Yardstick competition and public sector innovation

Abstract

The paper addresses the incentives of the public sector to implement new technologies in public service provision. I focus on the role of local governments under decentralization. Exploiting variation in the level of innovation in a large sample of US school districts, the impact of yardstick competition on the choice of public sector technologies is identified. It is shown that the impact of other districts' innovation activity on a district's innovation score is much stronger in communities where incumbents face a high risk of being elected out of office. This finding suggests that under decentralization, yardstick competition is an important force shaping the decisions of local governments to adopt new technologies.

1 Introduction

Providing public inputs involves choices about production technologies. The present paper deals with one aspect that is particularly important in this respect: the incentives of office-motivated representatives to adopt innovations in the sense that the technology by which public goods and services are produced is significantly altered. In particular, I focus on decentralized systems of governments and ask how the interplay between rent-seeking politicians and residents affects the choice of local governments between traditional and new public sector technologies. I argue that it is useful to focus on the incentives

of governments in a situation with asymmetric information on the quality of public sector production. The reason is twofold: Firstly, it is usually difficult for voters to determine whether public sector technologies actually chosen are optimal or not. Secondly, it is reasonable to assume that switching to a new technology is associated with high effort costs, leading office-motivated representatives to dislike public sector innovation irrespective of potential productivity gains.

The recent literature has pointed to the fact that in decentralized systems of government, the ability of voters to compare performance across jurisdictions is a force that may be used to tame office-motivated representatives. The argument is straightforward: When evaluating their representatives, voters can only imperfectly distinguish between bad performance and bad luck. In correlated environments, relative performance evaluation across jurisdictions will help the voter to improve the inference on the quality of public policies. The paper relates the problem of technological choice in the public sector to the argument on relative performance evaluation. As its main contribution, the paper offers empirical evidence on how public sector innovation is affected by yardstick competition among local jurisdictions.

To link the empirical analysis to the theoretical literature, I set up a simple agency model that adapts Besley and Case (1995) in order to allow for discrete technological choices. The model illustrates how yardstick competition distorts the incentives of local governments towards technologies chosen in reference jurisdictions. The intuition is the following: If state variables that determine which technology is optimal are positively correlated, incumbents anticipate to be elected out of office if reference jurisdictions adopt a new technology and the home jurisdiction does not. The main part of the paper is devoted to test whether this effect is present in local jurisdictions' innovation behavior. More specifically, the analysis deals with the introduction of charter schools as a new type of public schooling in US school districts. An estimation approach allowing for distortion towards choices of neighboring jurisdictions

is developed. The analysis concentrates on discriminating between yardstick competition and other potential sources for interaction in innovation activity such as learning externalities and Tiebout choice. I test for the presence and extent of distortion in public sector innovation in subsamples varying by the degree of perceived political competition. While significant spatial interaction is found in districts where the degree of perceived political competition is high, there are no interaction effects in districts where political competition is low. Among the potential explanations, only the presence of yardstick competition is consistent with these findings.

Apart from Besley and Case (1995), the paper is related to other theoretical contributions on political agency problems such as Coate and Morris (1995), Besley and Smart (2007), and Belleflamme and Hindriks (2005). The contributions of Kotsogiannis and Schwager (2006a, 2006b) focus on the incentives to experiment with new policies and are therefore particularly closely related. Furthermore, there is a large body of literature that finds empirical support for strategic behavior of local governments surveyed by Brueckner (2003) and Revelli (2005). The paper adds to this line of research by widening the somewhat narrow focus on tax instruments present in most contributions.[1]

A weakness of the empirical literature on strategic interaction among local governments is that many studies fail to discriminate between different potential explanations for such interaction. In particular, there are only a few papers that are able to separate yardstick competition from alternative explanations for choice variables of local governments to be interrelated. Exploiting the fact that governors in the US face binding term limits, Besley and Case (1995) show that state taxes in the US are positively correlated only in years when the governor is eligible for reelection. In a related study on Italian cities, Bordignon, Cerniglia, and Revelli (2003) find dependence in business tax rates

---

[1]There are noteworthy exceptions. See Case, Hines, and Rosen (1993), Bivand and Szymanski (1997), Brueckner (1998), Fredriksson and Millimet (2002), Baicker (2005), as well as Büttner (2006).

only in cities where majors run for reelection in uncertain contests.[2] Finally, Revelli (2006) exploits a natural experiment to identify the effects of yardstick competition in social service provision by local authorities in the UK. To the best of the author's knowledge, this paper is the first one to discriminate between yardstick competition and other forms of interaction in the choice of non-fiscal policy instruments.

The remainder of the paper is organized as follows. The following section sets up the theoretical model. The empirical strategy to test for the presence of yardstick competition is outlined in Section 3, while Section 4 presents the estimation results. The paper ends with some concluding remarks.

2 Yardstick competition and public sector innovation: A theoretical model

This section sets up a simple model that provides a theoretical framework for the empirical discussion, drawing from Besley and Case (1995). Instead of a general treatment of public sector innovation, I present a special model that serves to highlight some key aspects of local governments' choices of production technologies. In particular, the model introduces the decision to implement a new technology for public service provision as one affecting the reputation of governments in a framework with asymmetric information on the appropriateness of public policies.

2.1 The model

Consider two jurisdictions $i = 1, 2$. Each jurisdiction is populated by a representative resident and has a government which provides a public service. Let

---

[2]Solé Ollé (2003) and Allers and Elhorst (2005) provide similar evidence for Spain and the Netherlands.

p_i be the technology of service provision and π_i the payoff generated. Governments face the alternative to choose between an old technology o and a new technology n. Assume that a special effort $e > 0$ is required to adopt the new technology. 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 payoff of public service provision depends on technology and a set of random variables, $s \in \{l, h\}$ and $S_i \in \{L, H\}$. Depending on the outcome for the jurisdiction-specific random variable S_i, the payoff from the old technology is either π (if $S_i = L$) or $\pi + \Delta$ (if $S_i = H$). The payoff from the new technology is jurisdiction-specific and is jointly determined by s and S_i. The outcome for s determines, jointly for both jurisdictions, whether the new technology is superior to the old technology. If $s = h$, then the new technology generates higher payoff than the old technology, and payoffs are $\pi + 2\Delta$ (if $S_i = L$) or $\pi + 3\Delta$ (if $S_i = H$). If $s = l$, the new technology generates lower payoff than the old technology, and payoffs are $\pi - 2\Delta$ (if $S_i = L$) or $\pi - \Delta$ (if $S_i = H$).

Together with s and S_i, $i = 1, 2$, for each jurisdiction the type of an incumbent government $I_i \in \{G, B\}$ is drawn, where G stands for 'good' and B for 'bad' (rent-seeking). Let θ be the probability that the old technology 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]

The model has two time periods, where utility derived in the second period is discounted with factor $\delta \in (0, 1)$. The timing of events is as follows. At the beginning of period 1, the random variables are drawn. Incumbents in both jurisdictions then simultaneously choose a pair (p_i, τ_i). 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 once again choose a technology and a transfer to themselves.

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 technology, and that choosing the new technology 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 technology is being implemented. In the first situation, let the new technology also be the superior technology. 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 technology is being implemented despite the fact that the old technology 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 technology is being used, it will always be possible for the voter to learn about the superior technology.

To complete the description of the model, the distribution of information has

---

[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.

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 technology is superior, its own type and the type of the other jurisdiction's incumbent.[5] Furthermore, each incumbent observes the realization of S_i in his own jurisdiction. Thus, when simultaneously choosing technologies and transfers, incumbents in both jurisdictions are perfectly informed about the payoff their own technology will generate. With regard to payoffs in the other jurisdiction, incumbents know which technology is superior and the distribution of payoffs conditional on the incumbent's choice. Residents in both jurisdictions know neither incumbents' types nor the realizations of s and S_i, $i = 1, 2$. What residents observe is the technology in their own jurisdiction and their own utility, i.e. the payoff net of the transfer taken by the incumbent. If we allow for yardstick competition, residents also observe the technology and the resident's utility in the other jurisdiction.

2.2 Equilibrium in 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[6] 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 technology p_i and delivers utility v_i. As usually, along the equilibrium path the

---

[5] 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 distribution of innovations critically depends on this assumption.

[6] 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.

underlying beliefs with regard to the incumbent's type are tied down by the requirement 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 technology has been chosen in i, the voter in i believes that a bad incumbent holds office. 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 technology and the corresponding utility of the voter given a zero transfer.

Note that good incumbents' strategies are of a trivial form: they always choose the superior technology together with a zero transfer. For the sake of brevity, the 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 definition:

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

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

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*

8

strategies together with the specified beliefs constitute a Perfect Bayesian Equilibrium:

Bad incumbents choose the old technology 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 technology is chosen and the voter's utility is either $\pi + 3\Delta$ or $\pi + 2\Delta$, and if the old technology 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.

Proof. See the appendix.

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 technology is costly in terms of additional effort, bad governments prefer the old technology in all cases. For the voter, reelecting governments pretending that the old technology is superior and that payoffs are low is optimal as long as the corresponding probabilities are sufficiently high.

2.3   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, adopting the rule from the single-jurisdiction case, we specify that

9

(i) whenever, for some jurisdiction i, it is apparent that a transfer has been taken and/or it is apparent that the inferior technology 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 suggests 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 technology 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 technology only if the following three conditions hold: The new technology 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 technology 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 technology is chosen and the voter's utility is either $\pi + 3\Delta$ or $\pi + 2\Delta$, and if the old technology 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.

Proof. See the appendix.

The crucial point in Proposition 2 is that a rent-seeking incumbent with a benevolent counterpart chooses the new technology if it is superior and, at the same time, it is possible to extract a positive rent in the first period without being detected. It is comparative performance evaluation that makes this behavior worthwhile: Rent-seeking incumbents anticipate that the other jurisdiction's choice will reveal that the new technology is superior and that running the old technology would therefore result in electoral defeat. The equilibrium is supported in a situation where the (unconditional) probability of a low-productivity shock is sufficiently high. Furthermore, it must be sufficiently likely that the old policy is superior and that the incumbent is good. Finally, the discount factor must put a sufficiently high weight on incumbents' period-2 utility, and the positive correlation between jurisdiction-specific pro-

ductivity shocks must be sufficiently low.[7]

2.4 Model predictions on the distribution of innovation activity

What does the above analysis imply for public sector innovation activity in an economy with many jurisdictions? Since the extent to which comparative performance evaluation can produce valuable information on incumbent performance depends on the degree of correlation between payoffs, locally correlated shocks on payoffs will evoke corresponding patterns in incumbents' choices. More precisely, provided that shocks affecting the payoff of public policies are positively correlated across jurisdictions, it is likely that we observe local governments adopting policies that mirror policies in neighboring jurisdictions. This is because office-motivated governments under yardstick competition benefit, in terms of reputation, from adopting new technologies only if their neighbors do so.

It is worth noting that a similar pattern of spatial dependence in technological choices would also emerge in a situation where it is costly for residents to collect and review information about other jurisdictions. Such information costs will typically rise with distance, making it worthwhile to select nearby jurisdictions as a reference when evaluating the performance of local governments.

The remaining parts of this paper are devoted to an empirical test of the model predictions regarding the spatial distribution of public sector innovation activity.

---

[7]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.

3 Empirical strategy and data

3.1 Background

The yardstick competition model deals with horizontal interaction among governments, and empirical tests of the model predictions should therefore be based on cross-sectional data on innovation activity of local jurisdictions. Moreover, we should focus on policies that are relevant from the point of view of voters. Otherwise, the identification of the effects of interest will be a hopeless endeavor.

The empirical analysis in this study makes use of the fact that the US public school system provides outstanding opportunities for testing hypotheses on local governments' innovation behavior. In particular, the local level with its large number of school districts provides us with an ideal testing ground of local government behavior. The quality of public schooling is important to most households, either because it affects the children's human capital or because of capitalization in local property values.

The innovation under investigation in this study is the operation of charter schools. Charter schools are publicly funded schools, but they operate independently from the existing school district structure. They are exempt from selected state or local rules and regulations and benefit from substantial autonomy with regard to curriculum, teaching methods, hiring decisions, etc. As such, charter schools are discussed as a means of increased parental choice in the US public school system.

The data used in the empirical analysis are from California and cover the period from 1998 to 2002. The California charter school law from 1992 specifies that charter schools are mainly operated under the supervision of the local school districts.[8] A charter school is typically governed by a group or orga-

---

[8]Under certain conditions, also county and state authorities can run charter schools,

13

nization under a contract or charter with the charter authority. Contracts usually have a five-year term. Since state and district money in California basically follows students, the funding of charter schools is similar to those of traditional public schools.[9]

It is 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. In general, district officials may use charter schools to compete for non-resident students.[10] 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.

While it is generally acknowledged that charter schools expand the choice options of parents and pupils within the public school sector, there is dispute over the effects of that type of school choice on student achievement and school productivity.[11] Note that the empirical analysis is only concerned with adoption decisions of school districts and not with the productivity effects of charter schools.

Since later on I make use of variation in political competition across districts,

---

but the vast majority of charter schools has been established by local school districts.

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

[10]Rincke (2006) shows that competition for students was the driving force behind Michigan school districts opting for open enrollment policies after 1996.

[11]See Hoxby (2003) for a general discussion and empirical evidence.

a few words on the political institutions of school districts are warranted. School districts are special-purpose districts which serve to operate the local public primary and secondary schools. The school board is the governing body of the school district. It has the ultimate responsibility for the operation of the local public schools and employs the superintendent who manages the district's day-to-day operations. Depending on district size, boards consist of three, five or seven members.

The vast majority of school board members in California are locally elected. Only in a few cases, the county board of supervisors appoints members of district school boards. School boards are non-partisan. Board members serve for terms of four years. Most elections are held in November, and board members are elected individually. The terms are usually staggered, so in a typical district there will be openings every two years. There are no general term limits.

3.2 Estimation strategy

Innovations are the outcomes of discrete choice decisions. Consequently, the ideal empirical model for investigating innovation activity will usually be a model for discrete responses. Unfortunately, it is particularly complicated to incorporate horizontal interaction among decision making units in such models.[12] In the following, I suggest an estimation strategy that accounts for the restricted choice set of school districts but that avoids the technical problems of modeling discrete outcomes.

Consider the following structural equation for a corner solution model[13] with

---

[12]See Rincke (2007) for a detailed discussion and references.

[13]This model is also known as the standard censored Tobit model, but in the present case it is more appropriate to focus on the presence of corner solutions instead of data censoring.

innovation spillover

$$y_{it} = \max(0, \phi\, y_{-i,t} + \lambda_1 y_{i,t-1} + \cdots + \lambda_s y_{i,t-s} + x_{it}\beta + \eta_t + u_{it}), \tag{1}$$

where the dependent variable y_{it} is a continuous measure of innovation activity. $y_{-i,t}$ represents a linear combination of other districts' contemporaneous innovation scores, x_{it} is a vector of observable characteristics, η_t denotes a period effect and u_{it} is a residual.

In the estimations shown below, I use the share of public schools that operate under a charter as an intuitive and easy-to-construct measure of innovation activity. For brevity, let us call this measure the innovation score. The linear combination of other districts innovation scores can be written as

$$y_{-i,t} = \sum_{j=1}^{N} w_{ij}\, y_{jt}, \tag{2}$$

where the spatial weights w_{ij} determine the composition of $y_{-i,t}$. As usual, we require $w_{ij} = 0$ if $i = j$.

By explicitly modeling corner solutions, the model accounts for the fact the innovation score is bounded below by zero. Note that ignoring this and simply running a linear regression (using the entire sample or the subsample for which $y_{it} > 0$) would generally result in inconsistent parameter estimates. Note also that it is crucial to account for the history of a district's innovation activity because the five-year term of charter contracts induces dependence of innovation scores over time.

The parameters in Equation (1) cannot be estimated by the standard approach for corner solution models because $y_{-i,t}$ is determined simultaneously with y_{it}. Several estimators for corner solution models with endogenous explanatory variables have been suggested. Smith and Blundell (1986) have proposed a two-step estimator which is easy to implement but has the disadvantage that the computation of average partial effects is cumbersome. I therefore rely on the maximum likelihood approach discussed by Wooldridge (2002, p. 532-33) that avoids the two-stage estimation problem.

Consistent estimation of ϕ and β requires a reduced form equation for the endogenous regressor $y_{-i,t}$ in Equation (1). The complete model looks like

$$y_{it} = \max(0, \phi\, y_{-i,t} + \tilde{x}_{it}\tilde{\beta} + u_{it}) \tag{3}$$

$$y_{-i,t} = \tilde{x}_{it}\delta_1 + z_{it}\delta_2 + v_{it}, \tag{4}$$

where for convenience the first equation restates the structural equation of the model in more general notation and where z_{it} is a vector of instruments that are validly excluded from Equation (3). Note that, as in other econometric approaches involving reduced form equations for endogenous regressors, there is nothing necessarily structural about Equation (4). The errors (u_{it}, v_{it}) are assumed to be independently and identically distributed multivariate normal, independent of the instruments, $(\tilde{x}_{it}, z_{it})$. Under this assumption, maximum likelihood estimation can be based on the joint distribution of $(y_{it}, y_{-i,t})$ given $(\tilde{x}_{it}, z_{it})$.

As usual in applications of spatial econometrics where the identification of spatial effects relies on instrumental variables, the instruments are constructed as spatial transformations of exogenous characteristics $x_1, \ldots, x_K$, i.e. the average innovation score of a composite neighbor is instrumented by average exogenous characteristics of that composite neighbor.

Note that if the true model has lagged dependent variables among the regressors, the strict exogeneity assumption for the explanatory variables will no longer hold. This rules out estimation of standard unobserved effects Tobit models. The consistency of the estimates derived from Equations (3) and (4) thus relies on the assumption that, once we instrument for the composite neighbor's innovation score and account for the impact of observable contemporaneous characteristics x_{it} and the lagged innovation scores $y_{i,t-1}, \ldots, y_{i,t-s}$, the residual u_{it} does not contain any component which is correlated with any of the explanatory variables. In addition, it is assumed in the following that u_{it} itself is not spatially auto-correlated.[14] Since unobserved heterogeneity

---

[14]Note that for the standard linear model with a spatial lag and spatial auto-correlation

among districts may be a concern, I include a set of proxy variables for the unobserved propensity of school districts to implement innovative policies. The first proxy is the number of authorized charters in the pre-sample period, 1993-1997. The second is a dummy variable indicating the existence of a magnet program during the pre-sample period[15]. In addition, a dummy variable for an independent study program established during the pre-sample period is included.[16] Finally, to account for the fact that all proxies are bounded below by zero, I include a dummy indicating a zero value in all three proxies. The strategy to utilize the pre-sample history of innovation activity to account for unobserved district characteristics is discussed in more detail in a companion paper (Rincke, 2007).

At this point it is important to recall that it is not the aim of the analysis to provide evidence on spatial dependence in the choice of public sector technologies, but to identify the impact of yardstick competition on technology adoption. Identifying yardstick competition is complicated by the fact that there are at least two alternative explanations for the choice of public sector technologies being distorted towards that in neighboring jurisdictions: learning externalities and Tiebout choice.

Learning externalities may be important in cases where incumbents are not sure which technology is optimal. Observing choices made elsewhere may then drive incumbents to update their beliefs on the quality of technologies, leading to a sort of herd behavior in technological choice.[17] Alternatively, suppose

---

of errors consistent parameter estimates can be derived from IV-estimations even if the spatial correlation of errors is ignored (Kelejian and Prucha (1998)). However, I am not aware of comparable results for a spatial Tobit model.

[15]A magnet program is a program designed to attract students away from their school of residence. Often, districts intend to improve racial balance my introducing magnet programs or magnet schools. They operate on the basis of a particular curriculum and/or a particular instructional mode or structure.

[16]An independent study program means that individualized education plans are designed to meet the need of individual students. As part of this plan, at-risk students may enter into an agreement with the district to complete specific assignments.

[17]There is a broad literature on learning externalities and their effect on decisions under uncertainty. Hirshleifer and Teoh (2003), for instance, review the literature on herd behavior in capital markets, and many of the predictions regarding the interaction of decision makers

that, as assumed in the theory part, local governments dislike implementing new technologies irrespective of their productivity effects. If local governments use the choice of public sector technologies as an instrument to compete for mobile households, spatial dependence in the extent of public sector innovation may be an immediate result. Empirically, it would thus be difficult to distinguish patterns in public sector innovation resulting from learning externalities and/or Tiebout choice from patterns that are due to yardstick competition.

My strategy to identify the effect of yardstick competition on the school districts' technological choice makes use of variation in the perceived degree of political competition. To highlight the underlying idea, suppose that members of the school board in some school districts can be sure to prevail if they run for reelection as incumbent candidates in the next election. Since the reelection probability is independent of relative performance, school boards in these districts do not have any strategic incentive to mimic the behavior of other districts' local governments. School boards will thus choose their preferred technology, and when explaining the districts' innovation activity the activity of other districts should not have any explanatory power once we account for all relevant district characteristics. However, if voters engage in comparative performance evaluation and board members face the risk of being elected out of office, we expect the resulting strategic incentives to bias the choices of school boards towards the technology used in reference districts. More generally, if yardstick competition is present, we should find the extent to which the choice of technology is distorted being positively related to the exogenous risk of incumbent candidates to be elected out of office.[18] In contrast to that, in a situation without yardstick competition among school boards, there is no reason why the degree to which innovation activity is related to that of other districts should be affected by the extent of electoral competition faced

---

can be applied to public sector innovation.

[18]In fact, the relation may also be inversely U-shaped, since with the reelection probability approaching zero, the potential benefit from mimicking the behavior of other governments may decrease. In practice, however, the incumbency advantage is substantial, so we will rarely observe incumbents with very low reelection probabilities.

by incumbent candidates. In order to test for a differential effect of electoral competition on the degree of spatial correlation in innovation scores, a simple measure for electoral competition is derived from outcomes of past school district elections. The sample of districts is then divided into subsamples with low and high levels of electoral competition for incumbent candidates, and the estimates of the interaction effect are compared.

3.3   Empirical specification

Apart from the impact of yardstick competition, the willingness of school boards to operate charter schools can also be expected to depend on observable school district characteristics. Firstly, innovation activity may depend on size. Small districts can be expected to be more flexible in the adoption of new technologies. On the other hand, it is possible that small districts tend to avoid charter schools because, with a small number of public schools, even a single charter school (i.e., a 'marginal' innovation) will result in a substantial part of the student population being directly affected. Another characteristic that may influence the policy towards charter schools is revenues per student. Officials in high-revenue districts might find it easier to please residents by spending money on highly visible inputs such as teachers or physical infrastructure instead of implementing new technologies. On average, low-revenue districts can thus be expected to show a higher propensity to run charter schools. The descriptive literature on the charter school movement has pointed to the fact that most charter schools are established in urban school districts where the share of minority students and students living in low-income households is relatively high on average. Including the share of minority students, the share of students eligible for free or reduced price lunch as a measure for poverty, and a dummy variable for central city districts will account for the potential impact of these district characteristics. Another important control variable is student achievement. Charter schools

and alternative policies aiming at increased parental choice in public education are often seen as a means to improve the efficiency of public schooling and student achievement in particular. The political pressure to increase choice options of parents should therefore be lower in districts with high average test scores. Two additional dummy variables are included to account for the fact that some districts are specialized to serving elementary or high school students. With dummies for these district types included, the reference will be unified school districts serving all grade levels.

A further point that has to be addressed when selecting a specification for the corner solution model is the choice of spatial weights. It is common in the literature to define these weights according to a simple common border or contiguity criterion. I depart from this somewhat crude approach and suggest a more flexible weighting scheme that takes into account geographical conditions as well as the degree of economic interaction among jurisdictions. In particular, the approach is based on commuting flows between school districts. Intuitively, the weight w_{ij} describes the importance of j among all potential neighbors of district i in terms of commuting flows. Formally, weights are determined as

$$w_{ij} = \frac{1}{2}\left(\frac{n_{ij}}{\sum_{k=1} n_{ik}} + \frac{n_{ji}}{\sum_{k=1} n_{ki}}\right), \tag{5}$$

where n_{ij} (n_{ji}) is the number of commuters living in i (j) and working in j (i). Note that $n_{ij} = 0$ for $j = i$. A high weight thus reflects that either j receives a high share of all commuters from i or j is the home district of a substantial share of all commuters working in i, or both.

Since commuting data at district level are unavailable, I use data on commuting between Census designated places with a population of more than 2,500 to construct the weights. The data are obtained from the Bureau of Transportation Statistics and are based on the 2000 Census (CTPP 2000 Part 3 journey-to-work (JTW) tables). The weights utilize the information from more than 7 million individuals commuting between Census designated places

Table 1: School districts' innovation activity, 1998-2002

Period	1998	1999	2000	2001	2002	98-02
% districts running charter schools	0.100	0.132	0.169	0.196	0.220	0.163
% schools operated as charter schools	0.022	0.033	0.042	0.054	0.065	0.043
Nob[a]	976	984	982	981	978	4901

[a] Los Angeles City Unified school district has been excluded.

in California. In order to obtain approximate numbers for district-to-district commuting, Census-designated places are assigned to school districts, and flows to and from each district are computed from the corresponding place-to-place flows.[19] Although I use highly disaggregated commuting data, there are 314 (320) districts with no commuting to (from) other districts.[20] For these districts, I substitute uniform contiguity based weights. This is done by substituting the inverse of the number of contiguous districts, $1/n_i^c$, for the missing ratios in Equation (5) for contiguous districts, and zero otherwise.

3.4 Data

I use data for the years 1998-2002 to estimate the model described above. For the years prior to 1998, suitable data on standardized test scores at district level are unavailable. 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.

Data on California charter schools come from the California Department of Education.[21] Table 1 shows descriptive statistics on the diffusion of charter

---

[19]Census-designated 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.

[20]Note that commuting of individuals residing or working in settlements with less than 2,500 inhabitants is not captured in the Census workflow data.

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

schools for the years 1998-2002. The sample size varies over time due to variation in the total number of school districts. Note that in all descriptive statistics and regression results displayed, Los Angeles City Unified has been excluded 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 2 presents summary statistics for the covariates. Data used to construct the central-city dummy 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, minority students, students eligible for free or reduced price lunch, revenues, and student test scores averages by grade level are provided by the California Department of Education (California Basic Educational Data System (CBEDS), Public School Enrollment and Staffing Data Files, CalWORKS Data Files, files J-200 for financial data, and data files of the Standardized Testing and Reporting (STAR) program). The data from the Standardized Testing and Reporting program are used to derive an average test score variable that measures educational achievement. The measure describes the deviation of student achievement from the statewide average and is constructed as follows. In a first step, for each grade level the percent deviation of a district's average score from the statewide average is computed. In a second step, a district's average deviation is determined as the average deviation over all grade levels, weighted by the number of students tested. I use scores from a math test to derive the achievement variable because they are available for all grade levels. A problem with the STAR data is that there are missing values for a number of small districts. A simple imputation technique is used to solve this problem. For all observations with valid test scores, the score is regressed on a vector 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 59 (1998) to 66 (2001 and 2002). The proxies for unobserved

Table 2: Summary statistics for dependent variable and explanatory variables, 1998-2002

	Mean	S.D.	Min	Max
% charter schools	0.044	0.152	0	1
% charter schools $t-1$	0.035	0.134	0	1
% charter schools $t-2$	0.028	0.118	0	1
Neighbors' % charter schools	0.038	0.044	0	0.386
Enrollment ($\times$ 1,000)	5.76	10.5	0.036	141
Revenues per student ($\times$ 1,000)	6.88	2.31	0.646	85.4
% minority students	0.460	0.278	0	1
% students free lunch	0.428	0.262	0	1
Student achievement[a]	0.584	2.56	-6.31	9.77
Central city	0.131	0.337	0	1
Elementary school district	0.566	0.495	0	1
High school district	0.096	0.295	0	1
Number of charters in pre-sample period	0.131	0.588	0	10
Magnet school in pre-sample period	0.107	0.310	0	1
Independent study program in pre-sample period	0.606	0.488	0	1
Zero innovation score in pre-sample period	0.363	0.481	0	1

[a] Deviation of average math test score from statewide average in percent.
Pre-sample period is 1993-1997. Nob=4,399.

effects are based on CBEDS School Information Form (SIF), Sections G.

Below I make use of measures for the electoral success of incumbent candidates in past school district elections. The data come from the California Elections Data Archive, providing reports on county, city and school district elections. The reports contain information on election outcomes including the number of candidates to vote for, the number of candidates running for office, the incumbency status of candidates, the number of votes for each candidate, and the outcome for each candidate (elected or not elected).[22] Table 3 displays summary statistics for school district elections in California between 1996 and 2001. The first four rows refer to the level of individual candidates, showing that about one third of all candidates are incumbents and that 50 percent of all candidates are elected. Incumbents seem to benefit from a substantial electoral advantage over challengers, resulting in a 77 percent chance of reelection for the average incumbent. Consequently, about one half of all elected candidates are incumbents. The fifth and sixth row report numbers for the

---

[22]Online available at http://www.csus.edu/calst/cal_studies/CEDA.html.

Table 3: School district elections and election outcomes, 1996-2001

	Nob	Mean	S.D.	Min	Max
Individual candidate level (indicators):					
Incumbent candidates	8518	0.326	0.469	0	1
Candidates elected	8518	0.498	0.500	0	1
Elected candidates among incumbents	2781	0.770	0.421	0	1
Incumbents among elected candidates	4238	0.505	0.500	0	1
Variables at electoral race level:					
# of openings	2000	2.12	0.851	1	7
# of candidates	2000	4.26	2.12	1	28
School district level (indicator):					
Districts with incumbent candidate(s)	1699	0.858	0.349	0	1

level of electoral races.[23] The average number of open seats per race is 2.12, with 4.26 candidates on the ballot. Finally, the last row shows that in about 86 percent of the districts holding school district elections there is at least one incumbent seeking reelection.

4 Estimation and results

The empirical evidence on innovation among California school districts is presented in Tables 4 and 5. All results are based on the empirical model presented in Section 3. The estimation technique is one-step conditional maximum likelihood with instrumenting for the endogenous innovation score of neighboring districts.

As a first step in the empirical analysis, I estimated a baseline model. It is meant as a general test for the presence of interaction between school boards. The baseline model gives a first impression on the strength of the neighborhood influence on innovation activity and thereby serves as a point of reference for the following analysis of yardstick competition as a source of distortion in the

---

[23]Some districts are divided into so-called 'trustee areas'. In these districts, each board member represents a 'trustee area' and is elected only by the residents of this area. Hence, there are more electoral races than district-year observations.

districts' choices. The results are shown in the first two columns of Table 4. The first column has the estimated coefficients and corresponding standard errors. Note that the Tobit model is nonlinear and that, therefore, there is not much to learn from looking at coefficient estimates apart from their sign and the level of significance. What we are interested in are the partial effects of the explanatory variables y_{-i} and $\tilde{x}_k$ $(k = 1, \ldots, K)$ on the conditional expected value of y,

$$\frac{\partial E(y_{it}|y_{-i,t}, \tilde{x}_{it}; y_{it} > 0)}{\partial y_{-i,t}} \qquad \text{and} \qquad \frac{\partial E(y_{it}|y_{-i,t}, \tilde{x}_{it}; y_{it} > 0)}{\partial \tilde{x}_{itk}}. \tag{6}$$

For continuous variables, the marginal effects are evaluated at sample means. If $\tilde{x}_k$ is a dummy variable, the difference in the expected value of y with $\tilde{x}_k = 1$ and $\tilde{x}_k = 0$ (holding all other elements of $\tilde{x}$ fixed) is given. The partial effects and corresponding standard errors for the baseline model are displayed in the second column of Table 4. Most importantly, the baseline model confirms the presence of positive interaction in innovation activity among school districts. The partial effect of neighbors' innovation score is 0.3, indicating that, conditional on a strictly positive innovation score, a one percentage point increase in the share of charter schools in neighboring districts is associated with a 0.3 percentage point increase in a district's own expected share of charter schools. This finding is in line with the results presented in a previous study Rincke (2007) and means that the choice of a technology is strongly affected by corresponding choices in neighboring districts. The results also confirm that it is important to account for inertia in the process of technological choice. The one-period lag of the innovation score is highly significant, with a marginal effect of 0.25. Note that the estimate of the marginal effect for the two-period lag is not statistically different from zero. Hence, the effect of past innovation activity on the present innovation score is completely captured in the one-period lag. Furthermore, innovation activity is somewhat higher in larger districts. While revenues per student do not seem to have any impact, the partial effects of the share of minority students and poverty both are negative. I also find some evidence suggesting that charter schools are more likely to

26

Table 4: Spatial dependence in school districts' innovation activity (IV-Tobit)

Dependent variable: % charter schools

Sample restriction imposed[a]	No		Yes	
	Coefficient	$\frac{\partial E(y|x,y>0)}{\partial x_j}$	Coefficient	$\frac{\partial E(y|x,y>0)}{\partial x_j}$
Neighbors' % charter schools	1.59*	0.297*	1.68**	0.334**
	(0.861)	(0.160)	(0.821)	(0.161)
Own % charter schools in $t-1$	1.33***	0.249***	1.19***	0.237***
	(0.066)	(0.018)	(0.103)	(0.022)
Own % charter schools in $t-2$	0.071	0.013	0.094	0.018
	(0.087)	(0.016)	(0.102)	(0.020)
Enrollment	0.002***	0.000***	0.007***	0.001***
	(0.000)	(0.000)	(0.002)	(0.000)
Revenues per student	0.004	0.000	0.002	0.000
	(0.004)	(0.000)	(0.002)	(0.000)
% minority students	-0.083*	-0.015*	-0.117*	-0.023*
	(0.048)	(0.008)	(0.060)	(0.011)
% students free lunch	-0.133**	-0.024**	-0.156***	-0.031***
	(0.055)	(0.010)	(0.060)	(0.011)
Student achievement	-0.018***	-0.003***	-0.030***	-0.006***
	(0.005)	(0.001)	(0.007)	(0.001)
Central city	0.042**	0.008**	0.022	0.004
	(0.020)	(0.004)	(0.022)	(0.004)
Elementary school district	-0.032	-0.006	0.013	0.002
	(0.020)	(0.003)	(0.019)	(0.003)
High school district	-0.041	-0.007*	-0.039	-0.007
	(0.025)	(0.004)	(0.025)	(0.004)
FE proxy 1	0.016	0.003	.024	0.004
	(0.017)	(0.003)	(0.015)	(0.003)
FE proxy 2	0.062**	0.012**	.046*	0.009
	(0.024)	(0.005)	(0.027)	(0.006)
FE proxy 3	-0.026	-0.004	-.027	-0.005
	(0.043)	(0.008)	(0.039)	(0.008)
FE proxy 4	-0.089**	-0.016**	-.056	-0.010
	(0.044)	(0.007)	(0.043)	(0.008)
Wald test of exogeneity, $\chi^2(1)$	2.51 [0.113]		3.03 [0.081]	
# of censored (uncensored) obs.	3648 (751)		2474 (518)	

Huber-White Standard errors (allowing for clustering by school district) in parentheses. Estimations include a full series of period effects.
[a] Sample restricted to districts with less than 15,000 students and at least two schools.
* sign. at 10% level; ** sign. at 5% level; *** sign. at 1% level.

be established in central urban communities. The results also confirm the expectation that innovation scores are lower in districts with high student achievement, although the quantitative effect is rather small. Finally, the table also reports the coefficients of the proxies for the unobserved propensity to implement innovative schooling technologies. Two out of four proxy variables

are significant at the 5 percent level. Moreover, when excluding the proxies, the partial effects of the interaction term and the lagged dependent variables tend to be significantly larger (results not shown), suggesting that the proxies account for some variation in actual innovation activity that would otherwise (falsely) be attributed to structural explanatory variables.

Some of the results presented below are based on a restricted sample of school districts with less than 15,000 students and at least two schools. To provide a valid point of reference, Table 4 displays the results of the baseline model for both the unrestricted and the restricted sample. A short inspection reveals that excluding very large and very small districts from the sample does nothing to the main results. In particular, the estimate of the interaction effect is virtually unaffected by the sample reduction, both in terms of the size of the marginal effect and statistical significance.

As we have seen, the innovation activity of neighboring districts is endogenously determined by the model setup, and the estimations account for this by instrumental variables. A test of exogeneity of neighboring districts innovation scores is displayed at the bottom of Table 4. The test is easily constructed from the output of the conditional maximum likelihood procedure used to estimate the coefficients and partial effects (see Wooldridge 2002, p. 533 for details). The Test statistics (distributed $\chi^2(1)$) and p-values show that we are well advised to take the endogeneity problem seriously: the null hypothesis of exogeneity is rejected at the 12 percent level of significance in the first and at the 9 percent level of significance in the second specification.

We now turn to tests for the presence of yardstick competition as a potential driving force behind the bias towards technologies of neighbors. As discussed in Section 3, the idea behind the test is to exploit the differential effect of the degree of electoral competition on incumbents' incentives to engage in yardstick competition. I use the share of incumbent candidates who failed to get reelected in the last election taking place in the school district as a mea-

sure for the degree of electoral competition for incumbent candidates. The rationale for using past election outcomes to approximate the current degree of political competition for incumbents is the following. Incumbents seeking reelection will adjust their behavior to the expected degree of electoral competition in coming elections. Unfortunately, there is no perfect way to measure these expectations. However, with school district elections in a typical district taking place every two years, it seems reasonable to assume that incumbents' perceptions of the closeness of the next electoral race are strongly influenced by the degree of electoral success of incumbent candidates in the last district election.

Table 5 presents evidence on the effect of perceived political competition. It displays estimation results based on samples with and without the restriction on school district size being imposed. The reason for excluding very large and very small districts is that the political process determining innovation activity in these districts is potentially quite different from that of more typical districts. School districts with more than 15,000 students are usually unified school districts serving students in large urban areas such as, for instance, Los Angeles, San Francisco, and San Diego. Given the definition of spatial weights, the model links the choice of school boards in these large urban districts to choices of school boards in a large number of suburban school districts typically surrounding the central urban area. Given the distinct structural differences among large urban and suburban school districts, one may question whether this assignment of neighbors for large urban districts is appropriate. Regarding very small districts, the agency problem that underlies the potential presence of yardstick competition may simply not be relevant. With only a few students being instructed at a single schoolsite, the monitoring of school district officials by residents will be close to perfect, giving school board members little incentives to distort the technology of schooling production. Including very large and very small districts may thus obscure important forces affecting the school districts' choices.

29

Table 5: The effect of perceived political competition (IV-Tobit)

Dependent variable: % charter schools. Columns show marginal effects, $\frac{\partial E(y\mid x, y>0)}{\partial x_j}$.				
% incumbents not reelected in last election	$\geq 50\%$	$< 50\%$	$\geq 50\%$	$< 50\%$
Sample restriction imposed^a	No	No	Yes	Yes
Neighbors' % charter schools	0.736*	0.118	0.747**	0.258
	(0.392)	(0.181)	(0.306)	(0.189)
Wald test of exogeneity, $\chi^2(1)$	2.15 [0.142]	0.41 [0.523]	3.82 [0.050]	1.61 [0.204]
# of censored (uncensored) obs.	591 (160)	1688 (388)	450 (109)	1227 (266)

Huber-White Standard errors (allowing for clustering by school district) in parentheses. Additional controls (see Table 4) and period effects included. Districts without incumbent candidate running in last election excluded.
^{a} Sample restricted to districts with less than 15,000 students and at least two schools.
* sign. at 10% level; ** sign. at 5% level.

For brevity, in Table 5 the marginal effects of the control variables are omitted. The threshold for the share of incumbents who failed in the last election in defining subsamples with low and high perceived political competition is one half.[24] Without imposing a restriction on district size, this procedure assigns 751 observations to the subsample with high perceived electoral competition (leftmost column).[25] The partial effect of neighbors' innovation score is significant at the 10 percent level and much higher than in the corresponding baseline regression. In contrast to that, the interaction effect is not significantly different from zero in the sample of districts with less than 50 percent of incumbent candidates losing in the last school district election. After imposing the restriction on district size, the strong dependence of innovation scores on neighbors' activity in districts with high political competition is even more clearly visible. While the size of the estimated marginal effect is again robust to the sample reduction, it is now significant at the 5 percent level. The marginal effect indicates that, conditional on a strictly positive innovation score, an increase in the share of schools operated as charter schools among neighbors by one percentage point will cause a 0.75 percentage point increase in a district's own innovation score. Note that this is more than twice as

[24]Using different threshold levels gives broadly similar results. Even when assigning all districts with a strictly positive share of failing incumbents to the subsample with high political competition, the main results are unchanged.

[25]Note that districts without incumbent candidate running in the last school district election have been excluded.

much as the interaction effect of 0.33 estimated in the baseline model. More importantly, there is no significant impact of innovation activity in neighboring communities on the schooling technology in districts with low perceived electoral competition.

Table 5 points to differential effects of the perceived degree of political competition on the extent to which the schooling technology is distorted towards that of neighbors. Among the potential explanations for interaction effects in public sector innovation, only the presence of yardstick competition is consistent with this finding. Neither competition for mobile households nor learning externalities are able to explain the significant differences in estimated interaction effects.

5 Conclusion

The literature has to date paid little attention to the fact that the provision of public goods and services involves choices about production technologies. In particular, our knowledge on how decisions on the adoption of new technologies for public sector production are made is very limited. This paper sheds light on the choice of local governments between established and new technologies. Assuming that information on technologies is asymmetrically distributed between voters and their elected representatives and that representatives prefer business-as-usual technologies to innovation, the analysis deals with the question whether the ability of voters to compare government performance across jurisdictions affects innovation in a decentralized public sector. The empirical example is the operation of charter schools in California school districts. With almost 1,000 school districts observed over several years, the study exploits a data set that offers outstanding opportunities to provide insight into the political economy of public sector innovation. Based on the exceptional rich information on California's public school system, the

paper identifies the impact of yardstick competition on technology adoption in the public sector.

In a baseline specification of a spatial corner solution model accounting for the endogeneity of neighbors' technologies, strong positive spatial correlation in innovation scores is found. In order to test to what extent the positive interdependence of innovation activity is driven by yardstick competition, I use data on school district election outcomes to derive a measure for the perceived degree of political competition faced by members of local school boards. The estimation results suggest that in districts with high perceived political competition, innovation scores strongly depend on innovation activity in neighboring districts. In contrast to this, I do not find such dependence in school districts with low political competition. Among the potential explanations for public sector technologies to be interdependent, only yardstick competition is consistent with this finding.

The contribution of the paper to the literature is twofold. Firstly, the analysis suggests that public sector innovation is affected by forces that have to date been largely ignored. Secondly, the interaction between decision makers and voters should be taken into account when thinking about the incentives of agents in the public sector to implement best-practice technologies.

Appendix: 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 technology 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:

For convenience, let us first restate the strategies.

Bad incumbents set

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

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

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

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

and the representative voters sets

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

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

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

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

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 technology 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 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 technology 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 technology, yardstick competition does not provide the voter with additional information in this case compared to the situation with only a single jurisdiction, and technology 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 no longer optimal for a bad incumbent in i to choose (o, Δ) if $s = h$ and $S_i = H$ since the superior new technology 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.

Acknowledgements

References

ALLERS, M. A., AND J. P. ELHORST (2005). Tax mimicking and yardstick competition among local governments in the Netherlands. Internatonal Tax and Public Finance, 12, 493–513.

BAICKER, K. (2005). The spillover effects of state spending. Journal of Public Economics, 89, 529–544.

BELLEFLAMME, P., AND J. HINDRIKS (2005). Yardstick competition and political agency problems. Social Choice and Welfare, 24, 155–169.

BESLEY, T., AND A. CASE (1995). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. Amercian Economic Review, 85(1), 25–45.

BESLEY, T., AND M. SMART (2007). Fiscal restraints and voter welfare? Journal of Public Economics, 91(3-4), 755–773.

BIVAND, R., AND S. SZYMANSKI (1997). Spatial dependence through local yardstick competition: Theory and testing. Economics Letters, 55, 257–265.

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.

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.

BÜTTNER, T. (2006). The incentive effect of fiscal equalization transfers on tax policy. Journal of Public Economics, 90, 477–497.

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.

COATE, S., AND S. MORRIS (1995). On the form of transfers to special interests. Journal of Political Economy, 105, 1210–1235.

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.

HIRSHLEIFER, D., AND S. H. TEOH (2003). Herd behavior and cascading in capital markets: A review and synthesis. European Financial Management, 9(1), 25–66.

HOXBY, C. M. (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 C. M. Hoxby (ed.), The economics of school choice (pp. 287–341), Chicago: University of Chicago Press.

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, 60, 484–497.

————— (2006b). Political uncertainty and policy innovation. Journal of Public Economic Theory, 8, 779–805.

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, 90, 459–475.

RINCKE, J. (2006). Competition in the public school sector: Evidence on strategic interaction among US school districts. Journal of Urban Economics, 59(3), 352–369.

——— (2007). Policy diffusion in space and time: The case of charter schools in California school districts. Regional Science and Urban Economics, 37, 526–541.

SMITH, R. J., AND R. W. BLUNDELL (1986). An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. Econometrica, 54, 679–685.

SOLÉ OLLÉ, A. (2003). Electoral accountability and tax mimicking: The effects of electoral margins, coalition government, and ideology. European Journal of Political Economy, 19, 685–713.

WOOLDRIDGE, J. M. (2002). Econometric analysis of cross section and panel data. Cambridge (Mass.): The MIT Press.