Econ Gov (2010) 11:269–294
DOI 10.1007/s10101-009-0067-x

ORIGINAL PAPER

Corruption, decentralization and yardstick competition

Oguzhan C. Dincer · Christopher J. Ellis · Glen R. Waddell

Received: 15 December 2008 / Accepted: 28 August 2009 / Published online: 22 September 2009

Abstract Several empirical studies have found a negative relationship between corruption and the decentralization of the powers to tax and spend. In this paper we explain this phenomenon using a model of Yardstick Competition. Using data on federal corruption-related convictions in U.S. states, we also provide new evidence that points to the existence of a spatial autoregressive component to explaining corruption. We interpret this as consistent with the theoretical findings.

Keywords Corruption · Decentralization · Yardstick competition · Spatial autoregression

JEL Classification H20 · H29

1 Introduction

Recent research by several contributors points to a significant negative relationship between the degree to which the powers to tax and spend are decentralized in an economy and the overall level of governmental corruption (e.g., Treisman 2000; de Mello 2004; de Mello and Barenstein 2001; Fisman and Gatti 2002a; Arikan 2004).[1] On the surface this is quite surprising, as it might be anticipated that local politicians or bureaucrats would possess detailed knowledge of any opportunities for corruption that

[1] One recent paper that bucks this trend is Fisman and Gatti (2002b) who finds that expenditure decentralization does not reduce the level of corruption unless it is accompanied by revenue decentralization.

O. C. Dincer
Department of Economics, Illinois State University, Normal, IL 61790-4200, USA

C. J. Ellis (✉) · G. R. Waddell
Department of Economics, University of Oregon, Eugene, OR 97403-1285, USA
e-mail: cjellis@uoregon.edu

might arise in their jurisdictions. They might thus be expected to extract any available corruption rents more efficiently than their less-informed national counterparts.

In what follows, we both propose a new theoretical explanation for a negative correlation between empirical measures of corruption and decentralization. We also establish new evidence of this empirical regularity. Of course, the poor performance of a government in one jurisdiction might be attributed to a number of factors including corruption. However, we argue that because there are more independent taxation and expenditure decisions made in a decentralized economy, there are more opportunities for local populations, somewhat of a regulating force, to make cross-jurisdictional comparisons of politician or bureaucrat performance. An increase in the number of comparative observations made by the populations of jurisdictions has two effects that are relevant to the relationship between corruption and decentralization. First, as the number of comparative observations made by a jurisdiction's population increases the inferences they make about the causes of a particular observed outcome become increasingly precise. Second, as the number of observations increases the likelihood that particular inferences will be arrived at also changes. As we shall subsequently see, this latter effect proves to be crucial in generating a negative theoretical relationship between corruption and decentralization. Our explanation of the corruption-decentralization relationship depends crucially on there being imperfect information about a politician–bureaucrats type, this contrasts with the other stories found in the literature where Bertrand competition for mobile factors between jurisdictions involves reductions in the corruption rents accruing to agents setting regulatory barriers to economic activity. See Bardhan and Mookherjee for a nice survey of this literature.

The idea that the performance of local jurisdictional governments can be evaluated by cross jurisdictional comparisons was first proposed by Salmon (1987)[2] and popularized by Besley and Case (1995).[3] They compared the behaviors of state governors and found that if a particular policy was adopted by one governor and deemed successful, it was quickly copied by neighbors. Besley and Case concluded that local populations were making comparative evaluations of jurisdictional governments, who thus responded by engaging in what has come to be known as "yardstick competition." In our analysis, we modify the yardstick competition approach in such a way as to explain the empirical regularity we observe between decentralization and corruption. The attractiveness of the Besley-Case explanation for governmental behavior is that it is quite parsimonious with information; very little detailed knowledge is required to make evaluations. The population compares outcomes across jurisdictions, and draws statistical inferences about the politicians' or bureaucrats' level of competence or objectives.

There is considerable empirical support in recent work for the notion of yardstick competition between jurisdictions (e.g., Netherlands: Allers and Elhorst 2005, Spain: Sole Ole 2003, England: Revelli 2002) which we view as significant reason

[2] Salmon's key insight was that much of the thinking behind the theory of labor tournaments of Lazear and Rosen (1981) could be applied to intergovernmental competition. A nice discussion may be found in Breton et al. (2007).

[3] Good discussions of the foundation of this approach and a survey of the litrature can be found in Besley (2006).

to capture such responsiveness in any empirical model of corruption. In our own empirical analysis of corruption in U.S. states, we will fully embrace this notion of Besley and Case (1995) with the adoption of a spatial autoregressive model of corruption. Doing so, we will accomplish two things most directly. First, in estimating the effect of decentralization on corruption, we will hold constant the potential for one jurisdiction's corruption activity to directly influence that of another, potentially increasing the precision of estimated coefficients relating to decentralization. Second, we will identify if this endogenous responsiveness—yardstick competition—appears in the data, and if so, to what degree. In what follows, we provide new empirical evidence that is consistent with regularities suggested by the theory—reaffirming the negative relationship between decentralization and corruption found elsewhere, and documenting positive and significant neighborhood effects in corruption levels. In the end, regularities within the data suggest that the yardstick competition explanation cannot be easily dismissed.

In the sections that follow we first develop the fundamental relationship between corruption and decentralization assuming constant marginal costs of service provision. In Sect. 4 we reinterpret the analysis to show that it may have applicability to other forms of corruption, such as the extraction of bribes by politician–bureaucrats in return for the circumvention of regulations. In Sect. 5 we explore how proximity (in an informational sense) may lead to a greater likelihood of similar corruption experiences in adjoining jurisdictions. In Sect. 6 we provide further empirical support for the corruption-decentralization relationship, which will also serve to demonstrate the importance of controlling appropriately for potential relationships across geographic space. Finally in Sect. 7 we give a conclusion and further discussion.

2 Model

2.1 The story of events

To make the sequence of events and main arguments of our model clear we begin with a simple informal story temporarily neglecting all technical details. In so doing, consider an economy that lasts for two periods and is divided into a number of fiscally distinct jurisdictions. In the beginning of the first period each jurisdiction selects a single politician–bureaucrat and empowers her to raise taxes to fund the supply of local public services. These services might include the provision of roads, sewers, garbage collection, parks, snow removal and the like. Some politician–bureaucrats are honest social welfare maximizers, and some are dishonest individual selfish maximizers. Each politician–bureaucrat's type is private information known only to themselves. The first act of each politican-bureaucrat (at the beginning of the first period) is to levy taxes from the population of the jurisdiction. Honest types then apply these taxes to the provision of public services. Dishonest types then decide whether to pool with honest types and supply public services, or separate from them and steal all tax revenues for their own personal consumption. If taxes are applied to the production of public services, then the volume of services supplied in each jurisdiction depends on a stochastic production technology that is effected by both jurisdiction specific and trans-jurisdictional shocks. At the end of the first period the populations in each jurisdiction learn of the levels of public services supplied in all jurisdictions, but do not observe the independent components from which they are constituted.

At the beginning of the second period, based on their observations in period one, each member of each jurisdiction solves an inference problem that determines the probability that the politician–bureaucrat that runs their own jurisdiction is honest. Next, the members of each jurisdiction decide whether to continue with the current politician–bureaucrat or replace her with a new draw from the politician–bureaucrat population. For example, a politician may be voted out of office, or a public scandal may lead either to the resignation or dismissal of a bureaucrat. The sequence of events described for period one then repeats itself in the second period.

2.2 Formal structure of the economy

The objective here is to develop the inference problem on which our arguments are based, and establish the basic corruption-decentralization relationship. As such, we begin with a very simple two period model with constant marginal costs of public service provision. The structure here is similar to that of Smart and Besley (2007) with the key innovation that public service provision is decentralized to local jurisdictions that are subject to idiosyncratic technology shocks.

The economy is assumed to be divided into $i = 1, \ldots, n$ jurisdictions each with a population of size $\frac{m}{n}$. In each jurisdiction the local government levies a per person tax of $\tau_i(t)$ which it may use to supply a service $z_i(t) \in \{0, 1\}$ to every individual in the population, we may think of $\tau_i(t)$ as the cost of purchasing the inputs required to supply this service[4] The per person government budget constraint involves $\tau_i(t) = 1 \geq z_i(t)$ with an equality if the service supply is non-zero. The production of the service is given by the stochastic production technology

$$z_i(t) = f(\mu(t), \quad \varepsilon_i(t)) = \mu(t)\varepsilon_i(t), \tag{1}$$

where $\varepsilon_i(t)$ and $\mu(t)$ are jurisdiction specific and economy wide productivity shocks respectively.

We assume that

$$\varepsilon_i(t) = \begin{cases} 0 \text{ with probability } p \\ 1 \text{ with probability } 1 - p, \end{cases} \tag{2}$$

and

$$\mu(t) = \begin{cases} 0 \text{ with probability } q \\ 1 \text{ with probability } 1 - q. \end{cases} \tag{3}$$

[4] Public service provision is either zero or one, for example the snow in a jurisdiction is cleared or it is not. This ties down the level of taxes that may be levied and the maximal amount that may be stolen from the public purse. Allowing choice over the level of taxes levied and service provided adds little to our arguement except mathematical complexity.

This particular stochastic structure is chosen to bring the results into sharp relief. All that follows can also be obtained using continuous density functions for the random variables, however this adds little to the economics of our analysis and has been suppressed.

Each jurisdiction is run by a politician–bureaucrat drawn from a local population in which the proportions λ_i are honest and $1 - \lambda_i$ dishonest. We assume that the jurisdictions are ordered such that $\lambda_i \geq \lambda_j$ iff $i > j$. Each politician–bureaucrat lives for two periods and serves in each period at the discretion of the local population. Each politician–bureaucrat is paid a wage $w(t)$ for each period they retain office, and can, if they choose, supplement this wage by stealing tax revenues $r_i(t) \leq \left(\frac{m}{n}\right) \tau_i(t)$. Honest politician–bureaucrats never steal. Dishonest politician–bureaucrats do so if this raises their expected income. The expected utility of a politician–bureaucrat in jurisdiction i is assumed to be given by

$$V_i(t) = w_i(t) + r_i(t) + \delta\psi(t + 1 \mid a)[w_i(t + 1) + r_i(t + 1)], \tag{4}$$

where $0 < \delta < 1$ is a discount rate and $\psi(t + 1 \mid a)$is the probability that the politician–bureaucrat will be in office in the second period conditional on the action a taken in the first, and will be made precise in the sections that follow. a is an indicator variable taking the value 0 if the politician–bureaucrat does not steal tax revenues and 1 if they do. It follows that the tax revenues will be stolen if

$$r_i(t)+\delta\psi(t + 1 \mid 1)[w_i(t + 1)+r_i(t + 1)] \geq \delta\psi(t + 1 \mid 0)[w_i(t + 1) + r_i(t + 1)] \tag{5}$$

If this condition holds then honest politician–bureaucrats do not steal in period 1 but dishonest ones do. Expression (5) then determines if the equilibrium of the model involves pooling or separating between the types. Initially, for purposes of expositional clarity, we assume $w_i(t) = w_i(t + 1) = 0$. We shall relax this assumption in due time. Here, as inspection of (5) reveals, it serves to rule out the possibility of pooling and allows us to develop the fundamental inference problem that drives all the results. Later when we reintroduce positive payments to politician–bureaucrats it will be shown that the same inference problem remains key to explaining the relationship between decentralization and corruption even when some dishonest politicians engage in pooling behavior.

Given that $w_i(t) = w_i(t + 1) = 0$ ensures separating behavior, it will soon be obvious that if a politician–bureaucrat chooses to steal they will set $r_i(t) = \left(\frac{m}{n}\right) \tau_i(t)$ and $r_i(t+1) = \left(\frac{m}{n}\right) \tau_i(t+1)$. That is, they will steal all tax revenues available to them, which represent the difference between the costs of supplying the two levels of public service provision. That they steal all tax revenues is thus an artifact of the assumption that the lower level of public service provision is zero, and has no implications for our analysis other than to simplify the exposition.

The population of each jurisdiction has just one decision to make —whether to retain a politician–bureaucrat to run the jurisdiction in the second period of the politician–bureaucrat's life, or, replace her with a random draw from the population.

We assume the populations to be homogeneous with utility functions

$$U_i = x_i(t) + z_i(t) - \tau_i(t) - \eta_i(t) + \delta E[x_i(t+1) + z_i(t+1) - \tau_i(t+1)], \quad (6)$$

where $x_i(t)$, $x_i(t+1)$ are the populations' incomes, and $\eta_i(t)$ is the cost of replacing a politician–bureaucrat if this action is taken. For algebraic simplicity, we henceforth assume $x_i(t) = x_i(t+1) = 0$, which yields

$$U_i = z_i(t) - \tau_i(t) - \eta_i(t) + \delta E[z_i(t+1) - \tau_i(t+1)]. \quad (7)$$

We shall also assume that the public cannot directly observe the random variables $\mu(t)$ and $\varepsilon_i(t)$. All that the pubic may observe is whether the service is supplied or not. Therefore, they receive the signals

$$s_i(t) = \begin{cases} z_i(t) \in \{0, 1\} \text{ iff } r_i(t) = 0 \\ 0 \quad\quad\quad\quad\quad \text{otherwise.} \end{cases} \forall i \quad (8)$$

Notice that the population in jurisdiction i observes the supply of services in all other jurisdictions $s_j(t)\forall j$, and hence may make inferences based on yardstick competition. Should the population in a particular jurisdiction decide that it is sufficiently likely that their local government is corrupt they will decide to replace them with a new politician–bureaucrat, randomly drawn from the population. We address this next.

2.3 Replacing politician–bureaucrats

Variations in the level of corruption depend on the frequency with which tax/expenditure decisions are made by corrupt agents. This in turn depends on the frequency with which corrupt activities are detected and the perpetrators replaced. In our theoretical analysis, we do not make the form of replacement specific—politician–bureaucrats may be voted out of office, fired, or arrested and incarcerated. It follows that the cost of replacing a politician $\eta_i(t)$ is subject to several possible interpretations. If the politician is replaced following a corruption inquiry and subsequent legal action it may represent resources used in this process and also the possible disruption to services. If the politician is voted out of office it may represent the loss of "pork" that a politician with seniority can realize for their constituents but a new replacement cannot immediately obtain.

In our theory, the politics of the economy are assumed to be quite simple. Each jurisdiction selects a politician–bureaucrat via random draw from the pool of potential politician–bureaucrats at the start of period 1. After observing $s_j(t)\forall j$, the populations in each jurisdiction decide whether to retain the incumbent or take a new draw. All dishonest politician–bureaucrats will steal in the second period. All honest politician–bureaucrats will not. A new draw will be made if the expected value of doing so exceeds the expected value of retention. That is, if

$$\lambda_i(1-p)(1-q) - \eta \geq \overline{\lambda}_i(t+1)(1-p)(1-q) \quad (9)$$

or

$$\lambda_i - \frac{\eta}{(1-p)(1-q)} \geq \overline{\lambda}_i(t+1), \tag{10}$$

where $\overline{\lambda}_i(t+1)$ are the posterior beliefs about the type of the politician–bureaucrat in jurisdiction i following the observations $s_j(t)\forall j$. Clearly then, the decision to retain or replace a politician–bureaucrat depends on the updating of beliefs about the politician–bureaucrats' type, which takes place at the beginning of the second period.

2.4 Updating of beliefs

We assume the public are Bayesians who update beliefs about a politician–bureaucrats type according to Bayes Rule. Hence

$$\begin{aligned}\overline{\lambda}_i(t+1) &= \Pr\left[i^h \mid s_1(t), \ldots, s_n(t)\right] \\ &= \frac{\Pr\left[s_1(t), \ldots, s_n(t) \mid i^h\right]\Pr\left[i^h\right]}{\Pr\left[s_1(t), \ldots, s_n(t)\right]} \\ &= \left\{\frac{\Pr\left[s_1(t), \ldots, s_n(t) \mid i^h\right]}{\Pr\left[s_1(t), \ldots, s_n(t)\right]}\right\}\lambda_i, \end{aligned} \tag{11}$$

where we adopt the notation $\Pr\left[i^h\right]$ for the probability that i is honest and $\Pr\left[i^d\right]$ for the probability that i is dishonest. We thus need to compute the appropriate probabilities. Three information states can arise for any jurisdiction i. Either positive production of the service is observed in i, positive production of the service is not observed in i but is observed in another jurisdiction, or, positive production of the service is not observed in any jurisdiction. We term these information states $\lambda_i(t)$ -revealing, $\mu(t)$-revealing and non-revealing respectively.[5]

2.4.1 The $\lambda_i(t)$-revealing state

In the $\lambda_i(t)$-revealing state for jurisdiction i positive service production is observed, such that $s_i(t) = 1$. The population immediately deduce that $\mu(t) = \varepsilon_i(t) = 1$ and that the incumbent politician–bureaucrat is honest. Hence, updating yields

$$\overline{\lambda}_i(t+1 \mid s_i(t)=1) = \Pr\left[i^h \mid s_i(t)=1\right] = 1. \tag{12}$$

[5] The modeling strategy of choosing a multilicative stochastic production technology comprising of $\{0, 1\}$ shocks was precisely to give this simple three state information structure. This specification can be generalized without qualitatively effecting the conslusions that follow, but the cost in terms of algebra is quite large.

2.4.2 The $\mu(t)$-revealing state

Here the observation is made that the service is not produced in jurisdiction i, $s_i(t) = 0$. However, elsewhere production is positive, such that $s_j(t) = 1$ for some $j \neq i$. Hence, the population of jurisdiction i will immediately be able to deduce that there has been a positive economy wide shock, $\mu(t) = 1$. They will update their beliefs concerning the honesty of their own politician–bureaucrat according to (see Appendix for details)

$$\begin{aligned}\overline{\lambda}_i(t+1 \mid s_i(t) = 0 \cap \mu(t) = 1) &= \Pr[i^h \mid s_i(t) = 0 \cap \mu(t) = 1] \\ &= \frac{p\lambda_i(t)}{1-(1-p)\lambda_i(t)}.\end{aligned} \tag{13}$$

2.4.3 The non-revealing state

Finally, it is possible that service production is not observed in any jurisdiction, such that $s_i(t) = 0 \forall i$. In such an event, updating is quite complex and involves

$$\begin{aligned}\overline{\lambda}_i(t+1 \mid s_i(t) = 0\ \forall i) &= \Pr\left[i^h \mid s_i(t) = 0\ \forall i\right] \\ &= \left(\frac{\Pr[s_i(t) = 0\ \forall i \mid i^h]}{\Pr\left[s_i(t) = 0\ \forall i\right]}\right)\lambda_i(t).\end{aligned} \tag{14}$$

With a little manipulation (see Appendix) this can be shown to be equal to

$$\overline{\lambda}_i(t+1 \mid s_i(t) = 0\ \forall i) = \left(\frac{\Pi_{j\neq i}^n[1-\lambda_j(t)(1-p)]p(1-q)+q}{\Pi_{j=1}^n[1-\lambda_j(t)(1-p)](1-q)+q}\right)\lambda_i(t). \tag{15}$$

In the case of homogeneous jurisdictions, $\overline{\lambda}_j(t) = \overline{\lambda}_i(t) \forall i, j$, this can be further reduced to

$$\overline{\lambda}_i(t+1 \mid s_i(t) = 0\ \forall i) = \left(\frac{[1-\lambda(t)(1-p)]^{n-1}\ p(1-q)+q}{[1-\lambda(t)(1-p)]^n\ (1-q)+q}\right)\lambda_i(t). \tag{16}$$

It is not too difficult to show that

$$\frac{\Pi_{j\neq i}^n\left[1-\lambda_j(t)(1-p)\right]p(1-q)+q}{\Pi_{j=1}^n\left[1-\lambda_i(t)(1-p)\right](1-q)+q} \to 1 \text{ as } n \to \infty\,. \tag{17}$$

As such, $\overline{\lambda}_i(t+1 \mid s_i(t) = 0\forall i) \to \lambda_i(t)$ from below as $n \to \infty$. In other words, as the number of jurisdictions in which $s_i(t) = 0\forall i$ is observed increases, the probability that this is because $\mu(t) = 0$ approaches 1, so the observation contains no useful information on $\overline{\lambda}_i$.

Expressions (12), (13), and (15) describe the evolution of the beliefs of the population in jurisdiction i about the honesty of their politician–bureaucrat in each of the the three possible information states. These, together with the expected frequency with which the three information states will occur, allow us to study the expected level of corruption in an economy and it's relationship to decentralization.

2.5 Decentralization and the level of corruption

In this two period model, we know that dishonest politician–bureaucrats will engage in corruption in the first period. Hence, the level of expected total corruption in the first period is simply

$$E\left[C(t)\right] = \sum_{i=1}^{n} \left(\frac{m}{n}\right)(1 - \lambda_i(t)), \tag{18}$$

the sum over all the jurisdictions of the probabilities that taxes are appropriated by a corrupt politician–bureaucrat.

In the second period, after observing $s_i(t)\ \forall i$, each jurisdiction's population updates their beliefs of their own politician–bureaucrat's type. Those for whom

$$\lambda_i(t) - \frac{\eta}{(1-p)(1-q)} \geq \overline{\lambda}_i(t+1) \tag{19}$$

will choose to replace their current politician–bureaucrat with a new draw from the local population of potential politician–bureaucrats. Since replacement only takes place if the populations (correctly) believe that a random draw will reduce the frequency of corruption, it follows that if we can show that an increase in the number of jurisdictions increases the frequency of replacement then it also decreases the frequency of corruption.

Proposition 1 *The expected level of corruption in the second period is non-increasing in the level of decentralization provided that politician–bureaucrats are not fired when zero service production is observed in all jurisdictions. In the case where politician–bureaucrats are fired when zero service production is observed in all jurisdictions, the expected level of second period corruption is non-increasing in the level of decentralization for "almost all" levels of decentralization.*

Proof Appendix. □

As discussed above, there are three possible information states that can arise and for each the population in a jurisdiction i will update their beliefs accordingly using Bayes rule. In determining the relationship between the expected level of corruption and decentralization, we need to analyze both the effect of a change in the number of jurisdictions, n, on the updating process in each information state and on the frequency of each information state occurring. To provide intuitive insight into the proposition, we take these three information states in turn.

1. In the $\lambda_i(t)$-revealing information state a politician–bureaucrat is seen to be honest ($s_i(t) = 1$). Clearly, the level of decentralization cannot effect the level of expected corruption in i, as there is none. Further, the probability that this state will arise is independent of the number of other observations made (n).

2. In the $\mu(t)$-revealing information state, the public in jurisdiction i know that there is a positive economy-wide shock. Since they observe no service production in their own jurisdiction, they can only attribute this to an adverse local shock or a corrupt politician–bureaucrat and thus will reduce the probability they attribute to the politician–bureaucrat in i being honest. Hence, provided condition (10) is met, a replacement will occur. While the updating of $\lambda_i(t)$ is independent of n, the probability that this information state will occur is increasing in n. That is, as the number of jurisdictions becomes large, it becomes quite unlikely that *every* jurisdiction will either be run by a corrupt politician–bureaucrat and/or draw a zero jurisdiction-specific productivity shock. Hence, the likelihood of the $\mu(t)$-revealing information state occurring and of a replacement occurring is increasing in n. Expected corruption is therefore decreasing in n.
3. In the non-revealing information state an increase in n increases the probability that the population in i attaches to their politician–bureaucrat being honest. However, the likelihood that this information state will arise is decreasing in n, giving an ambiguous effect of decentralization on expected corruption.

The first part of Proposition 1 tells us that if direct cross-jurisdictional comparisons are required for politician–bureaucracts to experience replacement (i.e., if they occur only in the $\mu(t)$-revealing information state) then corruption is non-increasing in decentralization. Here, the only impediment to obtaining the stronger result of corruption strictly decreasing in the level of decentralization is the discreteness of the problem. The second part of the proposition is required because of a "knife edge" property of the analysis. There is a level of decentralization, n^*, at which the public in jurisdiction i are just indifferent between the replacement of a politician–bureaucrat in both the $\mu(t)$-revealing and non-revealing information states, and replacement in the $\mu(t)$-revealing information state only. At n^*, expected corruption can (but is not guaranteed to) discontinuously increase in decentralization (see the Appendix).

3 Decentralization and the level of corruption with pooling

In the preceding section we developed the inference problem behind the negative relationship between corruption and decentralization. The intuition was really quite simple; as the number of jurisdictions increased the likelihood that the information state in which honest politician–bureaucrats replaced dishonest ones also increased. In this section we allow for the possibility that decentralization may decrease corruption as a consequence of dishonest politician–bureaucrats choosing to modify their first period behavior so as to increase their chances of retaining office in the second period. That is the dishonest will pool with the honest in the first period. Notice that if more dishonest politician–bureaucrats choose to pool rather than separate overall corruption must fall, this follows from the fact that those that choose to separate steal with probability one on the first period, then with some probability retain office and steal again in the second, or are replaced with a new politician–bureaucrat who if they are also dishonest will themselves steal in that period. Alternatively a dishonest politician–bureaucrat that pools will not steal in the first period then will seal with probability one in the second. Clearly, the more pooling that takes place the lower will be the level of overall corruption. It follows then that if decentralization increases pooling it decreases corruption, which we shall see to be the case.

To facilitate this analysis we relax the assumption that $w_i(t) = w_i(t+1) = 0$ which immediately allows for the possibility that dishonest politician–bureaucrats will choose to pool with honest ones by not stealing tax revenues in the first period, that is the inequality in expression (5) may be reversed

$$r_i(t)+\delta\psi(t+1 \mid 1)[w_i(t+1)+r_i(t+1)] < \delta\psi(t+1 \mid 0)[w_i(t+1)+r_i(t+1)] \tag{20}$$

To investigate this channel for the effects of decentralization on corruption we need to explore how an increase in the number of jurisdiction effects whether or not dishonest politician–bureaucrats choose to pool. Clearly this works through the effects on the probabilities $\psi(t+1 \mid 1)$ and $\psi(t+1 \mid 0)$. Notice first that if a politician–bureaucrat chooses to pool, and it is known that they face these incentives then $\psi(t+1 \mid 0) = 1$, those jurisdiction where pooling will be chosen retain their politican-bureaucrat into the second period with probability one, it then follows that the pooling condition reduces to $r_i(t) + \delta\psi(t+1 \mid 1)[w_i(t+1)+r_i(t+1)] < \delta[w_i(t+1)+r_i(t+1)]$, hence rearranging slightly we get pooling if

$$\psi(t+1 \mid 1) < w_i(t+1) + r_i(t+1) - \frac{r_i(t)}{\delta} \tag{21}$$

Hence, if $\psi(t+1 \mid 1)$ is decreasing in the level of decentralization then the number of dishonest politician–bureaucrats that pool is increasing in the level of decentralization, we have

Proposition 2 *The number of dishonest politican-bureaucrats that choose to pool is non-decreasing in the level of decentralization, and hence corruption is non-increasing in the level of decentralization.*

Proof Appendix. □

The intuition behind this proposition follows immediatley from the basic inference problem. As the number of jurisdictions increases the likelihood that a politician–beaurocrat that separates being detected stealing increases (see Sect. 2.5 above) hence the probability of retaining office into the second period $\psi(t+1 \mid 1)$ falls and for some jurisdictions this will imply that condition (21) is newly satisfied.

4 Alternative interpretations of the theory

Corruption can take many forms. While our analysis is developed in terms of the direct appropriation of tax revenue by a politician–bureaucrat, it can be relatively easily modified to give insights into other forms of corruption.[6] For example, the taking of bribes

[6] We thank Eckhard Janeba for suggesting the need to explore this possibility.

to circumvent regulations is a frequently cited form of governmental or bureaucratic corruption. Suppose that there exist regulations which are both costly to enforce and comply with and that the return to the activity they permit is stochastic—being subject to both jurisdiction specific and economy wide shocks. For example, consider the regulations as those governing construction, where shocks arise as both jurisdiction-specific and economy-wide shifts in the demand for housing. Suppose, further, that the housing supply can either be in a high or low state. The low state occurs in a jurisdiction i if there is either a negative local shock, a negative national shock, or if the jurisdiction has a dishonest politician/bureaucrat who restricts construction permits to create rents that they may corruptly appropriate via personal bribes. The population in jurisdiction i may observe one of three information states—the high state in their own jurisdiction, the high state in some other jurisdictions but not locally, or the low state everywhere. That such can be captured and explained by the model developed above suggests that the analysis has some general applicability.

5 Proximity

The main results in our analysis of the relationship between corruption and decentralization are driven by the effects of more observations (jurisdictions) on the inference problems being solved by the populations of each jurisdiction. The key element that makes these inter-jurisdictional comparisons informative is the sharing of a common productivity shock $\mu(t)$. It is then a small step to think of an economy as being divided up into many informational islands each of which might share a common productivity shock. For example, in the USA the western states of Oregon, Washington and California all devote considerable areas of land to the production of timber, an activity which is heavily regulated. It seems reasonable to assume, then, that there are common shocks across this pacific northwest timber island, that are not relevant to, say, the wheat island of the Midwest plains states. Many of these informational islands are geographically contiguous, both because of terrain and because of the benefits to certain activities of agglomeration. If follows that observations on the provision of services in Massachusetts might be relevant to the population of New York but not those in Arizona (snow clearance comes to mind). Informational islands share common experiences. If the populations on island 1 observe the $\mu_1(t)$-revealing information state, while the populations on island 2 observe the non-revealing state, then we would predict that most of the politician–bureaucrats on island 1 would share the common experience of separation (together with other sanctions), while those on island 2 would continue with business as normal.

6 Empirical analysis

The preceding theoretical analysis supports the empirical regularity that corruption and decentralization should be negatively correlated. This is therefore a focus of our empirical analysis. To the extent that informational islands span state boundaries, however, an interpretation of the theory justifies one's neighbors' decentralizations being put on the right-rand-side of explaining one's corruption. For example, in the case of

Oregon, one might argue that California's and Washington's decentralization could be partially explanatory to Oregon's corruption, as the potential for common shocks to these three states implies that their corruption behavior itself will be correlated. After introducing baseline specifications, we will control for this potential econometrically through the estimation of a spatial autoregressive component to corruption.

6.1 Data

The existing literature has documented a negative relationship between decentralization and corruption (e.g., de Mello and Barenstein 2001; Fisman and Gatti 2002a). In confronting the issue with U.S. state-level data, where we exploit both time-series and cross-sectional variation, we document empirical regularities that are consistent with these relationships. In so doing, for each state-year we define the dependent variable, $Corruption$, as the number of corruption-related federal convictions per 100,000 population. A similar measure is used in Goel and Rich (1989); Fisman and Gatti (2000b); Fredrikson et al. (2003) and Glaeser and Saks (2004). As is standard in the literature, we measure fiscal decentralization as the non-central government share of total state expenditures, equal to the total expenditure of local government divided by the total expenditure by all levels of government (state and local).

To a large extent, we follow the literature in controlling for other variation in state characteristics that may explain corruption. For example, Fisman and Gatti (2002a) suggest that less developed countries are likely to be more corrupt.[7] Hence, while we expect much less significance in relative development measures explaining relative corruption, we control for the level of development across states with the inclusion of real gross state product. Following Glaeser and Saks (2006) and Dincer (2008), we also control for ethnic fractionalization, which may increase corruption by decreasing the popular will to oppose corruption. Specifically, we use data from the 1990 Census to calculate the ethnic fractionalization index; defined as $1 - \sum_i s_{si}^2$, where s_{si} is the population share of group i in state s. Ethnic categories included in the calculation of the index are Hispanic, White, Black, Asian, American Indian and Eskimo, Hawaiian and Pacific Islander and Others, and can be interpreted as the probability that two randomly selected individuals in state s belong to two different ethnic groups. We also control for government wages deflated by personal income, as cross-sectional variation in government wages may correlate with incentives for government officials to extract personal rents through corruption. For example, Goel and Rich (1989) and van Rijckeghem and Weder (2001) find a negative relationship between the level of government wages and the level of corruption.

Finally, we control for state population, manufacturing employment, education levels (i.e., the share of the population with less than high-school and less than college) and a measure of income inequality in state-level Gini coefficients (following Uslaner 2008). In all specifications the sample contains the 48 contiguous U.S. states over the

[7] Relative to studies that rely on cross-sectional variation in cross-country samples, our analysis may be viewed as less sensitive to bias due to unobserved country-specific heterogeneity.

Table 1 Summary statistics, 48 contiguous states, 1987–2001, ($N = 698$)

Variable	Mean	Std. Dev.	Min.	Max.
Corruption[a]	0.317	0.276	0	2.131
Ethnic fractionalization[b]	0.282	0.149	0.037	0.593
Gross state product (billions)[c]	152.242	182.424	11.701	1260.041
Government wage relative to personal income[d]	1.247	0.147	0.892	1.842
Population (millions)[b]	5.339	5.732	0.454	34.6
Manufacturing employment[c]	0.397	0.399	0.009	2.226
Share of population with less than high school[e]	0.485	0.046	0.393	0.616
Share of population with less than college[e]	0.86	0.032	0.752	0.935
Income inequality (Gini coefficient)[e]	0.35	0.035	0.274	0.452
South	0.335	0.472	0	1
Northeast	0.185	0.388	0	1
Midwest	0.256	0.437	0	1

[a] Source: U.S. Justice Department, Report to Congress on the Activities and Operations of the Public Integrity System, defined separately for each state as the number of government officials convicted for federal crimes relating to corruption per 100,000 population

[b] Source: US Census Bureau. (Note that ethnic fractionalization is only available as part of the decennial U.S. Census. We adopt 1990 levels)

[c] Source: Bureau of Economic Analysis. (Real Gross State Product measured in 1996 dollars)

[d] Source: Bureau of Labor Statistics

[e] Source: Frank (2009). (Income inequality is derived from Internal Revenue Service data. Education variables are constructed from the Current Population Survey)

1987–2001 period. Summary statistics for all variables are provided in Table 1 where we also provide additional detail on sample construction.

6.2 Specification and results

Our base specification is as follows:

$$Corruption_{st} = \beta_0 + \beta_1 Decentralization_{st} + \beta_2 X_{st} + \epsilon_{st}, \tag{22}$$

where for state s in year t, X_{st} represents the set of other state characteristics that influence corruption (i.e., government wages, gross state product, ethnic fractionalization, population, manufacturing employment, education, income inequality and quadratic trend) and ϵ_{st} represents the error term, where we allow errors to cluster around state observations.

The baseline results from a pooled OLS specification are given in the first column of Table 2. As anticipated, the point estimate on *Decentralization* is negative and significant at conventional levels, suggesting that higher decentralization is associated with lower levels of corruption. Point estimates on all control variables yield expected signs. As in the existing literature, estimated coefficients of gross state product and government wages are negative and the estimated coefficients of ethnic fractionalization,

Table 2 Corruption and fiscal decentralization, US States, 1987–2001

	(1)	(2)	Uncorrected for spatial endogeneity (3)	Corrected for spatial endogeneity (4)	Corrected for spatial endogeneity and instrument decentralization (5)[a]
Fiscal decentralization[b]	−0.439*	−0.303	−0.308	−0.311	−0.659*
	(0.261)	(0.379)	(0.363)	(0.209)	(0.365)
Spatially weighted neighborhood			0.273***	0.419**	0.473**
corruption			(0.080)	(0.194)	(0.190)
Ethnic fractionalization[c]	0.030	0.187	0.225	0.245*	0.256*
	(0.254)	(0.287)	(0.282)	(0.147)	(0.151)
Ln (gross state product)	−0.046	−0.049	−0.064	−0.072	−0.074
	(0.187)	(0.194)	(0.181)	(0.116)	(0.129)
Government wage relative to	−0.109	0.012	−0.016	−0.031	−0.073
personal income	(0.097)	(0.114)	(0.113)	(0.077)	(0.083)
Ln (State population)	0.175	0.210	0.226	0.234**	0.255*
	(0.155)	(0.168)	(0.152)	(0.117)	(0.138)
Ln (Manufacturing employment)	−0.090	−0.151**	−0.151**	−0.151***	−0.151***
	(0.076)	(0.063)	(0.061)	(0.038)	(0.039)
Share of population with less than	1.042	1.203*	1.214*	1.221***	1.109**
high school	(0.792)	(0.625)	(0.608)	(0.416)	(0.496)
Share of population with less than	0.334	0.069	0.001	−0.036	0.078
college	(0.780)	(0.753)	(0.744)	(0.480)	(0.560)
Income inequality (Gini coefficient)	0.662	0.890	0.765	0.698	0.800
	(0.852)	(0.638)	(0.647)	(0.509)	(0.598)
Linear trend (1987 = 1)	−0.022*	−0.020*	−0.013	−0.008	−0.010
	(0.013)	(0.012)	(0.012)	(0.010)	(0.013)

Table 2 continued

	(1)	(2)	Uncorrected for spatial endogeneity (3)	Corrected for spatial endogeneity (4)	Corrected for spatial endogeneity and instrument decentralization (5)[a]
Quadratic trend	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
South		0.112	0.069	0.046	0.011
		(0.071)	(0.072)	(0.054)	(0.097)
Northeast		0.161**	0.147**	0.139***	0.104
		(0.065)	(0.063)	(0.043)	(0.106)
Midwest		0.209***	0.186***	0.174***	0.164***
		(0.057)	(0.056)	(0.045)	(0.047)
Constant	−2.821*	−3.748**	−3.890***	−3.966***	−4.101***
	(1.473)	(1.514)	(1.384)	(1.151)	(1.256)
Observations	698	698	698	698	698
R^2	0.10	0.14	0.16	0.16	0.15

Using data available from the US Justice Department, Report to Congress on the Activities and Operations of the Public Integrity Section,the dependent variable, *Corruption*, is defined separately for each state as the number of government officials convicted for federal crimes relating to corruption per 100,000 population. Columns (1)–(3) are OLS with errors clustered at the state level. Columns (4) and (5) correct for the endogeneity of the spatial autoregressive term and fiscal decentralization via two-stage-least squares

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses

[a] The Wooldridge (1995) robust score test cannot reject the null that OLS would produce consistent estimates for the same regression. (The test statistic is $F(2{,}680) = 1.355$, $p = 0.259$)

[b] Fiscal decentralization is not available in the year 2000. Thus, in order to preserve the consistent time-series available in other dimensions of the data, we impute fiscal decentralization (separately for each state) for the year 2000 using a linear imputation rule. However, the fiscal decentralization result is robust to a sample that drops the year-2000 observations entirely

[c] Ethnic fractionalization (EFI) is drawn from the decennial U.S. Census. Likely due to there being so little time-series variation in EFI within states (e.g., the correlation coefficient between 1990 and 2000 is above 0.98), the specifications are not sensitive to including either an EFI level or to drawing a line through the observations of EFI available to us to approximate the change implied between the Census observations. Here, we report the results when including the 1990 level of EFI, although the results are robust to such an imputation method between 1990 and 2000

population and the share of population with less than high school or college education are positive. However, these coefficients tend to be measured somewhat imprecisely.

In Column (2) we add Census-region indicator variables, which are intended to capture any variation in corruption levels that is specific to regional-specific level effects. While the point estimate on $Decentralization$ remains negative, the estimate is imprecise and statistical significance is lost. One might interpret this as suggestive of some geographic variation in corruption and decentralization that is particular to regions. That said, while the empirical pattern identified in Column (1) appears to be sensitive to controlling for geographic region, regional controls may actually proxy for other factors that introduce some commonality to signals within regions.[8]

While this lack of significance of $Decentralization$ in predicting $Corruption$ is somewhat surprising initially, the potential for yardstick competition suggests that any empirical pattern in $Corruption$ (such as that identified in columns 1 and 2) may well depend on that observed in proximate states.[9] Moreover, if corruption increases where it is known that others are corrupt, not accounting for this influence dependency may attenuate a negative relationship between corruption and decentralization as decentralization varies across states. Given the indication (from Column 2) that variation in the data is, in part, explained by geographic region, investigating the potential neighbor-to-neighbor interactions may be a fruitful exercise.

6.2.1 Spatial endogeneity

In Column (3) of Table 2, we report the model that includes the average corruption levels among neighboring states as an explanatory variable in predicting each state-year observation of corruption. This is similar in approach to Fisman and Gatti (2002b), where average corruption levels in neighboring states are included as a control variable, with little or no explanatory power. However, the simple inclusion of this "neighborhood" average without proper account for the spatial-endogeneity bias can be misleading. Before correcting for bias, the results can be interpreted in light of the theory as a type of yardstick competition insofar as corruption levels in neighboring states positively predict one's corruption, which is consistent with information islands spanning state boundaries in a way that ties state-level outcomes (such as corruption) together. (Note that the estimated coefficient on fiscal decentralization remains statistically imprecise.)

Correcting for the spatial endogeneity implies estimating a model of the form:

$$Corruption_{st} = \beta_0 + \beta_1 Decentralization_{st} + \beta_2 X_{st} + \rho W \cdot Corruption_{st} + \epsilon_{st}, \tag{23}$$

[8] The inclusion of state fixed effects sufficiently captures variation in corruption levels so as to leave no additional explanatory power to measured levels of decentralization. As such, the patterns we report are appropriately thought of as cross-sectional in nature.

[9] For example, the concentration of the lumber industry in the Pacific Northwest likely introduces some value to Oregon and Washington learning from each others' experiences. If such is the case, outcomes in these two proximate states should appear, ex post, to correlate.

where the addition of $\rho W\cdot$ Corruption reflects the spatial autoregression term itself, W is the spatial-lag weighting matrix, and ρ is a parameter to be estimated, speaking to the sign and strength of any spatial relationship in Corruption. We adopt a simple weighting scheme of strict state-contiguity, such that element $w_{rc} = 1$ if $r \neq c$ and state r is contiguous to state c and $w_{rc} = 0$ otherwise. As in the simple model of Column (3), $W \cdot$ Corruption$_{st}$ can be interpreted as the average level of corruption in state s's contiguous states in year t. In the spatial econometrics literature, such weighting schemes are quite common, in part as it imposes little structure on the spatial relationship and is easily interpretable.[10] As measured by the I statistic of Moran (1950), one should note that there is significant positive spatial correlation in per capita personal income, GSP, and education across contiguous states. Moran's I values are positive and statistically significant at a one percent level for all co-variates in Table 2 but GSP (which is significant at three percent). In other words, geographic closeness is a good proxy for closeness in a broad sense. The effects of geographic closeness on economic variables such as the growth rate of income attract a great deal of attention in the literature as well. Garrett et al. (2007), for example, found a positive spatial correlation in income growth across neighboring states.

To estimate (23), we employ classical two-stage, least-squares procedures common to the spatial econometrics literature, where WX are employed as instruments for $W \cdot Corruption_{st}$.[11] The estimated coefficients from (23) are reported in Column (4) of Table 2. While we will employ geographic contiguity as our exogenous metric for assigning closeness, particular stories could be told to support other assignment rules. However, we proceed acknowledging that to the extent geography misses the real measure of closeness, the estimated influence of one state on a neighboring state should attenuate.

Here, we note several important regularities. First, there is a general increase in the predictive power of control variables resulting from the estimated spatial autoregressive model. In all cases, point estimates tend to be more-precisely estimated and often increase in magnitude. In all cases, they remain of the anticipated sign. Of particular note, controlling for the spatially autoregressive dependent variable yields increased precision in our estimate of the direct effect of $Decentralization$ on corruption, although the point estimate itself is quite stable and remains insignificant ($p = 0.13$). If the true model of $Corruption$ is spatially autoregressive, models that fail to account for this relationship (e.g., columns 1 through 3) may well introduce bias into the estimation of other covariates—in particular, in other spatially-correlated control variables.

[10] In a spatially lagged dependent variable model as the one we adopt, spatial dependence is something of substance in the sense that the dependent variable in one state is potentially influenced by the value of the dependent variable in other states. To the contrary, adopting a spatial-error model would be to treat spatial correlation primarily as nuisance—as something to be eliminated, and solely as an estimation problem. Put another way, the spatial-error model focuses on estimating the underlying parameters of interest and disregards the possibility that the observed correlation in dependent variables may reflect something meaningful about the data generating process.

[11] At this point we will direct the reader to Anselin (1988) for further discussion of this and other appropriate methods of correcting for spatial endogeneity in linear models.

Second, we interpret our results as an indication of the importance of appropriately accounting for the presence of spatially autoregressive relationships in *Corruption* and that state attributes alone are insufficient to capture the complexity of the apparent relationships within U.S. data. Controlling for average differences by Census region, the spatial-lag model yields a positive and significant lag coefficient, implying that corruption levels are generally increasing in the corruption of neighboring states. We interpret this as the presence of significant "yardstick competition" in the data. We also note that correcting for the endogeneity of spatially-weighted *Corruption* levels increases the suggested importance of neighborhood corruption, as compared to the uncorrected model (in Column 3). After correcting for bias, we interpret this result in light of the theory as a type of yardstick competition insofar as corruption levels in neighboring states positively predict state-level corruption. That is, we interpret such empirical regularity as consistent with information "islands" spanning state boundaries in a way that ties state-level outcomes (such as corruption) together.

In the end, the point estimates of Column (4) imply that the number of corruption convictions per 100,000 population decreases by roughly 12 percent (or 0.04 convictions or every 100,000 of population) over the inner-quartile range of fiscal decentralization. Further, the statistically significant spatial-lag coefficient is meaningful as it implies that the average state's predicted corruption rate increases some 29%—from 0.27 per 100,000 in population to 0.35 per 100,000—across the inner-quartile range of neighboring corruption levels.

6.2.2 Instrument for decentralization

The above analysis has presumed fiscal decentralization to be exogenous to corruption. However, as suggested in related literature, one may suspect that corrupt governments themselves oppose fiscal decentralization insofar as it hampers their ability to steal. In particular, then, we follow Arikan (2004) and instrument for decentralization with the surface area of the state in square miles. As argued by Arikan (2004), while the surface area is often used as an explanatory variable for decentralization, it is not likely to explain the level of corruption—that is, land area correlates well with the potentially endogenous variable yet remains uncorrelated with the error term, thus yielding a reasonable instrument. These results are reported in Column (5) of Table 2.

Overall, instrumenting for *Decentralization* suggests a degree of robustness to the general patterns evident in the earlier specifications. The most notable change, however, is the marked increase in the point estimate on the instrumented variable itself, *Decentralization*. In fact, the point estimates of Column (5) suggest that the number of corruption convictions per 100,000 population decreases by 22% over the inner-quartile range of fiscal decentralization (or 0.8 convictions for every 100,000 of population).

In terms of testing for whether the suspect endogenous regressor in the model is in fact exogenous, we employ Wooldridge (1995) robust score test given our robust standard errors. We cannot reject the null that OLS would produce consistent estimates for the same regression, which suggests that there may be no need to instrument. In a sample of countries, Arikan (2004) arrives at the same conclusion. Adopting a different exclusion restriction, Fisman and Gatti (2002a) also report an increase in the estimated coefficient but conclude that "although there might be *a priori* good reasons to expect a potential endogeneity and/or mismeasurement problem, the resulting bias in the OLS coefficient is not significant."[12] Similarly, then, we take away that it is "conceptually important to bear the endogeneity issue in mind as research on the link between fiscal decentralization and corruption proceeds," Arikan (2004).

In interpreting the estimated coefficient on $Decentralization$ in Column (5), one should bear in mind, however, that that the effect of $Decentralization$ is revealed only for the subsample that is "responsive" to the cross-sectional variation in surface area, and that the subsample which varies most with the instrument will have the largest effects on the magnitude of the IV estimate itself.[13]

7 Conclusion

We present a model which uses the yardstick competition approach proposed by Salmon (1987) and Besley and Case (1995) to explain the negative relationship between decentralization and corruption. The model demonstrates that as the population in a jurisdiction make observations across jurisdictions, the level of corruption is decreasing in decentralization. Further, the model can easily be reinterpreted to predict that jurisdictions close together will tend to experience similar levels of corruption.

The empirical regularities we demonstrate are largely consistent with the patterns suggested by the model. Using federal corruption-related convictions in U.S. states for the period 1987–2001, we find suggestive evidence of the role of fiscal decentralization in explaining corruption, with the level of corruption declining with decentralization. However, the role of geography – something which can proxy for information flows—is somewhat complex. Ultimately, we identify a significant spatially autoregressive relationship in corruption, indicating that one state's corruption is positively related to the level of corruption in adjacent states. Where relevant information spans state borders, this suggests that yardstick competition is a good candidate explanation for the overall patterns observed in the data.

8 Appendix

8.1 Derivation of (13)

We have

$$\overline{\lambda}_i(t+1) = \Pr[i^h \mid s_i(t) = 0 \cap \mu(t) = 1]$$

[12] We also considered each state's status as a post revolution civil-law state, in the spirit of Fisman and Gatti (2002a), but find no predictive power in this instrument.

[13] For example, this empirical approach does not inform the researcher in any way about the effect of $Decentralization$ among states who would either always or never be decentralized regardless of surface area. (See Imbens and Angrist (1994) for additional discussion.)

since

$$\Pr[i^h \mid s_i(t) = 0 \cap \mu(t) = 1]\Pr[s_i(t) = 0 \mid \mu(t) = 1]\Pr[\mu(t) = 1]$$
$$= \Pr[s_i(t) = 0 \mid i^h \cap \mu(t) = 1]\Pr[i^h \mid \mu(t) = 1]\Pr\,[\mu(t) = 1]$$

it follows that

$$\begin{aligned}\overline{\lambda}_i(t+1) &= \Pr[i^h \mid s_i(t) = 0 \cap \mu(t) = 1] \\ &= \frac{\Pr[s_i(t) = 0 \mid i^h \cap \mu(t) = 1]\Pr\left[i^h \mid \mu(t) = 1\right]}{\Pr[s_i(t) = 0 \mid \mu(t) = 1]} \\ &= \frac{p\lambda_i(t)}{1 - (1-p)\lambda_i(t)}\end{aligned}$$

8.2 Derivation of (15)

We have

$$\begin{aligned}\overline{\lambda}_i(t+1) &= \Pr[i^h \mid s_j(t) = 0\ \forall j] \\ &= \left(\frac{\Pr[s_j(t) = 0\ \forall j \mid i^h]}{\Pr[s_j(t) = 0\ \forall j]}\right)\lambda_i(t),\end{aligned}$$

so

$$\begin{aligned}\Pr[s_j(t) = 0\ \forall j] &= \Pr[s_j(t) = 0\ \forall j \mid \mu(t) = 1]\,\Pr[\mu(t) = 1] \\ &\quad + \Pr[s_j(t) = 0\ \forall j \mid \mu(t) = 0]\,\Pr[\mu(t) = 0] \\ &= \Pr[s_j(t) = 0\ \forall j \mid \mu(t) = 1](1-q) + q\end{aligned}$$

and

$$\begin{aligned}\Pr[s_j(t) = 0\ \forall j \mid \mu(t) = 1] &= \Pi_{j=1}^{n}[\Pr[\varepsilon_j = 1 \cap \tau_j = 1] + \Pr[\varepsilon_j = 0]] \\ &= \Pi_{j=1}^{n}[\Pr[\tau_j = 1 \mid \varepsilon_j = 1]\,\Pr[\varepsilon_j = 1] + \Pr[\varepsilon_j = 0]] \\ &= \Pi_{j=1}^{n}[(1-\lambda_j(t))(1-p) + p)]\end{aligned}$$

So

$$\Pr\left[s_j(t) = 0\ \forall j\right] = \Pi_{j=1}^{n}\left[(1-\lambda_j(t))(1-p) + p\right](1-q) + q$$

also

$$\begin{aligned}\Pr[s_j(t)=0\ \forall j \mid i^h] &= \Pr[s_j(t)=0\ \forall j \mid \mu(t)=1\cap i^h]\Pr[\mu(t)=1]\\&\quad+\Pr[\mu(t)=0]\\&=\Pi^n_{j\neq i}[\Pr[\tau_j=1\mid \varepsilon_j=1]\Pr[\varepsilon_j=1]\\&\quad+\Pr[\varepsilon_j=0]]\Pr[\varepsilon_i=0]\Pr[\mu(t)=1]+\Pr[\mu(t)=0]\\&=\Pi^n_{j\neq i}[1-\lambda_j(t)(1-p)]p(1-q)+q\end{aligned}$$

So finally

$$\overline{\lambda}_i(t+1)=\left(\frac{\Pi^n_{j\neq i}\left[1-\lambda_j(t)(1-p)\right]p(1-q)+q}{\Pi^n_{j=1}\left[1-\lambda_j(t)(1-p)\right](1-q)+q}\right)\lambda_i(t)$$

If $\overline{\lambda}_j(t)=\overline{\lambda}_i(t)\forall i, j$ then

$$\overline{\lambda}_i(t+1)=\left(\frac{[1-\lambda_j(t)(1-p)]^{n-1}p(1-q)+q}{[1-\lambda_j(t)(1-p)]^{n}(1-q)+q}\right)\lambda_i(t)$$

8.3 Proof of Proposition 1

To prove this proposition we begin with a couple of simple lemmas

Lemma 3 *The frequencies at which the three information states arise are given by*

1. $\Pr[s_i(t)=1]=(1-p)(1-q)\lambda_i(t)$
2. $\Pr[s_i(t)=0\cap s_j(t)=1 \text{ some } j]=\left[1-\left(\Pi^n_{j\neq i}\left[1-\lambda_j(t)(1-p)\right](1-q)+q\right)\right]$
   $\times[1-(1-p)(1-q)\lambda_i(t)]$
3. $\Pr[s_i(t)=0\forall i]=\left(\Pi^n_{j=1,j\neq i}\left[1-\overline{\lambda}_j(t)(1-p)\right](1-q)+q\right)$
   $\times[1-(1-p)(1-q)\lambda_i(t)]$

Lemma 4 *No separation occurs in jurisdiction i if the information state* $s_i(t)=1$ *occurs, nor is the frequency of this information state effected by the level of decentralization.*

Proof Follows immediately from $s_i(t)=1$ which reveals the politician–bureaucrat/bureaucrat i is honest. □

Lemma 5 *If for a given set of parameter values a separation occurs in jurisdiction i for the information set* $s_i(t)=0\forall i$ *then it also occurs for the information set* $s_i(t)=0\cap s_j(t)=1$ *some* j*. But the converse is not true.*

Proof Involves demonstrating

$$
\begin{aligned}
&\left(\frac{\Pi_{j\neq i}^{n}[1-\lambda_j(t)(1-p)]p(1-q)+q}{\Pi_{j=1}^{n}[1-\lambda_j(t)(1-p)](1-q)+q}\right) \\
&\quad > \left(\frac{p}{1-(1-p)\lambda_i(t)}\right) \\
&\Rightarrow \left(\Pi_{j\neq i}^{n}[1-\lambda_j(t)(1-p)]p(1-q)+q\right)(1-(1-p)\lambda_i(t)) \\
&\quad > p\left(\Pi_{j=1}^{n}\left[1-\lambda_j(t)(1-p)\right](1-q)+q\right) \\
&\Rightarrow \Pi_{j=1}^{n}\left[1-\lambda_j(t)(1-p)\right]p(1-q)+q\,(1-(1-p)\lambda_i(t)) \\
&\quad > \Pi_{j=1}^{n}\left[1-\lambda_j(t)(1-p)\right]p(1-q)+pq \\
&\Rightarrow q\,(1-(1-p)\lambda_i(t)) > pq \\
&\Rightarrow 1 > p
\end{aligned}
$$

as required. □

Lemma 6 *The frequency with which the information state $s_i(t) = 0 \cap s_j(t) = 1$ some j occurs is equal to one minus the frequency with which the information state $s_i(t) = 0 \forall i$ occurs. Hence the effects of an increase in decentralization on the frequency that information state $s_i(t) = 0 \cap s_j(t) = 1$ some j occurs is equal to minus its effect on the frequency with which the information state $s_i(t) = 0 \forall i$ occurs.*

Proof By definition $\Pr[s_i(t) = 0 \cap s_j(t) = 1 \text{ some } j] = 1 - \Pr[s_i(t) = 0 \forall i]$ and the proof is immediate. □

Proof of the Proposition We now know there are three possibilities

1. No separation in jurisdiction i—hence

$$
\begin{aligned}
\left(\frac{\Pi_{j\neq i}^{n}[1-\lambda_j(t)(1-p)]p(1-q)+q}{\Pi_{j=1}^{n}[1-\lambda_j(t)(1-p)](1-q)+q}\right)\lambda_i(t) &> \left(\frac{p}{1-(1-p)\lambda_i(t)}\right)\lambda_i(t) \\
&> \lambda_i - \frac{\eta}{(1-p)(1-q)}
\end{aligned}
$$

   here n clearly has no effect on corruption.
2. Separation in jurisdiction i only in the information state $s_i(t) = 0 \cap s_j(t) = 1$ some j—hence

$$
\begin{aligned}
\left(\frac{\Pi_{j\neq i}^{n}[1-\lambda_j(t)(1-p)]p(1-q)+q}{\Pi_{j=1}^{n}[1-\lambda_j(t)(1-p)](1-q)+q}\right)\lambda_i(t) &> \lambda_i - \frac{\eta}{(1-p)(1-q)} \\
&> \left(\frac{p}{1-(1-p)\lambda_i(t)}\right)\lambda_i(t)
\end{aligned}
$$

so an increase in n has no effect on corruption *within* either the information state $s_i(t) = 0 \cap s_j(t) = 1$ some j or *within* the information state $s_i(t) = 0 \forall i$, but it increases the relative frequency with which information state $s_i(t) = 0 \cap s_j(t) = 1$ some j arrises. Since separations occur only in this information state it follows that the frequency of separations increases and hence expected corruption falls as decentralization increases.

3. Separation in jurisdiction i occurs in both information states $s_i(t) = 0 \cap s_j(t) = 1$ some j and $s_i(t) = 0 \forall i$—hence

$$\lambda_i - \frac{\eta}{(1-p)(1-q)} > \left(\frac{\Pi^n_{j \neq i}[1 - \lambda_j(t)(1-p)]p(1-q) + q}{\Pi^n_{j=1}[1 - \lambda_j(t)(1-p)](1-q) + q} \right) \lambda_i(t)$$
$$> \left(\frac{p}{1-(1-p)\lambda_i(t)} \right) \lambda_i(t)$$

here since separations occur in both information states n does not effect the frequency of separations and hence decentralization does not effect corruption. To demonstrate the caveat in the proposition we note that $\frac{\Pi^n_{j \neq i}[1-\lambda_j(t)(1-p)]p(1-q)+q}{\Pi^n_{j=1}[1-\lambda_j(t)(1-p)](1-q)+q}$ is increasing in n hence there is an n^* at which there is a flip from case 3 to case 2. This can lead to a drop in expected separations if the effect of moving between the two cases outweighs the effect on the probability of the two information states occurring. □

8.4 Proof of Proposition 2

To prove the proposition we need to show that $\psi(t+1 \mid 1)$ is non-increasing in the number if jurisdictions n. We consider only the case where separations occur in jurisdiction i in the information state $s_i(t) = 0 \cap s_j(t) = 1$ some j. From the proof of proposition 1 we know that the frequency with which this information state arrises may be written

$$\Pr[s_i(t) = 0 \cap s_j(t) = 1 \text{ some } j]$$
$$= [1 - \left(\Pi^n_{j \neq i}[1 - \lambda_j(t)(1-p)](1-q) + q\right)][1 - (1-p)(1-q)\lambda_i(t)],$$

being careful to note that we must set $\lambda_j = 1$ in those jurisdictions in which the politician–bureaucrats are engaging in pooling behavior. In choosing whether or not to separate or pool a bad politician in jurisdiction i knows their own type and hence in forming their expectation sets $\lambda_i(t) = 0$ which then gives the probability that they expect to suffer a separation, hence we have

$$\Pr[s_i(t) = 0 \cap s_j(t)$$
$$= 1 \text{ some } j \mid i^d] = \left[1 - \left(\Pi^n_{j \neq i}[1 - \lambda_j(t)(1-p)](1-q) + q\right)\right].$$

By definition, the probability that a dishonest politician–bureaucrat in jurisdiction i will retain office for the second period is

$$\begin{aligned}\psi(t+1 \mid 1) &= 1 - \Pr[s_i(t) = 0 \cap s_j(t) = 1 \text{ some } j \mid i^d] \\ &= \Pi_{j \neq i}^{n} \left[1 - \lambda_j(t)(1-p)\right](1-q) + q,\end{aligned}$$

which is non-increasing in n as required. □

Acknowledgments We thank Eckhard Janeba, Ron Davies, Joe Stone, Jon Fiva and participants at the IIPF 2005 conference in Korea and the PET 2005 conference in Marseille for many useful comments.

References

Allers M, Elhorst J (2005) Tax mimicking and yardstick competition among local governments in the Netherlands. Int Tax Public Finance 12:493–513

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Boston

Arikan GG (2004) Fiscal decentralization: a remedy for corruption? Int Tax Public Finance 11:175–195

Bardhan P, Mookherjee D (2009) "Decentralization, corruption and government accountability: an overview" in the handbook of economic Corruption, In: Susan RA (ed) Edward Elgar (in press)

Besley T (2006) Principled agents? The political economy of good government, the lindahl lectures. Oxford University Press, Oxford

Besley T, Case A (1995) Incumbent behavior: vote seeking, tax setting, and yardstick competition. Am Econ Rev 85:25–45

Breton A, Gianluigi G, Pierre S, Ronald W (2007) The economics of transparency in politics. Ashgate Publishing Ltd, Farnham

de Mello L (2004) Can fiscal decentralization strengthen social capital?. Public Finance Rev 32:4–35

de Mello L, Barenstein M (2001) Fiscal decentralization and governance: a cross country analysis. IMF Working Paper WP/01/71

Dincer OC (2008) Ethnic and religious diversity and corruption. Econ Lett 99:98–102

Fisman R, Gatti R (2002a) Decentralization and corruption: evidence across countries. J Public Econ 83:325–345

Fisman R, Gatti R (2002b) Decentralization and corruption: evidence from U.S. Federal Transfer Programs. Public Choice 113:25–35

Frank MW (2009) Inequality and growth in the United States: evidence from a new state level panel of income inequality measures. Econ Inq 47(1):55–68

Fredrikson PG, List JA, Millimet DL (2003) Bureaucratic corruption, environmental policy and inbound US FDI: theory and evidence. J Public Econ 87:1407–1430

Garrett TA, Wagner GA, Wheelcock DC (2007) Regional disparities in the spatial correlation of state income growth. Annals Reg Sci 41(3):601–618

Glaeser E, Saks RE (2006) Corruption in America. J Public Econ 90:1053–1072

Goel RK, Rich DP (1989) On the economic incentives for taking bribes. Public Choice 61:269–275

Imbens G, Angrist J (1994) Identification and estimation of local average treatment effects. Econometrica 62:467–476

Lazear EP, Rosen S (1981) Rank-order tournaments as optimum labor contracts. J Political Econ 89:841–864

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–33

Revelli F (2002) Testing the tax mimicking versus expenditure spillover hypotheses using English data. Appl Econ 34:1723–1731

Salmon P (1987) Decentralization as an incentive scheme. Oxf Rev Econ Policy 3(2):24–43

Smart M, Timothy B (2007) Fiscal restraints and voter welfare. J Public Econ 91:755–773

Sole Ole A (2003) Electoral accountability and tax mimicking: the effects of electoral margins, coalition government, and ideology. Eur J Political Econ 19:685–713

Treisman D (2000) Decentralization and the quality of government. Mimeo

Uslaner Eric M (2008) Corruption, inequality, and the rule of law: the bulging pocket makes the easy life. Cambridge University Press, Cambridge

van Rijckeghem C, Weder B (2001) Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much? J Dev Econ 65:307–331

Wooldridge JM (1995) Score diagnostics for linear models estimated by two stage least squares. In: Maddala GS, Phillips PCB, Srinivasan TN (eds) Advances in econometrics and quantitative economics: essays in honor of Professor C. R. Rao. Blackwell, Oxford, pp 66–87