Policy diffusion in space and time: The case of charter schools in California school districts

Johannes Rincke*

Final version, February 2007

Abstract

The literature on policy diffusion and emulation has gathered descriptive evidence on the behavior of state and local governments in the adoption of policy innovations, but structural evidence is largely missing. Using spatial econometric models for limited dependent variables, this paper presents evidence on the diffusion of charter schools among California school districts. Strong positive spatial correlation either in the districts' predisposition towards the establishment of charter schools or between the districts' predisposition and lagged adoptions of neighbors is found. The results support the view that in decentralized systems of government the diffusion of new political technologies is driven by horizontal interactions among jurisdictions.

Keywords: policy innovation, policy diffusion, spatial correlation, charter schools

JEL Classification: D78, H77

* Department of Economics, University of Munich.
Postal address: Akademiestr. 1/II, D-80799 Munich. Phone: +49/89/2180-6753, Fax: +49/89/2180-6296, email: johannes.rincke@lrz.uni-muenchen.de.

1 Introduction

The ability of decentralized systems of government to adjust to newly emerging problems of governance often depends critically on how local governments interact with one another in the adoption of new policies. Suppose, for instance, that local governments have private information on the prospect of some new political technology, and that each government's private information is valuable for other jurisdictions' governments. If some jurisdictions adopt the new policy and governments can observe each other's choices, this will affect the remaining governments' inference on the quality of the new policy. While it is in general unclear whether this kind of interaction is beneficial or not for the system as a whole,[1] it is important to know how the diffusion of new policies is affected by interdependencies among adoption decisions of local governments.

However, despite the fact that the literature in political science has assembled extensive descriptive evidence on what has been labeled "political lesson drawing", "policy diffusion", "policy transfer" or "policy band-wagoning",[2] our understanding of the diffusion of new political technologies is still limited. One particular difficulty in producing robust structural evidence on policy diffusion is caused by the fact that observable characteristics as well as unobservable shocks are often correlated across

[1]Note that the example involves a learning externality that could, for instance, lead to a situation where adoption decisions are inefficiently postponed.

[2]See Dolowitz and Marsh (1996) and Stone (1999) for references.

jurisdictions. Therefore, to gain insight into what drives policy diffusion, it is necessary to disentangle interaction between local governments from policies being related to one another across jurisdictions due to correlated environments. Appropriate techniques to identify such interaction among local governments have been widely applied in empirical work focussing primarily on spatial effects in the choice of tax rates, public goods provision, and expenditures.[3] In addition, there are a few studies presenting structural evidence on the diffusion of non-fiscal policy instruments, among them Brueckner (1998), dealing with the stringency of growth controls in California cities, and Fredriksson and Millimet (2002), addressing environmental policies of the US states. In many cases, however, policy diffusion takes the form of governments adopting policies which have not been available before. In federal political systems, for instance, we regularly observe higher tiers of government triggering processes of policy diffusion by transferring certain responsibilities to governments at lower tiers. Similarly, governments at higher tiers often start new programs and let local governments independently decide whether they want to participate in these programs.[4] Such adoption decisions are most adequately modeled using a

[3] See, e.g., Besley and Case (1995), Brett and Pinkse (1997), Hayashi and Boadway (2000), Büttner (2001), Brueckner and Saavedra (2001), and Bordignon et al. (2003) for strategic interaction in the choice of tax rates. Bivand and Szymanski (1997) estimate reaction functions for the cost of public goods provision, while Case et al. (1993) and Baicker (2005) deal with expenditures. See Brueckner (2003) for a survey.

[4] A well known example is the introduction of voluntary public school choice in the US. In many states, the state legislatures have adopted laws concerning inter-district transfers of students, and the local school districts must decide whether they admit non-resident

framework for limited dependent variables. While the growing literature on spatial econometrics has suggested various techniques to identify parameters of spatial interaction in limited dependent variable models, applications focussing on the diffusion of new political technologies in decentralized systems of government are largely missing. The paper contributes to the literature in providing structural evidence on interactions among local governments in the adoption of newly invented policies.

The policy innovation under consideration in this paper is the establishment and operation of charter schools in California 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 California charter school law is from 1992 and specifies the conditions under which charter schools may be established under the authority of local school districts.

The aim of the paper is to provide evidence on interdependencies among the school districts' policies towards charter schools. The evidence is based on a unique data set providing information on all California charter schools established between 1993 and 2002. Two models are used to test for

students at their local public schools. Rincke (2006a) and Rincke (2006b) provide evidence on the diffusion of school choice policies among US school districts.

interdependencies among the districts' charter school policies. While the first model accounts for contemporaneous interaction in the propensity to establish charter schools, the second model allows for interaction involving time lags. The evidence from both models points to positive interaction among school districts in the adoption of charter school policies. The results suggest that, irrespective of the fact that the conditions for adopting policies supportive to charter schools were similar for many districts either because of common shocks or similar social and economic characteristics, interaction among local decision makers has contributed significantly to the diffusion of charter schools.

The paper proceeds as follows. Section 2 briefly describes charter schools as a school district policy innovation in California and discusses several potential driving forces for adoption decisions to be interrelated. Section 3 sets out the empirical strategy of the paper. In Section 4, the results are presented and discussed, and Section 5 concludes.

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 regulations and, in addition, may negotiate with the sponsor district exemptions from certain district policies. Since state and district funding basically follows students, the funding of charter schools is similar to those of traditional public schools.

As explicitly mentioned in the Charter School 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".[5]

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 2002, among the almost 1,000 school districts in California, 219 signed at

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

least one charter contract. By the end of 2002, the total number of charters under the authority of local school districts was 490.

One can think of various reasons for school district officials to implement policies supportive to charter schools. Firstly, as Kotsogiannis and Schwager (2006) point out in their theoretical analysis of local decision makers' career concerns, implementing new policies is often more demanding than running 'business as usual'. More able decision makers will then, on average, be more successful in using new policies than less able decision makers. In cases where residents are only imperfectly informed about the ability of local representatives, decision makers may take advantage of the opportunity to signal above-average ability by implementing innovative policies. The signalling incentive may thus drive the diffusion of policy innovations among local jurisdictions. Hence, in the present example, operating charter schools may be seen as an opportunity by local decision makers to signal their willingness and ability to serve residents with high-quality public schooling. Interaction among local decision makers may then quite naturally arise either from learning externalities (this is the case discussed by Kotsogiannis and Schwager 2006) or from yardstick competition. While learning externalities provide decision makers with incentives to postpone the adoption of new policies until information on outcomes in other jurisdictions becomes available, yardstick competition may accelerate the diffusion by rewarding decision makers for running similar policies as in reference jurisdictions. However, even in a situation where decision makers behave non-strategically, information spillovers between jurisdictions can lead local decision makers to adopt policies from other jurisdictions. The simplest case is learning from the behavior or experience of others: If local officials are not sure which policy is optimal, they may extract information from what policies are chosen in reference jurisdictions or from the outcomes of policies observed elsewhere.

Another force that may lead local governments to adopt novel policies is Tiebout choice. Suppose that households welcome the introduction of charter schools for the expanded choices that are provided by a new type of public school. School districts that are concerned about local property values may then adopt policies supportive to charter schools in order to increase the attractiveness of the local public school system.

It is also worth noting that charter schools do not have a specified attendance area and may even enroll students who do not live within the school district. The charter school idea is thus closely related to other forms of increased parental choice within the public school sector such as intra- and inter-district school choice. Since districts obtain additional state funds if they enroll non-resident students, district officials may use charter schools to compete for non-resident students. Note, however, that in case of over-enrollment, resident students are preferred over transfer students. Furthermore, charter schools must specify means by which the school's student body will reflect the racial composition of the general population living in the school district. In practice, the opportunity to attract transfer students by means of charter schools may therefore be quite limited.

Furthermore, school districts in California may elect to participate in a statewide voluntary inter-district choice program. Districts can thus compete for non-resident students without the need to run a charter school.

Apart from the incentives discussed so far, an empirical investigation of district policies towards charter schools should take account of a number of potentially relevant district characteristics. The establishment of charter schools will change the rules and regulations under which public schooling is provided at least for some part of the student body. Thus, the establishment of charter schools has an immediate impact on the 'technology' school districts are using in their educational production. The better the outcomes of that production, the less inclined a district should be towards changes in the production technology. We should therefore expect the districts' willingness to establish charter schools to be negatively related to average student achievement. Looking at the input side of educational production, we may expect the preferences of district officials towards charter schools to also depend on certain key characteristics of the student body. In general, the more favorable average student characteristics are in terms of potential academic achievement, the less district officials should value the additional opportunities and the greater flexibility provided by charter schools compared to traditional public schools.

3 Empirical strategy and data

3.1 Estimation strategy

Inherently, the diffusion of any innovation is both a dynamic and a spatial process. Empirical models of innovation activity should therefore allow for both dynamic and spatial effects. While it is straightforward to account for pure time dependence, the modeling of spatial dependence is more involved. As mentioned above, there are a variety of underlying forces which might result in cross sectional dependence of innovation behavior. Depending on the focus of the analysis, one might want to adjust the modeling of interaction among local governments. For instance, while yardstick competition should lead to contemporaneous interaction, the potential presence of learning externalities suggests to account for time lags in decision makers' reactions to adoptions observed elsewhere. In principle, it would also be desirable to account for a variety of interaction effects (contemporaneous vs. time lagged, different functional forms, etc.) in an integrated approach to test for their explanatory power.

As will be discussed below, in a discrete choice framework it is difficult to allow for the presence of different spatial effects within one model. The empirical evidence on the diffusion of charter schools among California school districts is therefore derived from two distinct empirical models. While the first model (presented in Subsection 3.2) allows for contemporaneous interaction in the districts's predisposition towards charter schools, the second model (discussed in Subsection 3.3) accounts for spatial effects involving time lags.

Before going into details, however, it seems appropriate to make some general remarks on the estimation of dynamic models of innovation activity. While looking at the diffusion of some new political technology among local jurisdictions, we are likely to 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.

Unfortunately, the standard estimation techniques for discrete choice models involving unobserved heterogeneity rely on a strict exogeneity assumption with respect to the explanatory variables. In dynamic models with (functions of) lagged dependent variables among the explanatory variables, this assumption does not hold. For some fairly special cases, in particular models with multiplicative unobserved effects, GMM estimators have been proposed to circumvent the problem (Wooldridge 1997, Montalvo 1997). It is unclear, however, whether these approaches can be extended to dynamic models incorporating spatial effects.

In the light of these difficulties, I use an alternative approach to account for unobserved effects proposed by Blundell et al. (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 to the context of this study, 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 the permanent unobserved effect. 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 all estimations presented below, the period 1993-1997 is defined as the pre-sample period, and the period 1998-2002 is used in estimation. Data from the pre-sample period are used only to derive proxies for unobserved school district effects.

Since only 8 percent of all 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 unobserved district effects. 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 to indicate 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.

3.2 A finite spatial lag model for the diffusion of charter schools

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, with a special focus on the diffusion of charter schools as a new school district policy. The model draws on Case (1992), who discussed the interaction among farmers in rural Java in the adoption of a new harvesting tool. Case's approach is extended to cover general spatial structures and applied to panel data of innovations instead of pure cross-section data.[6]

A straightforward way to formalize interdependence among the charter school policies of local school districts is to introduce contemporaneous spatial correlation into a standard latent variable framework for limited dependent variables. To illustrate the approach, let y^*_{it} describe school district i's predisposition to establish a charter school in period t. The predisposition determines actual adoptions y_{it} according to

$$y_{it} = 1[y^*_{it} > 0] \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where $1[\cdot]$ is the indicator function. For the latent variable, consider the

[6] Case (1992) assumes a block-diagonal matrix of spatial weights in order to derive an estimable equation. This is equivalent to assuming uniform weights within certain groups of jurisdictions and restricting the dependence between jurisdictions from different groups to zero.

following linear specification incorporating a spatial lag,

$$y_{it}^* = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, y_{jt}^* + x_{it}\beta + \alpha_i + u_{it}, \tag{2}$$

where w_{ij} is an appropriately defined weight for district j among all districts $j \neq i$, x_{it} is a vector of characteristics including a constant and a full series of period effects, α_i is an unobserved district effect, and u_{it} is an error term. Note that we allow x_{it} to contain functions of lagged y_{it} and that the weights define a composite neighbor for each district. Eq. (1) and (2) define a model where the diffusion of charter school policies is driven by contemporaneous correlation in the predisposition of local governments towards adoption. With $\phi > 0$, the model describes a situation where, ceteris paribus, a school district is more inclined towards running charter schools the stronger the propensity of the composite neighbor to do so. The model with contemporaneous interaction in latent variables is consistent with yardstick competition among local governments as well as policy competition induced by Tiebout choice.

For the reasons discussed above, the structural form Eq. (2) is modified by substituting the unobserved effect, α_i, by the proxy variables derived from the pre-sample period. This gives

$$y_{it}^* = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, y_{jt}^* + x_{it}\beta + c_i\gamma + u_{it}, \tag{3}$$

where c_i is the (1×4) vector of proxies and γ is the corresponding vector of coefficients. Note that the latent variable model cannot be estimated as stated in Eq. (3) since the predisposition of the composite neighbor is unobservable. In order to obtain an estimable reduced form equation, the N equations for each period are stacked. Using matrix notation, this gives

$$y_t^* = \phi W \, y_t^* + x_t\beta + c\gamma + u_t. \tag{4}$$

Note that W is now a $(N \times N)$ matrix of weights. Repeated substitution provides us with the reduced form

$$y_t^* = (I + \phi W + \phi^2 W^2 + \cdots)(x_t\beta + c\gamma + u_t). \tag{5}$$

For $|\phi| < 1$ and appropriately defined weights, the reduced form describes a stable relationship between i's predisposition and own characteristics as well as the characteristics of the composite neighbor, the composite neighbor's neighbors, etc. As mentioned above, Case (1992) proposed an approach to estimate the coefficients from the reduced form under the assumption that W is block-diagonal and contains block-wise uniform weights. A simple way to generalize Case's approach is to approximate the true reduced form by the finite spatial lag model

$$y_t^* = D\,(x_t\beta + c\gamma + u_t), \tag{6}$$

where

$$D \equiv I + \phi W + \phi^2 W^2 + \phi^3 W^3. \tag{7}$$

The difference between the true reduced form and the finite spatial lag model is that in the finite spatial lag model, the impact of neighbors of forth (and higher) order has been suppressed.

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

$$u_{it} = \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} \, \epsilon_{jt} + \epsilon_{it}, \tag{8}$$

where ϵ_{it} is assumed to be homoscedastic and serially uncorrelated. In matrix notation, this gives

$$u_t = (I + \rho W)\epsilon_t. \tag{9}$$

Note that this form of spatial error dependence assumes zero correlation among errors of neighbors of second and higher order.

The main advantage of the finite spatial lag model compared to Case (1992) is that estimation does not require to put any restriction on the W matrix.[7] Note furthermore that probit estimation of the true reduced form taking full account of the structure in the variance covariance matrix imposed by the spatial dependence among latent variables would require to evaluate a multi-dimensional normal probability. Given that there is no analytical solution even for a univariate normal cdf, various simulators have been used in spatial probit applications.[8] However, if one is willing to ignore the structure in the off-diagonal elements of the variance covariance matrix, estimation of the parameters from Eq. (6) can rely on standard maximum likelihood techniques without any need for simulations.

A final issue with respect to the composite error $v = D\,(I + \rho W)\,\epsilon$ and its variance covariance matrix $\Omega = E(vv')$ must be addressed. Note that parameter estimates based on probit or logit specifications using Eq. (6) would be inconsistent due to v being heteroscedastic. Therefore, as proposed by Case (1992), Eq. (6) is pre-multiplied by the inverse of the square root of diag(Ω) before estimation in order to restore homoscedasticity. The model parameters can be estimated using any of the standard maximum likelihood techniques for limited dependent variables. To estimate the model, one simply has to plug the rows of $diag(\Omega)^{-1/2}D\,(x_t\beta + c\gamma)$ into the usual likelihood function.

[7] 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.

[8] See Beron and Vijverberg (2004) for a discussion.

3.3 Spatial effects with time lags

The diffusion model presented in the previous section is not without alternatives. In contrast to i's predisposition depending on the contemporaneous predisposition of a composite neighbor, interaction among local governments could also occur with a time lag. The structural form for the latent variable could then look like

$$y^*_{it} = \phi \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}\, \tilde{y}_{jt} + x_{it}\beta + c_i\gamma + u_{it}, \tag{10}$$

where $\tilde{y}_{jt}$ is an indicator for lagged adoptions in district j. Suppose, for instance, that

$$\tilde{y}_{jt} = \begin{cases} 1: & \sum_{s=1}^{3} y_{j,t-s} > 0 \\ 0: & \text{otherwise} \end{cases} . \tag{11}$$

Eq. (10) accounts for the impact of adoption decisions of the composite neighbor from $t-1$ to $t-3$. A more flexible specification would allow for a separate effect of indicators for each lag. Instead of $\sum_{j \neq i} w_{ij}\, \tilde{y}_{jt}$, the structural form would then incorporate $\sum_{j \neq i} w_{ij}\, y_{j,t-1}$, $\sum_{j \neq i} w_{ij}\, y_{j,t-2}$, and $\sum_{j \neq i} w_{ij}\, y_{j,t-3}$ as separate explanatory variables. Note that spatial effects with time lags are consistent with policy competition induced by Tiebout choice and learning on optimal policies from the observation of other districts' choices.

Technically, the estimation of the model with a lagged impact of neighbors' choices is less involved. As long as we are willing to make the assumption that the composite neighbor's lagged choices are exogenous to i's current predisposition, parameter estimation can directly be based on Eq. (10). Furthermore, given that the model accounts for unobserved effects on the predisposition towards adoption, one should not be too concerned about spatial error dependence. The reason is that spatial correlation in time-invariant unobserved effects is probably the most important source for correlation between u_{it} and $u_{j,t-1}, \ldots, u_{j,t-3}$. If the time-invariant unobserved effect is successfully removed from the error, it is unlikely that the parameter estimates are biased due to spatial dependence in unobserved components.

As mentioned above, it would be desirable to incorporate the contemporaneous and the time lagged interactions into a single model of innovation behavior, allowing to test what type of interaction is actually driving diffusion. In practice, however, the problem to jointly identify the various spatial effects in a single equation model is paramount because, implicitly, *neighbors'* realizations for spatially transformed lagged adoptions would have to be used as instruments for neighbors' actual predispositions.

3.4 Spatial weights

The model requires the definition of appropriate spatial weights for the definition of composite neighbors. 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, while the average household as well as district officials will consider information on the quality of other districts' public schools to be relevant in general, they will focus their attention on districts within commuting distance to the average household's workplace. As long as (re-)location decisions of residents 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 more similar social conditions, amenities and housing prices, the more interested residents will be in the other district's public school system.

Based on the preceding discussion, I derive spatial weights using information on geographic contiguity (neighbors are those districts sharing a border) and the similarity between districts in terms of median household income. Although contiguity is a relatively crude measure for geographical proximity, it will, at least in most cases, provide us with a reasonable approximation of local educational markets. The weights are determined in three steps. As a first step, a measure for the similarity of median household income in 1999 (*MHI*, \$ 10,000's) is constructed.[9] For two districts i and j, it is computed as

$$\tilde{w}_{ij} = exp(-|MHI_i - MHI_j|). \tag{12}$$

As a second step, the $\tilde{w}_{ij}$'s are interacted with indicators b_{ij} for a shared border between i and j or some common territory[10]. Finally, the diagonal elements of the weight matrix $W = \{w_{ij}\}_{i,j=1}^{N}$ are set to zero, and the off-diagonal elements are row-standardized such that $\sum_j w_{ij} = 1$. This gives (for $i \neq j$)

$$w_{ij} = b_{ij}\, \tilde{w}_{ij} \left(\sum_{\substack{k=1 \\ k \neq i}}^{N} b_{ik}\, \tilde{w}_{ik} \right)^{-1}. \tag{13}$$

[9]Data on district median household income is from the School District Demographics System of the National Center for Education Statistics (NCES), online availabe at http://nces.ed.gov/surveys/sdds/downloadmain.asp.

[10]Elementary and high school districts in California often share some common territory, i.e. they overlap. Of course, "contiguity" indicators must account for that.

3.5 School district characteristics and data

As discussed in Section 2, the vector of explanatory variables should contain educational achievement as an output measure as well as social and economic characteristics of the student body which are relevant for educational production in terms of the quality of input factors.

Educational achievement is covered by an average test score variable. It 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.[11] A problem with the STAR data is that there are missing values for a number of small districts. To solve this problem, a simple imputation technique is used. The score is regressed (for all observations with valid test scores) on a large number of district characteristics, and the estimated coefficients are then used to predict missing test scores. The number of districts for which test scores are imputed ranges from 55 (1999) to 64 (2001 and 2002).

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

Social and economic characteristics of the student body are taken into account in estimations by controlling for the share of Hispanic students and the share of students living in poverty. Including the share of Blacks and other races does not alter the results, while the estimated coefficients for these groups are not statistically different from zero. In order to construct a suitable poverty measure, the number of students who are eligible for free or reduced price lunch is related to the total number of students. Finally, since large school districts are more likely to operate charter schools even if charters are evenly distributed among all public schools, the estimations control for school district size measured by enrollment. Data on enrollment, 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 Data Files and CalWORKS Data Files).[12]

Data on district policies are also provided by the California Department of Education. The dependent variable is constructed using the Charter School Locator, the Department's online database on California charter schools.[13] Data on alternative education programs used to approximate unobserved effects are from the CBEDS School Information Form (SIF), Sections G. Table 1 presents summary statistics for all the variables.

[12] Online available at http://www.cde.ca.gov/ds/sd/cb.

[13] Online available at http://www.cde.ca.gov/ds/si/cs.

4 Results

This section presents estimation results for a probit specification of the finite spatial lag model as well as for the diffusion model incorporating time-lagged spatial effects. Let us first turn to results from the finite spatial lag probit.

Table 2 presents results for probit estimations explaining the establishment of charter schools by California school districts in the years 1998-2002. As motivated and discussed in Section 3.4, the weight matrix used in the underlying regression defines for each district a composite neighbor based on the interaction between a measure for the similarity of median household income and indicators for common borders. The results displayed are based on a balanced panel of 941 districts.[14]

The table reports coefficients, standard errors and average partial effects (slopes). Note that the vector of explanatory variables includes an indicator

[14]Due to the homoscedasticity-restoring transformation discussed in Section 3.2, it is computationally involved to estimate an unbalanced panel. In addition, there was a relatively large number of changes in district boundaries between 1993 and 2002, mostly due to consolidation of small districts. Consolidations resulted in the number of local school districts decreasing from 1,012 in 1993 to 984 in 2002. Since the construction of the FE proxies requires information on district policies in the pre-sample period 1993-1997, most of the newly created consolidated districts could not be included in the estimation anyway. Therefore, districts with incomplete records are treated as missing observations in all years 1998-2002, i.e. the impact of these districts on their neighbors' policies is ignored.

for at least one additional charter school in the previous three periods. The model thus accounts for the potential of lagged own adoptions to affect a district's propensity to establish additional charter schools.

The interaction coefficient, 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 shows a positive sign and is significantly different from zero at the 5 percent level. This suggests that there is positive interdependence in the school districts' attitudes towards charter schools. As a given district's neighbors' anticipated benefits from the establishment of charter schools increase, the original district's willingness to set up charter schools is found to be significantly increased.

Apart from the positive interaction coefficient, it is instructive to have a brief look on the parameter estimates for the controls and the fixed effect proxies. The coefficient of the indicator for lagged establishments of charter schools is positive and highly significant. The average partial effect indicates that districts which signed a charter in the previous three years are 4.5 percentage points more likely to authorize additional schools. Somewhat surprisingly, I do not find any size effect on the districts' willingness to establish charter schools. Thus, although large districts operate more schools, the hypothesis that they are not more likely to run charter schools than small districts cannot be rejected at any reasonable level of significance. Statistically significant effects are found for the remaining controls. Both the share of Hispanic students and the share of students eligible for free or reduced price lunch negatively affect the adoption probability. This suggests that the more favorable the social conditions under which local public school producers operate, the more likely is the establishment of additional charter schools. This is an interesting result, given that proponents of the charter school movement often claim that charter schools are likely to benefit socially disadvantaged students. Furthermore, 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. A decrease in the percent difference between a district's average test score and the statewide average by one percentage point increases the probability for the establishment of charter schools by 0.008. Thus, more successful schooling producers are less often experimenting with charter schools as a new production technology.[15]

It is also worth noting that three 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. Similarly, districts which have experience in managing magnet schools are significantly more likely to set up charter schools. Finally, districts which did not use any of the included

[15]The reader might worry about the correlation between the share of students receiving free lunch and achievement (controlling for ethnicity). As a robustness check, I estimated the model excluding the share of students receiving free lunch. The coefficient estimates were found virtually unchanged.

innovative policies during the pre-sample period are significantly less likely to innovate. 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.

As an additional check for the performance of the finite spatial lag model, 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 -775.53 compared to -772.47 for the finite spatial lag probit. A likelihood ratio test reveals that this difference is significantly different from zero at the 5 percent level.

Unfortunately, since ϕ is the coefficient of a latent variable, no average partial effect can be computed for the spatial lag, and we cannot say anything about the quantitative effect of neighbors' predispositions on the probability that districts establish additional schools. In contrast to this, it is easy to derive a partial effect from the diffusion parameter in the model with time-lagged interaction.

Table 3 displays the results for two specifications of a probit estimation with time-lagged spatial effects. The partial effects (changes in probabilities) presented in the first column are based on the latent variable model given in Eq. (10) with innovations of neighboring districts entering the estimation equation as specified in Eq. (11).

The partial effect of neighbors' lagged policies on the adoption probability indicates a strong interaction between districts in the choice of charter school policies. A one percentage point increase in the share of neighbors with at least one charter school established in the previous three years brings about a 0.05 percentage point increase in the probability that the district establishes a charter school in the current period. Thus, a district with maximum exposure to innovating neighbors is about 5 percentage points more likely to open a charter school compared to a district with the minimum exposure. This is a substantial diffusion effect, given that the overall innovation rate is, on average, 4.8 percent per year. Note also that the effect of neighbors' lagged policies on a district's adoption probability is only slightly smaller than the effect of own lagged decisions. The remaining average partial effects are well in line with the results displayed in Table 2 and need not be discussed again.

Let us now turn to the results depicted in column (3). They are from a specification with separate explanatory variables derived from adoptions of the composite neighbor in $t-1$, $t-2$, and $t-3$.[16] While their is no statistically significant impact of neighbors' adoptions in $t-1$, the policies of neighboring districts in $t-2$ and $t-3$ do affect the current innovation probability. A one percentage point increase in the measure for neighbors'

[16]Incorporating further lags of neighbors' adoptions does not have any impact on the results, and the corresponding coefficients are not statistically different from zero.

innovation activity in $t-2$ increases the current innovation probability by roughly 0.0004. The impact of adoptions among neighboring districts in $t-3$ is even more pronounced. Compared to a district with the minimum exposure, a district with the maximum exposure to innovating neighbors in $t-3$ is 5.4 percentage points more likely to open a charter school. Although it is tempting to speculate that the absence of any reaction on neighbors' decisions in $t-1$ points to districts learning from *outcomes* of charter school policies adopted by neighboring districts instead of neighbors' decisions, caution is warranted in interpreting the results displayed in column (3). Given that substantial effort is needed to implement charter school policies, it could well be the case that district decision makers quickly respond to policies of neighboring districts, but that it takes one or more school years until a charter school is in operation.

A further point that is worth noting is that the estimates for the interaction effects are not critically affected if the indicator for lagged own adoptions is dropped from the estimation equation. Despite the fact that the presence of time-invariant unobserved effects is taken into account by the inclusion of the fixed effect proxies, serial correlation in the error term cannot be ruled out completely. The potential of such correlation could give rise to concerns about the exogeneity of the indicator for lagged innovation activity. However, the results displayed in columns (2) and (4) of Table 3 show that dropping the potentially endogenous variable from the set of explanatory variables does not have any significant effect neither on the interaction effects nor on the controls. This finding should make us confident that the fixed effect proxies are successfully removing any significant amount of serial correlation from the errors.

5 Conclusion

This paper belongs to a small body of literature dealing with strategic interaction among local governments with non-fiscal variables as instruments. Specifically, the paper provides evidence on the diffusion of charter schools among California school districts in the years 1998-2002. Extending the approach of Case (1992), a discrete choice panel data model for the diffusion of new political technologies is proposed. The model incorporates a spatial lag as well as lagged own adoption decisions and accounts for spatial error correlation. In addition, the pre-sample history of innovative district policies is used to derive proxy variables for district fixed effects. The approach accounts for contemporaneous interaction in the adoption of policy innovations and is thus consistent with yardstick competition among local governments as well as policy competition induced by Tiebout choice. The model with contemporaneous interaction is contrasted to a model where policy diffusion involves a time lag. The lag accounts for the fact that it is time consuming for local governments to implement new policies that are observed elsewhere, and that learning from other jurisdictions' choices is therefore likely to lead to time lags in the diffusion process.

In both models, the presence of positive interaction among school districts in the adoption of charter school policies is supported by the data. Based on the two models, it is thus not possible to discriminate between yardstick competition, policy competition induced by Tiebout choice, and learning as potential driving forces for the diffusion of charter schools. While in the model with contemporaneous interaction it is not possible to derive the quantitative impact of interaction the innovation activity, the estimation results for the model with time-lagged spatial effects suggest a strong effect of neighbors' lagged adoptions on the current innovation probability. Taken together, the evidence suggests that the school districts' behavior when choosing their charter school policies was significantly affected either by interdependencies among the districts' attitudes towards the new policy instrument or by the behavior of neighboring districts in previous periods.

Acknowledgements

I am indebted to Thiess Büttner and Robert Schwager for their advice and would like to thank Richard Arnott and two anonymous referees for their comments and suggestions. All errors are my own.

References

Baicker, K., 2005, The spillover effects of state spending, Journal of Public Economics 89, 529-544.

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

Besley, T. and A. Case, 1995, Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition, Amercian Economic Review 85, 25-45.

Bivand, R. and S. Szymanski, 1997, Spatial dependence through local yardstick competition: Theory and testing, Economics Letters 55, 257-265.

Blundell, R., R. Griffith, and J. Van Reenen, 1995, Dynamic count data models of technological innovation, Economic Journal 105, 333-344.

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.

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, 438-467.

Brueckner, J.K., 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, 203-229.

Büttner, T., 2001, Local business taxation and competition for capital: The choice of the tax rate, Regional Science and Urban Economics 31, 215-245.

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.

Dolowitz, D. and D. Marsh, 1996, Who learns what from whom? A review of the policy transfer literature, Political Studies 44, 343-357.

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, 101-122.

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.

Kotsogiannis, C. and R. Schwager, 2006, On the incentives to experiment in federations, Journal of Urban Economics 60, 484-497.

Montalvo, J.G., 1997, GMM estimation of count-panel-data models with fixed effects and predetermined instruments, Journal of Business and Economic Statistics 15, 82-89.

Rincke, J., 2006a, Competition in the public school sector: Evidence on strategic interaction among US school districts, Journal of Urban Economics 59, 352-369.

Rincke, J., 2006b, Policy innovation in local jurisdictions: Testing for neighborhood influence in school choice policies, Public Choice 129, 189-200.

Stone, D., 1999, Learning lessons and transferring policy across time, space and disciplines, Politics 19, 51-59.

Wooldridge, J.M., 1997, Multiplicative panel data models without the strict exogeneity assumption, Econometric Theory 13, 667-678.

Table 1: Summary statistics

	Mean	S.D.	Min	Max
Indicator for at least one additional charter	0.048	0.214	0.00	1.00
Enrollment[a]	6.16	25.5	0.007	747
% Hispanic students	0.315	0.267	0.00	1.00
% students free lunch	0.431	0.266	0.00	1.00
Student achievement[b]	0.638	2.55	-6.32	9.77
Fixed effect (FE) proxies				
FE 1: Number of charters	0.133	0.639	0.00	10.0
FE 2: Magnet program	0.102	0.303	0.00	1.00
FE 3: Independent study program	0.573	0.495	0.00	1.00
FE 4: Zero innovation activity	0.399	0.490	0.00	1.00

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

[a] scaled by 1/1000.

[b] Deviation of average math test score from statewide average in percent.

Table 2: Finite spatial lag probit

	Coefficient	S.E.	dP/dX
Composite neighbor's predisposition	0.259 **	0.102	-
Charter in previous three years	0.543 ***	0.098	0.045
Enrollment	0.002	0.001	0.000
% Hispanic students	-0.557***	0.167	-0.046
% students free lunch	-0.481**	0.211	-0.040
Student achievement	-0.091***	0.025	-0.008
FE 1	0.144 ***	0.046	0.012
FE 2	0.448 ***	0.095	0.037
FE 3	-0.026	0.167	-0.002
FE 4	-0.338*	0.185	-0.028

Probit on balanced panel of 941 districts (1998 to 2002).
Estimation accounts for spatial error correlation and year effects.
* 10% significance level.
** Idem., 5%.
*** Idem., 1%.

Table 3: Probit with time-lagged spatial effects

	(1)	(2)	(3)	(4)
Composite neighbor's adoptions	0.050 ***	0.053 ***	-	-
in previous three periods	(0.013)	(0.013)		
Composite neighbor's	-	-	-0.010	-0.007
adoptions in $t-1$			(0.021)	(0.021)
Composite neighbor's	-	-	0.038 *	0.039 *
adoptions in $t-2$			(0.021)	(0.021)
Composite neighbor's	-	-	0.054 **	0.055 **
adoptions in $t-3$			(0.021)	(0.022)
Charter in previous three years	0.054 ***	-	0.055 ***	-
	(0.014)		(0.014)	
Enrollment	0.000 **	0.000 **	0.000 **	0.000 **
	(0.000)	(0.000)	(0.000)	(0.000)
% Hispanic students	-0.045 ***	-0.048 ***	-0.046 ***	-0.049 ***
	(0.013)	(0.013)	(0.013)	(0.013)
% students free lunch	-0.029 *	-0.036 **	-0.030 *	-0.037 **
	(0.016)	(0.016)	(0.016)	(0.016)
Student achievement	-0.007 ***	-0.008 ***	-0.007 ***	-0.008 ***
	(0.002)	(0.002)	(0.002)	(0.002)
FE 1	0.010 ***	0.014 ***	0.010 ***	0.014 ***
	(0.003)	(0.004)	(0.003)	(0.003)
FE 2	0.047 ***	0.056 ***	0.047 ***	0.057 ***
	(0.013)	(0.014)	(0.013)	(0.014)
FE 3	0.000	0.001	-0.001	-0.000
	(0.012)	(0.012)	(0.012)	(0.012)
FE 4	-0.020	-0.022	-0.022 *	-0.024 *
	(0.012)	(0.012)	(0.012)	(0.012)

Probit on balanced panel of 941 districts (1998 to 2002).
Estimation accounts for year effects.
Table shows changes in probabilities and corresponding standard errors in parentheses.
* 10% significance level of underlying coefficient.
** Idem., 5%.
*** Idem., 1%.