstly, early movers may be more experienced with innovative policies, and this may positively affect their willingness to sign additional charters. Secondly, the proxy variables may account for the impact of unobservable, time-invariant district effects.

Corresponding results for contiguity based spatial weights are displayed in Table 6.3. Again, the model fails to detect any interdependence in the predisposition towards the establishment of charter schools if differences in median household income among districts belonging to the same local educational market are ignored. As before, a weakly significant positive effect is found with income differences taken into account. The coefficients and average partial effects of the lagged dependent variable and the FE proxies are similar to those derived before and need not be discussed again.

From the results discussed so far it would be difficult to assess whether the diffusion of charter schools among California school districts was significantly affected by interactions among local school boards. Have the school districts by and large been unaffected by their neighbors, or did the model fail to detect an interaction among districts which was actually present? The results displayed in Table 6.4 suggest a straight answer to this question. Now, the weights are based on the similarity of district-to-district commuting patterns. As before, the null that spatial interaction among districts is absent can not be rejected if differences in median household income are not taken into account. However, in contrast to the results discussed so far, the coefficient of the spatial lag is found to be significant at the 1 percent level with spatial weights being defined as interactions between commuting-based indicators for reference districts and the measure for the similarity of median household in-

Table 6.3: Spatial probit, composite neighbor based on contiguity

	Coefficient	S.E.	Slope[a]
Weighting scheme C: Weights do not account for income differences			
Spatial lag, ϕ	0.090	0.173	-
Charter in previous three years	0.474***	0.122	0.040
FE 1:	0.139**	0.059	0.012
FE 2:	0.172*	0.105	0.015
FE 3:	-0.143	0.228	-0.012
FE 4:	-0.429*	0.231	-0.036
Weighting scheme D: Weights do account for income differences			
Spatial lag, ϕ	0.230*	0.132	-
Charter in previous three years	0.491***	0.118	0.041
FE 1:	0.138**	0.057	0.011
FE 2:	0.161	0.106	0.013
FE 3:	-0.136	0.223	-0.011
FE 4:	-0.448*	0.231	-0.037

Probit on balanced panel of 941 districts (1998 to 2002).
Standard errors (S.E.) are robust for serial correlation in scores.
Estimations account for spatial error correlation.
Other covariates include district type dummies, city and suburb dummies, a Los Angeles City Unified dummy, enrollment, student-teacher ratio, % minority students, % students free lunch, and year effects.
Log of likelihood: -766.32 (weighting scheme C), -764.85 (weighting scheme D).
[a] Average of estimated individual changes in probabilities.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

come. Thus, once the set of reference districts is appropriately defined, there is strong evidence for the districts' attitudes towards charter schools being positively interdependent.

As an additional check for the significance of the differences in spatial effects, I estimated a simple probit with all exogenous explanatory variables but without any interaction among equations for different districts. The log-likelihood for this model is -766.54. Likelihood ratio (LR) tests reveal that this value is not significantly different from the likelihoods of the spatial probit model if weighting schemes B or D are used. However, using weighting

Table 6.4: Spatial probit, composite neighbor based on commuting patterns

	Coefficient	S.E.	Slope[a]
Weighting scheme E: Weights do not account for income differences			
Spatial lag, ϕ	0.224	0.249	-
Charter in previous three years	0.493 ***	0.121	0.040
FE 1:	0.137 **	0.059	0.011
FE 2:	0.164	0.109	0.013
FE 3:	-0.154	0.232	-0.012
FE 4:	-0.429*	0.237	-0.035
Weighting scheme F: Weights do account for income differences			
Spatial lag, ϕ	0.390 ***	0.123	-
Charter in previous three years	0.484 ***	0.112	0.041
FE 1:	0.127 **	0.057	0.011
FE 2:	0.148	0.106	0.013
FE 3:	-0.140	0.230	-0.012
FE 4:	-0.415*	0.236	-0.035

Probit on balanced panel of 941 districts (1998 to 2002).
Standard errors (S.E.) are robust for serial correlation in scores.
Estimations account for spatial error correlation.
Other covariates include district type dummies, city and suburb dummies, a Los Angeles City Unified dummy, enrollment, student-teacher ratio, % minority students, % students free lunch, and year effects.
Log of likelihood: -766.00 (weighting scheme E), -763.30 (weighting scheme F).

[a] Average of estimated individual changes in probabilities.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

scheme F, the LR test statistic is 6.49. With two degrees of freedom, this is significantly different from zero at the 4 percent level. Hence, among the weight matrices accounting for differences in median household income, the matrix based on district-to-district commuting patterns clearly outperforms the weight matrices where the definition of local educational markets rests on county borders or contiguity of districts. Taken together, the results suggest that the most appropriate way to define composite neighbors is to select reference districts based on district-to-district commuting patterns and, at the same time, to account for the degree of similarity in social and economic conditions.

As mentioned in Section 3.3, a distinctive disadvantage of limited dependent variable models with interaction in latent variables is that it is not possible to derive from the parameter estimates an average partial effect for the spatial lag. Hence, we cannot say anything about the magnitude of the effect of neighbors' predispositions on the probability that districts establish additional schools. The next subsection presents additional evidence on interaction effects in district policies towards charter schools produced from a model which does not suffer from this disadvantage.

6.5 Applying the instrumental variables probit

6.5.1 Estimation approach

Compared to the data on policy innovation used in Chapters 4 and 5, the available information on the school districts' charter school policies in California is particularly rich. The analysis provided in this chapter makes use of this information by presenting evidence on interdependencies among district policies towards charter schools based on a second estimation approach, the IV probit. In contrast to the panel data approach discussed before, the dependent variable is constructed as an indicator taking value one if, in the given school year, the school district operates at least one charter school.[8] Since charter school closures are rare events, pooling observations from various years would lead to strong serial correlation in the dependent variable. To avoid any problem associated with such correlation, the identification of the parameters relies exclusively on cross section variation.

[8] Recall that in Section 6.4, the dependent variable is an indicator for the establishment of *additional* charter schools.

The structural equation of the IV probit is written as

$$y_i^* = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_j + x_i \beta + u_i. \tag{6.11}$$

As mentioned in Section 3.5, the linear combination of other districts' policies on the right hand side of Eq. (6.11) is endogenously determined. Spatially transformed exogenous characteristics are used as instruments in the IV probit estimation.

6.5.2 Data

As before, data for the years 1998-2002 are used to estimate the IV probit. An advantage of using this data instead of data from earlier years is the availability of data on standardized test scores at the district level. With data on student test scores at hand, it is possible to control for the effect of average student achievement on the willingness of local decision makers to experiment with new policies.[9] High school dropout rates are readily available and, in general, could be used as a substitute. Note, however, that dropout rates are not defined for about 60 percent of school districts in California which operate only elementary schools.

Since the definition of the dependent variable differs from the one used for the estimation of the finite spatial lag model, Table 6.5 shows descriptive statistics for the indicator on charter school operation in California school districts for the years 1998-2002. The numbers shown for the various cross-sections are based on the samples of districts which are used in estimation.

[9] I did not use the test score data to control for student achievement when estimating the finite spatial lag probit because the balanced panel approach would have suffered from too many missing values in the test score variable. The reason for missing values in the test score data is that information on student test scores is not available for a number of small districts.

Table 6.5: Summary statistics for dependent variable, 1998-2002

Year	1998	1999	2000	2001	2002
Mean	0.106	0.140	0.178	0.203	0.230
S.D.	0.308	0.348	0.382	0.403	0.421
Nob	925	926	918	915	912

Table 6.6: Summary statistics for covariates, 1998-2002 (Nob=4596)

	Mean	S.D.	Min	Max
Enrollment[a]	5.71	10.3	0.036	142
Student-teacher ratio	19.9	2.69	7.17	50.6
% minority students	0.462	0.277	0	1.00
% students free lunch	0.430	0.261	0	1.00
Student achievement[b]	0.567	2.54	-6.32	9.77
Central city	0.137	0.344	0	1.00
Elementary school district	0.552	0.497	0	1.00
High school district	0.100	0.300	0	1.00

[a] In 1000 students.
[b] Deviation of average math test score from statewide average in percent.

The sample size varies over time due to variation in the total number of school districts and missing values in covariates. Note that Los Angeles City Unified has been excluded from the sample as an influential observation. It served more than 740,000 students in 2002, which is more than five times the number of students in the second largest district, San Diego City Unified.

Table 6.6 provides descriptive statistics for the covariates including the test score variable which is supposed to control for student achievement. The test score variable describes the deviation of student achievement from the statewide average and is constructed as follows. First of all, for each grade level the percent deviation of a district's average score from the statewide average is computed. The district's average deviation is then determined as the average deviation over all grade levels, weighted by the number of students tested. I use the math test scores to derive the achievement variable because they are available for all grade levels. Using reading scores gives very similar estimation results. Test score data are from the Standardized Testing and Reporting (STAR) program of the California Department of Education.[10]

6.5.3 Results

In the following, estimation results for the IV probit model are presented. As mentioned in Section 3.5, the IV probit can be estimated by conditional maximum likelihood (CML) and by two-stage procedures. It should be stressed that results derived from two-stage procedures are somewhat difficult to interpret since the coefficients are only estimated up to scale. Therefore, the coefficients from Newey's two-stage procedure should not be directly compared to those derived from the CML approach. However, the two-step estimates can still be used to test for significant relationships. The informational content of the estimate for the interaction coefficient is thus comparable to the informational content of the corresponding estimate from the autoregressive model.

The first set of results for the IV probit model displayed in Table 6.7 illustrates the endogeneity problem associated with regressing an indicator for innovation activity on the contemporaneous activity of a composite neighbor. The table reports only results for 2002. The first two columns show average partial effects, i.e. changes in probability associated with a change in explanatory variables evaluated at sample means, from naive probit estimates of Eq. (6.11). The average partial effects shown in the last two columns have been estimated by IV probit using the CML approach. Neighbors' innovation activity is instrumented by neighbors' exogenous characteristics as described in Section 3.5.[11] The weighting schemes used to define composite neighbors

[10] Available online at http://www.cde.ca.gov/ta.

[11] Due to missing values for student test score averages, student achievement cannot be used in constructing instruments. Therefore, composite neighbors' enrollment, student-

Table 6.7: Marginal changes in probabilities, 2002: Naive probit vs. IV probit

Estimation approach	Probit		IV probit (CML)	
Weighting scheme	*C*	*D*	*C*	*D*
Neighbors' innovations	0.295***	0.307***	-0.085	0.057
	(0.065)	(0.058)	(0.191)	(0.181)
Enrollment[a]	0.012***	0.012***	0.013***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)
Student-teacher ratio	0.004	0.004	0.006	0.005
	(0.006)	(0.006)	(0.006)	(0.006)
% minority students	-0.335***	-0.326***	-0.372***	-0.358***
	(0.081)	(0.081)	(0.083)	(0.084)
% students free lunch	-0.303***	-0.299***	-0.349***	-0.334***
	(0.104)	(0.103)	(0.106)	(0.106)
Student achievement	-0.061***	-0.059***	-0.063***	-0.062***
	(0.010)	(0.010)	(0.011)	(0.011)
Central city	0.068	0.064	0.084*	0.078*
	(0.044)	(0.043)	(0.045)	(0.045)
Elementary school district	-0.024	-0.025	-0.028	-0.027
	(0.033)	(0.033)	(0.034)	(0.034)
High school district	-0.042	-0.044	-0.064	-0.057
	(0.046)	(0.046)	(0.045)	(0.045)
Test of exogeneity: $\chi^2\,(1)$	-	-	4.65	2.19
Test of exogeneity: p-value	-	-	0.031	0.139

Standard errors of average partial effects in parentheses.
[a] In 1000 students.
* 10% significance level of underlying coefficient.
*** Idem., 1%.

are C (1st and 3rd column) and D (2nd and 4th column). Table 6.7 clearly shows that naive estimates not accounting for the endogeneity of neighbors' innovation activity can produce very misleading results. While the average partial effects derived from a simple probit are highly significant, the estimated positive impact of neighbors' innovation activity vanishes once an IV probit approach is used. Using a Wald test of exogeneity of neighbors' innovation activities,[12] the null is rejected at the 5 percent level of significance for the

teacher ratio, percent minority students, percent students free lunch, and weighted averages of the dummy variables are used as instruments.

[12] See Wooldridge (2002) p. 476 for details on tests of exogeneity based on CML estimation

estimation with weighting scheme C. Based on the IV probit with weighting scheme D, the p-value is 0.139, which should still be interpreted as evidence against the null of exogeneity. The results of IV probit estimations using the cross-sections from other years and with weighting schemes A, B and E are similar to those presented in Table 6.7. In all cases it turns out that a highly significant effect of neighbors' innovation activity derived from a naive probit disappears once the simultaneity of y_i and $\sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_j$ is accounted for. So far, the findings confirm the results derived from the finite spatial lag model.

We now turn to results for weighting scheme F (interaction between indicators for similar commuting patterns and the measure for similarity in household income). Table 6.8 presents coefficient estimates derived from Newey's two-stage procedure. The model has been estimated using the cross sections for the years 1998-2002. The main result is that the probability for running at least one charter school is significantly and positively related to charter school policies in neighboring districts. However, based on the coefficients displayed we cannot say anything about the magnitude of this effect. Note that the first stage regression in the IV probit procedure produces significant coefficients for at least six out of seven instruments. Hence, in all cross-sections the partial correlation among the endogenous regressor and the instruments is strong.

It is more interesting to look at results showing the strength of the interaction effect. Table 6.9 reports average partial effects evaluated at sample means and the corresponding standard errors derived from the CML procedure. For the sake of brevity, the underlying coefficients are omitted.

Again, the estimated effects of neighbors' innovation activity show a positive sign in all estimations. Moreover, in all estimations, the null of no dependence between neighbors' innovation activity and a districts' attitude towards charter schools can be rejected at the one percent level of significance. The

of the IV probit.

Table 6.8: Coefficient estimates derived from Newey's two-stage procedure

	1998	1999	2000	2001	2002
Neighbors' innovations	2.98 ***	3.12 ***	1.93 **	1.92 ***	1.86 ***
	(1.06)	(1.02)	(0.754)	(0.630)	(0.601)
Enrollment[a]	0.030 ***	0.029 ***	0.046 ***	0.046 ***	0.045 ***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
Student-teacher ratio	0.027	0.031	-0.011	0.001	0.013
	(0.027)	(0.022)	(0.024)	(0.024)	(0.023)
% minority students	-0.748 **	-0.785 **	-0.915 ***	-0.960 ***	-0.990 ***
	(0.349)	(0.354)	(0.327)	(0.308)	(0.314)
% students free lunch	-0.653	-0.560	-1.30 ***	-1.40 ***	-1.09 ***
	(0.488)	(0.381)	(0.419)	(0.411)	(0.384)
Student achievement	-0.087 *	-0.085 **	-0.188 ***	-0.203 ***	-0.200 ***
	(0.048)	(0.042)	(0.043)	(0.043)	(0.041)
Central city	0.306 *	0.483 ***	0.419 ***	0.368 **	0.192
	(0.169)	(0.158)	(0.150)	(0.150)	(0.144)
Elementary school district	-0.122	-0.147	-0.010	-0.012	-0.061
	(0.146)	(0.135)	(0.130)	(0.126)	(0.121)
High school district	-0.509 **	-0.382 *	-0.223 **	-0.215 **	-0.205
	(0.252)	(0.215)	(0.200)	(0.195)	(0.189)

Standard errors in parentheses.
[a] In 1000 students.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

average partial effects are between 0.434 and 0.587, indicating that a one percentage point increase in the share of neighbors operating at least one charter school is estimated to increase the innovation probability by 0.43 to 0.59 percentage points. The school districts' behavior toward charter schools is also affected by a number of other factors. As expected, larger districts are more likely to offer charter schools. Furthermore, districts with a higher share of minority students are consistently less likely to operate charter schools. However, there is also weak evidence for the notion that charter schools are more common in central urban areas, and that high school districts are somewhat less inclined towards signing charter school contracts. Finally, there is strong evidence suggesting that districts with lower student achievement are more likely to run charter schools than districts with higher average test scores. More successful schooling producers are less often experimenting with charter

Table 6.9: Marginal changes in probabilities derived from IV probit (CML)

		1998	1999	2000	2001	2002
Neighbors' innovations		0.456 ***	0.587 ***	0.434 ***	0.477 ***	0.507 ***
		(0.165)	(0.188)	(0.164)	(0.152)	(0.163)
Enrollment[a]		0.005 ***	0.005 ***	0.010 ***	0.012 ***	0.012 ***
		(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Student-teacher ratio		0.004	0.006	-0.002	0.000	0.004
		(0.004)	(0.004)	(0.005)	(0.006)	(0.006)
% minority students		-0.111**	-0.145**	-0.207***	-0.239***	-0.272***
		(0.052)	(0.064)	(0.073)	(0.076)	(0.086)
% students free lunch		-0.100	-0.106	-0.297***	-0.350***	-0.302***
		(0.072)	(0.070)	(0.092)	(0.100)	(0.104)
Student achievement		-0.013*	-0.016**	-0.043***	-0.051***	-0.055***
		(0.007)	(0.008)	(0.009)	(0.010)	(0.011)
Central city		0.054	0.111 ***	0.111 ***	0.103 **	0.056
		(0.034)	(0.042)	(0.044)	(0.046)	(0.043)
Elementary school district		-0.018	-0.028	-0.003	-0.003	-0.017
		(0.022)	(0.025)	(0.029)	(0.031)	(0.033)
High school district		-0.058***	-0.059**	-0.046	-0.049	-0.052
		(0.020)	(0.026)	(0.036)	(0.040)	(0.044)
Test of exogeneity:	χ^2 (1)	1.21	1.29	0.42	0.68	0.04
	p-value	0.271	0.256	0.515	0.408	0.833

Standard errors of average partial effects in parentheses.
[a] In 1000 students.
* 10% significance level of underlying coefficient.
** Idem., 5%.
*** Idem., 1%.

schools as a new production technology.

As mentioned in Chapter 3, a simple check of the robustness of the results of the spatial probit is to run a 2SLS regression. Apart from its obvious limitations, in many applications the linear probability model gives reasonable estimates of the partial effects. Results for the 2SLS approach are presented in Table 6.10. Although in four out of five estimations the coefficients of neighbors' innovation activity are somewhat smaller than the average partial effects derived from the probit model, the results are remarkably similar. Again, the impact of charter schools operating in neighboring districts is strongest in 1999. A one percentage point increase in the share of neighbors running charter schools is estimated to increase the probability for the operation of such

Table 6.10: Coefficient estimates derived from two-stage least squares

	1998	1999	2000	2001	2002
Neighbors' innovations	0.430 **	0.506 ***	0.470 ***	0.467 ***	0.456 ***
	(0.214)	(0.160)	(0.135)	(0.123)	(0.124)
Enrollment[a]	0.008 ***	0.007 ***	0.011 ***	0.011 ***	0.011 ***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Student-teacher ratio	0.003	0.006	-0.001	0.000	0.003
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
% minority students	-0.142**	-0.167***	-0.187***	-0.191***	-0.223***
	(0.061)	(0.056)	(0.063)	(0.063)	(0.070)
% students free lunch	-0.062	-0.061	-0.218***	-0.288***	-0.237**
	(0.076)	(0.068)	(0.074)	(0.084)	(0.092)
Student achievement	-0.012**	-0.014**	-0.034***	-0.042***	-0.044***
	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)
Central city	0.049	0.107 **	0.100 **	0.088 *	0.048
	(0.037)	(0.042)	(0.045)	(0.046)	(0.044)
Elementary school district	-0.016	-0.029	-0.010	-0.009	-0.024
	(0.024)	(0.027)	(0.030)	(0.031)	(0.033)
High school district	-0.066*	-0.066	-0.038	-0.043	-0.044
	(0.037)	(0.043)	(0.047)	(0.049)	(0.050)
Overidentifying $F(6)$	6.86	7.50	7.41	4.67	4.77
restrictions: p-value	0.334	0.277	0.285	0.587	0.574

Robust standard errors in parentheses.
[a] In 1000 students.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

schools by 0.43 to 0.51 percentage points. It is also worth noting that in all estimations the overidentifying restrictions are not rejected at any reasonable level. This finding suggests that we can have some confidence in the overall set of instruments used to identify the spatial interaction in charter school policies.

The results from the IV probit and the spatial LPM confirm the findings derived from the finite spatial lag model presented in the previous section. Once reference districts are defined according to commuting patterns and similarity in household income, in selecting their charter school policies school boards are found to be heavily affected by corresponding policies in reference districts.

The empirical findings support the notion that strategic interaction among local governments is an important driving force for the diffusion of new political technologies.

Appendix to Chapter 6: Technical description of spatial weights

In this chapter, I use six different weighting matrices to define a composite neighbor for each district. Three of these matrices are based on a measure for the similarity of median household income in 1999 (*MHI*, in 10,000 \$). For two districts i and j, it is computed as

$$\tilde{w}_{ij} = exp(-|MHI_i - MHI_j|). \tag{6.12}$$

In addition, three different measures for geographic proximity are used. In the following, the construction of all weighting matrices $W = \{w_{ij}\}_{i,j=1}^{N}$ is described. Note that in all matrices $w_{ii} = 0$, and that weights are row-standardized such that $\sum_j w_{ij} = 1$.

Weight matrix A (Affiliation to counties): Indicators c_{ij} for districts belonging to the same county as i are used to compute the weights as $w_{ij} = c_{ij} / \sum_{j \neq i} c_{ij}$.

Weight matrix B (Interaction between indicators for affiliation to the same county and the measure for similarity in household income): Weights are based on interactions between indicators c_{ij} and $\tilde{w}_{ij}$,

$$w_{ij} = c_{ij}\, \tilde{w}_{ij} \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} c_{ij}\, \tilde{w}_{ij} \right)^{-1} . \tag{6.13}$$

Weight matrix C (Geographic contiguity): Indicators b_{ij} for a shared border between i and j or some common territory[13] are used to compute the weights

[13]Elementary and high school districts in California often share some common territory,

as $w_{ij} = b_{ij} / \sum_{j \neq i} b_{ij}$.

Weight matrix D (Interaction between contiguity indicators and the measure for similarity in household income): Weights are based on interactions between contiguity indicators b_{ij} and $\tilde{w}_{ij}$,

$$w_{ij} = b_{ij}\,\tilde{w}_{ij}\left(\sum_{\substack{j=1\\ j\neq i}}^{N} b_{ij}\,\tilde{w}_{ij}\right)^{-1} \tag{6.14}$$

Weight matrix E (Similarity of commuting patterns): First of all, a district-to-district commuting matrix is computed. Since commuting data at district level are unavailable, Census-designated places are assigned to school districts, and Census data on place-to-place commuting in 1999 are used to compute commuting flows between districts.[14] As a next step, a distance δ for each pair of districts is computed as the difference in commuting shares going to all other districts,

$$\delta_{ij} = \sum_{\substack{k=1\\ k\neq i,j}}^{N} \left| \frac{n_{ik}}{\sum_{l\neq i,j} n_{il}} - \frac{n_{jk}}{\sum_{l\neq i,j} n_{jl}} \right|, \tag{6.15}$$

where n_{ik} is the number of commuters from i working in k. Note that by construction, $\delta \in [0, 2]$. By applying the distance-decay function $exp(-\delta)$, distance measures are transformed into spatial weights. Since many δ's are close to the maximum of 2, it is useful to truncate the resulting weight matrix. After row-standardization, this is done by a simple row-wise procedure: starting with the smallest and moving on to the bigger, weights are set to zero as

i.e. they overlap. Of course, 'contiguity' indicators must account for that.

[14] Places and school districts are non-nested geographical structures. Often, commuting to or from some place may be assigned to various school districts. I solve this problem by proportionally assigning commuters to districts, where the proportion for each of the districts involved is given by the share of the place's area that is covered by the district.

long as the sum of the remaining weights exceeds the threshold of 0.1. After this procedure, the median number of non-zero weights is 22.

For 311 out of 941 districts in the sample, the imputed number of commuters is too low to obtain reliable weights.[15] For these districts, I substitute contiguity-based weights as described above.

Weight matrix F (Interaction between indicators for similar commuting patterns and the measure for similarity in household income): In order to obtain a matrix of indicators for similar commuting patterns, all strictly positive weights in the commuting-based weight matrix are set to unity. Indicators are then interacted with the measure for similarity in household income, $\tilde{w}_{ij}$, and the usual row-standardization is applied.

[15] I construct commuting-based weights only for districts with at least 50 out-commuters.

Chapter 7

Concluding remarks

This book sheds light on policy innovation in local jurisdictions. Based on the example of US school districts it provides evidence on strategic interaction among local governments in the adoption of new policies. The evidence suggests that inter-jurisdictional competition is a driving force for political change. The picture that emerges from the empirical chapters is one where local governments actively engage in relative comparison and performance evaluation across jurisdictions. This leads to a situation of policy diffusion, where the implementation of a new policy in one jurisdiction triggers adoptions of that policy in other jurisdictions. While in the long run the ultimate success or failure of the new policy will decide on its permanent implementation or abandonment, inter-jurisdictional competition seems to speed up the process of initial adoption and experimentation with the new policy. While the empirical investigations presented in the previous chapters, for practical reasons, focus on situations where the adoption of some new policy at the local level (i.e. the school districts) is triggered by political decisions at a higher tier of government (i.e. the states) in a federal political system, the empirical results suggest that inter-jurisdictional competition will also provide governments with incentives to independently experiment with innovative solutions to various problems of governance. The material presented in this book is therefore supportive to the view of 'laboratory federalism' in the sense that a large number of independent jurisdictions offers favorable conditions for finding innovative solutions to difficult political problems.

Since the book (to the best of my knowledge) is the first systematical approach to deal empirically with the issue of strategic interaction and policy innovation among local jurisdictions, it is worth noting that the findings are derived from various empirical models and methods. The evidence on positive spatial dependence in adoption decisions is fairly robust across models and estimation strategies. Moreover, the evidence is based on two different local policies, adoption of inter-district school choice and establishment of independent public schools, so-called charter schools. Finally, the school districts under investigation in the empirical parts of the book are from six US states: California, Idaho, Kansas, Massachusetts, Michigan, and Ohio. The robustness of the main findings across estimation techniques, policies, and samples makes me believe that for understanding the behavior of local governments in public sector innovation it is crucial to take into account interdependencies among jurisdictions.

Given that local governments seem to interact with one another when it comes to the implementation of new policies, quite naturally the question arises whether it is possible to discriminate between competing theoretical explanations for that interaction. In general, it seems to be promising to exploit natural experiments to achieve identification in this context. One could, for instance, argue that learning externalities are the driving force for the observed interaction among local governments. To check the validity of the argument, one could estimate the interaction effect in a situation where learning is absent. A situation similar to the one the Michigan school districts experienced in 1996, when the state initiated a voluntary inter-district choice program, lends itself to test the learning hypothesis, since without any history of adoptions there are no effects from learning by definition.

To gain insight into the relative importance of information spillovers versus resource flows and (re-)location decisions of households, it is useful to exploit barriers to mobility to achieve identification. An example has been provided in Chapter 4, where restrictions to the mobility of students in inter-district transfers are exploited. The results show that the districts' willingness to participate in the school choice program was positively related to the level of participation among neighboring districts because an environment more open to inter-district transfers induces stronger competition for students, i.e. for mobile resources.

The examples given above show that natural experiments are unequivocally promising to learn more about the incentives for public sector innovation. In many situations, however, there is not sufficient exogenous variation to achieve identification, and convincing strategies for discriminating between competing theoretical explanations for policy interdependencies are unavailable.

The issues discussed in this book are closely related to the long-lasting debate whether political decentralization is beneficial or not. Instead of entering the debate I would like to point to the fact that evidence on public sector innovation in one setting (say decentralized decision making) is uninformative with regard to the benefits of decentralization as long as corresponding evidence for the other setting (centralized decision making) is missing. Since only one setting has been analyzed the book does not contribute to the decentralization debate. It does, however, show that the forces which are necessary to make decentralization the better choice with respect to the ability of the public sector to produce and implement policy innovations are at work.

References

Anselin, L. (1988): *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Dordrecht/Boston/London.

Anselin, L., and R. J. Florax (1995): "Small sample properties of tests for spatial dependence in regression models: Some further results," in *New directions in spatial econometrics*, ed. by L. Anselin, and R. J. Florax, pp. 22–74. Springer, Berlin.

Arsen, D., D. Plank, and G. Sykes (1999): *School choice policies in Michigan: The rules matter*. Michigan State University, East Lansing.

Baicker, K. (2005): "The spillover effects of state spending," *Journal of Public Economics*, 89, 529–544.

Belfield, C. R., and H. M. Levin (2002): "The Effects of Competition on Educational Outcomes: A Review of US Evidence," mimeo, Columbia University.

Belleflamme, P., and J. Hindriks (2005): "Yardstick competition and political agency problems," *Social Choice and Welfare*, 24, 155–169.

Beron, K. J., and W. P. M. Vijverberg (2004): "Probit in a spatial context: A Monte Carlo analysis," in *Advances in spatial econometrics*, ed. by L. Anselin, R. J. Florax, and S. J. Rey, pp. 169–195. Springer, Berlin.

Berry, F. S. (1994): "Sizing up state policy innovation research," *Policy Studies Journal*, 22, 442–56.

Berry, F. S., and W. D. Berry (1990): "State lottery adoptions as policy innovations: An event history analysis," *The American Political Science Review*, 84(2), 395–415.

——— (1992): "Tax innovation in the states: Capitalizing on political opportunity," *American Journal of Political Science*, 36(3), 715–742.

Besley, T., and A. Case (1995): "Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition," *The Amercian Economic Review*, 85(1), 25–45.

Besley, T., and M. Smart (2002): "Does tax competition raise voter welfare?," CEPR Discussion Papers 3131.

Blundell, R., R. Griffith, and J. Van Reenen (1995): "Dynamic count data models of technological innovation," *Economic Journal*, 105(429), 333–344.

Bogart, W. T., and B. A. Cromwell (1997): "How much more is a good school district worth?," *National Tax Journal*, 50, 215–232.

——— (2000): "How much is a neighborhood school worth?," *Journal of Urban Economics*, 47, 280–305.

Bolton, P., and C. Harris (1999): "Strategic experimentation," *Econometrica*, 67, 349–374.

Boockman, B. (2001): "The ratification of ILO conventions: A hazard rate analysis," *Economics and Politics*, 13, 281–309.

Bordignon, M., F. Cerniglia, and F. Revelli (2003): "In search of yardstick competition: A spatial analysis of Italian municipality property tax setting," *Journal of Urban Economics*, 54, 199–217.

——— (2004): "Yardstick competition in intergovernmental relationships: Theory and empirical predictions," *Economics Letters*, 83, 325–333.

Bradbury, K. L., C. J. Mayer, and K. E. Case (2001): "Property tax limits, local fiscal behavior, and property values: evidence from Massachusetts under proposition 2 1/2," *Journal of Public Economics*, 80, 287–311.

Brett, C., and J. Pinkse (1997): "Those taxes are all over the map: A test for spatial dependence of municipal tax rates in British Columbia," *International Regional Science Review*, 20, 131–151.

Brueckner, J. K. (1998): “Testing for strategic interaction among local governments: The case of growth controls,” *Journal of Urban Economics*, 44(3), 438–467.

——— (2003): “Strategic interaction among governments: An overview of empirical studies,” *International Regional Science Review*, 26, 175–188.

Brueckner, J. K., and L. A. Saavedra (2001): “Do local governments engage in strategic property tax competition?,” *National Tax Journal*, 54(2), 203–229.

Büttner, T. (1999): “Determinants of tax rates in local capital income taxation: A theoretical model and evidence from Germany,” *Finanzarchiv*, 56, 363–388.

——— (2001): “Local business taxation and competition for capital: The choice of the tax rate,” *Regional Science and Urban Economics*, 31, 215–245.

——— (2003): “Tax base effects and fiscal externalities of local capital taxation: Evidence from a panel of German jurisdictions,” *Journal of Urban Economics*, 54, 110–128.

Canon, B., and L. Baum (1981): “Patterns of adoption of tort law innovations: An application of diffusion theory to judicial doctrines,” *The American Political Science Review*, 75, 975–987.

Case, A. (1992): “Neighborhood influence and technological change,” *Regional Science and Urban Economics*, 22, 491–508.

Case, A., J. R. Hines, and H. S. Rosen (1993): “Budget spillovers and fiscal policy interdependence: Evidence from the States,” *Journal of Public Economics*, 52, 285–307.

Caudill, S. B., J. M. Ford, F. G. Mixon Jr., and T. C. Peng (1995): “A discrete-time hazard model of lottery adoption,” *Applied Economics*, 27, 555–561.

Clotfelter, C. (1975): “The effect of school desegregation on housing prices,” *Review of Economics and Statistics*, 57, 446–451.

Coate, S., and S. Morris (1995): "On the form of transfers to special interests," *Journal of Political Economy*, 105, 1210–1235.

Colwell, P. F., and K. L. Guntermann (1984): "The value of neighborhood schools," *Economics of Education Review*, 3, 177–182.

Couch, J. F., and W. F. Shugart II (1993): "Private school enrollment and public school performance," *Public Choice*, 76, 301–312.

Coughlin, C. C., T. A. Garrett, and R. Hernández-Murillo (2003): "Spatial probit and the geographic patterns of state lotteries," Federal Reserve Bank of St. Louis, Working Papers: 2003-042.

Dee, T. S. (1998): "Competition and the quality of public schools," *Economics of Education Review*, 17, 419–427.

Dubin, R. (1995): "Estimating logit models with spatial dependence," in *New directions in spatial econometrics*, ed. by L. Anselin, and Raymond J.G.M. Florax, pp. 229–242. Springer, Berlin.

Egger, P., M. Pfaffermayr, and H. Winner (2005): "An unbalanced spatial panel data approach to US state tax competition," *Economics Letters*, 88, 329–335.

Epple, D., and R. E. Romano (1998): "Competition between private and public schools, vouchers, and peer-group effects," *The American Economic Review*, 88, 33–62.

Fredriksson, P. G., and D. L. Millimet (2002): "Strategic interaction and the determination of environmental policy across U.S. states," *Journal of Urban Economics*, 51(1), 101–122.

Garret, T. A., G. A. Wagner, and D. C. Wheelock (2003): "A spatial probit analysis of state banking regulation," Federal Reserve Bank of St. Louis, Working Papers: 2003-044 A.

Gill, H. L. (1983): "Changes in city and suburban house prices during a period of expected school desegregation," *Southern Economic Journal*, 50, 169–184.

Gray, V. (1973): "Innovation in the States: A diffusion study," *American Political Science Review*, 67, 1174–85.

Grosskopf, S., K. J. Hayes, L. L. Taylor, and W. L. Weber (2001): "On the determinants of school district efficiency: Competition and monitoring," *Journal of Urban Economics*, 49, 453–478.

Hautsch, N., and S. Klotz (2003): "Estimating the neighborhood influence on decision makers: Theory and an application on the analysis of innovation decisions," *Journal of Economic Behavior and Organization*, 52, 97–113.

Hayashi, M., and R. Boadway (2000): "An empirical analysis of intergovernmental tax interaction: The case of business income taxes in Canada," *Canadian Journal of Economics*, 34, 481–503.

Holmstrom, B. R. (1982): "Moral hazard in teams," *Bell Journal of Economics*, 13, 324–40.

Hoxby, C. M. (1994): "Do private schools provide competition for public schools?," NBER Working Paper No. 4978.

——— (1999): "The productivity of schools and other local public goods producers," *Journal of Public Economics*, 74, 1–30.

——— (2000): "Does competition among public schools benefit students and taxpayers?," *The American Economic Review*, 90, 1209–1238.

——— (2001): "All school finance equalizations are not created equal," *Quarterly Journal of Economics*, 116(4), 1189–1231.

——— (2003): "School choice and school productivity (or, could school choice be a rising tide that lifts all boats?)," in *The economics of school choice*, ed. by C. M. Hoxby, pp. 287–341. University of Chicago Press, Chicago.

Jud, G. D., and J. M. Watts (1981): "Schools and housing values," *Land Economics*, 57, 459–470.

Kelejian, H. H., and I. R. Prucha (1998): "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *Journal of Real Estate Finance and Economics*, 17(1), 99–121.

Kotsogiannis, C., and R. Schwager (2006a): "On the incentives to experiment in federations," *Journal of Urban Economics*, forthcoming.

——— (2006b): "Political uncertainty and policy innovation," *Journal of Public Economic Theory*, forthcoming.

Krasker, W. S., E. Kuh, and R. E. Welsch (1983): "Estimation for dirty data and flawed models," in *Handbook of econometrics, Volume I*, ed. by Z. Griliches, and M. D. Intriligator, pp. 651–698. North-Holland Publishing Company, New York.

LeSage, J. P. (2000): "Bayesian estimation of limited dependend variable spatial autoregressive models," *Geographical Analysis*, 32, 19–35.

McMillen, D. P. (1992): "Probit with spatial autocorrelation," *Journal of Regional Science*, 32, 335–248.

——— (1995): "Spatial effects in probit models: A Monte Carlo investigation," in *New directions in spatial econometrics*, ed. by L. Anselin, and R. J. Florax, pp. 189–228. Springer, Berlin.

Michigan Department of Treasury (2002): *School finance reform in Michigan. Proposal A: Retrospective.* Office of Revenue and Tax Analysis.

Mintrom, M. (1997): "Policy entrepreneurs and the diffusion of innovation," *American Journal of Political Science*, 41(3), 738–770.

Montalvo, J. G. (1997): "GMM estimation of count-panel-data models with fixed effects and predetermined instruments," *Journal of Business and Economic Statistics*, 15(1), 82–89.

Mukand, S. W., and D. Rodrik (2002): "In search of the holy grail: Policy convergence, experimentation and economic performance," CEPR Discussion Paper No. 3525.

Nalebuff, B. J., and J. E. Stiglitz (1983): "Prizes and incentives: Towards a general theory of compensation and competition," *Bell Journal of Economics*, 14, 21–43.

National Center for Education Statistics (1996): *1993-94 Schools and Staffing Survey: Data file user's manual.* U.S. Department of Education, Office of Educational Research and Improvement, Washington D.C.

——— (1998): *The Schools and Staffing Survey (SASS) and Teacher Followup Survey (TFS) CD-ROM: electronic codebook and public-use data for three cycles of SASS and TFS.* U.S. Department of Education, Office of Educational Research and Improvement, Washington D.C.

Newey, W. K. (1987): "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics*, 36(3), 231–250.

Newmark, C. M. (1995): "Another look at whether private schools influence public school quality: Comment," *Public Choice*, 82, 367–373.

Oates, W. E. (1999): "An essay on fiscal federalism," *Journal of Economic Literature*, 37, 1120–1149.

Pinkse, J., L. Shen, and M. Slade (2003): "Dynamic spatial probit with fixed effects using one step GMM: An application to mine operating decisions," unpublished manuscript.

Pinkse, J., and M. E. Slade (1998): "Contracting in space: An application of spatial statistics to discrete-choice models," *Journal of Econometrics*, 85, 125–154.

Rapp, G. C. (2000): "Agency and choice in education: Does school choice enhance the work effort of teachers?," *Education Economics*, 8, 37–63.

Reback, R. (2005): "House prices and the provision of local public services: Capitalization under school choice programs," *Journal of Urban Economics*, 57(2), 275–301.

Revelli, F. (2005): "On spatial public finance empirics," *International Tax and Public Finance*, 12, 475–492.

——— (2006): "Performance rating and yardstick competition in social service provision," *Journal of Public Economics*, forthcoming.

Rivers, D., and Q. H. Vuong (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

Sandström, M., and F. Bergström (2005): "School vouchers in practice: Competition will not hurt you," *Journal of Public Economics*, 89, 351–380.

Savage, R. L. (1985): "Diffusion research traditions and the spread of policy innovations in a federal system," *The Journal of Federalism*, 15, 1–27.

Strumpf, K. S. (2002): "Does government decentralization increase policy innovation?," *Journal of Public Economic Theory*, 4(2), 207–241.

Tiebout, C. M. (1956): "A pure theory of local expenditures," *The Journal of Political Economy*, 64, 416–424.

Tyran, J.-R., and R. Sausgruber (2005): "The diffusion of policy innovations - an experimental investigation," *Journal of Evolutionary Economics*, 15, 423–442.

Vedder, R., and J. Hall (2000): "Private school competition and public teacher salaries," *Journal of Labor Research*, 21, 161–168.

Walker, J. L. (1969): "The diffusion of innovations among the American states," *The American Political Science Review*, 63(3), 880–899.

Weimer, D. L., and M. J. Wolkoff (2001): "School performance and housing values: Using non-contiguous district and incorporation boundaries to identify school effects," *National Tax Journal*, 54, 231–253.

Wildasin, D. E. (1989): "Interjurisdictional capital mobility: Fiscal externality and a corrective subsidy," *Journal of Urban Economics*, 25, 193–212.

Wilson, J. D. (1986): "A theory of interregional tax competition," *Journal of Urban Economics*, 19, 296–315.

——— (1999): "Theories of tax competition," *National Tax Journal*, 52, 269–304.

Wooldridge, J. M. (1997): "Multiplicative panel data models without the strict exogeneity assumption," *Econometric Theory*, 13, 667–678.

——— (2002): *Econometric analysis of cross section and panel data.* The MIT Press, Cambridge (Mass.).

Wrede, M. (2001): "Yardstick competition to tame the Leviathan," *European Journal of Political Economy*, 17, 705–721.

Zodrow, G. R., and P. Mieszkowski (1986): "Pigou, Tiebout, property taxation, and the underprovision of local public goods," *Journal of Urban Economics*, 19, 356–370.