

Corruption, Decentralization and Yardstick Competition

Christopher J. Ellis* and Oguzhan C. Dincer**

March 2005.

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. Further, using data on government corruption in US states, we provide some new evidence that supports the theoretical findings.

Keywords: Corruption, decentralization, yardstick competition

JEL classification: H20, H29.

*Corresponding Author:
Department of Economics,
University of Oregon,
Eugene, OR 97403-1202
USA.
Email: cjellis@oregon.uoregon.edu

**Department of Commerce,
Massey University,
Private Bag 102 904,
Auckland
New Zealand.

Acknowledgements:

1 Introduction

Recent works by several contributors point 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 corruption¹ (Treisman 2000, de Mello 2000, de Mello and M. Barenstein 2001, Fisman and Gatti 2002a, Arikan 2004) On the surface this is quite surprising. It might be predicted that local politician-bureaucrats and bureaucrats possess detailed local knowledge of any opportunities for corruption that might arise in their jurisdictions. They may thus be expected to extract any corruption rents available more efficiently than less well informed national politician-bureaucrats.

In what follows we propose both a theoretical explanation for, and find further evidence in support of, the observed negative empirical relationship between corruption and decentralization. We argue that because there are more independent taxation and expenditure decisions made in a decentralized economy, then there are more opportunities for local populations to make cross-jurisdictional comparisons of politician-bureaucrats' performances. The poor performance of a government in one jurisdiction might be attributed to a number of factors including corruption, but as the number of observations made by the populations of jurisdictions increases, the inferences they make about the causes of a particular jurisdiction's poor performance become increasingly precise. This idea that the performance of local jurisdictional governments can be evaluated by cross jurisdictional comparisons was first proposed by Besley and Case (1995). They compared the behaviors of state governors and found that if a particular policy was adopted by one governor, and deemed to be successful, it was rapidly copied by his 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". The attractiveness of this explanation for governmental behavior is that it is very parsimonious with information; to make evaluations very little detailed knowledge is required. The population simply compares outcomes across jurisdictions, and draws simple statistical inferences about politician-bureaucrats' levels of competence or objectives.

In our analysis we modify the "yardstick competition" approach and show that it can be fruitfully applied as an explanation for the decentralization/corruption relationship. Further, we provide additional empirical support for the predictions of the theory.

A key role of local governments is to collect taxes and supply to their constituency's goods or services such as roads, sewers, parks, garbage collection and the like. The process by which tax revenues are transformed into the supply of various publicly provided goods might be envisaged as a stochastic one, where the final supply of the good is subject to both jurisdiction specific and common random shocks. The failure of one jurisdiction to produce an adequate supply of goods and services from tax revenue may be due to a poor draw from the distributions of the relevant random variables, or, it may be due to corruption. The more observations are made across jurisdictions, the better the inferences that may be made over the probability that poor performance was due to corruption. If the populations in jurisdictions are able to better detect corruption either politician-bureaucrats will refrain from this activity, or, corrupt politician-bureaucrats will more frequently be removed from office. Hence the observed decentralization/corruption relationship arises.

However this immediately raises other questions, if greater decentralization reduces corruption why not decentralize further? Where should the process stop and why? It seems likely that the answer lies in the technology of service production. Scale economies can be realized if some goods

¹One 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.

are produced by larger jurisdictional units. Indeed scale economies are often the justification for removing the provision of some goods and services from the private sector. Hence the loss of productive efficiency may have to be traded off against the reduction in corruption. The optimal level of decentralization then follows from equating the marginal impacts of these countervailing effects.

2 Model.

We begin with a very simple model with constant marginal costs of public service provision. The objective here is to develop the inference problem on which our arguments are based, and establish the basic corruption/decentralization relationship. In section 3 we introduce decreasing costs and explore optimal decentralization.

The economy is assumed to be divided into $i = 1, \dots, n$ jurisdictions each with a population of size $\frac{m}{n}$. In each jurisdiction the local government supplies a service $z_i(t) = \{0, 1\}$ to every individual in the population, and funds this provision by levying the per person tax of $\tau_i(t)$. The production of the service is given by a stochastic production technology

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

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

We assume that

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

and

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

Each jurisdiction is run by a politician-bureaucrat drawn from a local population in which fractions λ_i are honest and $1 - \lambda_i$ dishonest respectively. We assume that the jurisdictions are ordered such that $\lambda_i \geq \lambda_j$ if $i > j$. Each politician-bureaucrat lives for two periods and serves in each period at the discretion of the local population. The politician-bureaucrats are paid a wage $w(t)$ for each period they retain office, and can, if they choose, supplement this wage by stealing tax revenues $r_i(t)$. politician-bureaucrats that are honest never steal, those that are dishonest do so if this raises their expected income. The expected utility of a politician-bureaucrat in jurisdiction i is given by

$$V_i(t) = w_i(t) + r_i(t) + \delta E[w_i(t+1) + r_i(t+1)], \quad (4)$$

where $0 < \delta < 1$ is a discount rate. Initially for purposes of expositional simplicity we assume $w_i(t) = w_i(t+1) = 0$, giving

$$V_i(t) = r_i(t) + \delta E[r_i(t+1)]. \quad (5)$$

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.

The population of each jurisdiction have just one decision to make, whether to retain a politician-bureaucrat to run the jurisdiction in the second period of the politician-bureaucrats life or replace

them with a fresh draw from the population. We assume the population 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 $x_i(t) = x_i(t+1) = 0$ giving

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

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

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

Notice that the population in jurisdiction i observe 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 likley that their local government is corrupt they will decide to replace them with a new draw from the population of potential politician-bureaucrats. We address this next.

2.1 Replacing politician-bureaucrats.

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

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

or

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

where $\bar{\lambda}_i(t+1)$ is the public's 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.

2.2 Updating of Beliefs

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

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

where we adopt the notation $\Pr [i^h]$ for $\Pr [i \text{ is honest}]$ and $\Pr [i^d]$ for $\Pr [i \text{ is dishonest}]$.

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 revealing, partially-revealing and non-revealing respectively².

2.2.1 Revealing State

In the revealing state for jurisdiction i service production is observed, that is $s_i(t) = 1$, in which case the population immediately deduce $\mu(t) = \varepsilon_i(t) = 1$ and that the incumbent politician-bureaucrat-bureaucrat is honest. Hence updating involves

$$\bar{\lambda}_i(t+1 | s_i(t) = 1) = \Pr [i^h | s_i(t) = 1] = 1. \quad (12)$$

2.2.2 Partially-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, $s_j(t) = 1$ some $j \neq i$. Hence the population of jurisdiction i will immediately be able to deduce that for the 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)

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

2.2.3 Non-Revealing State

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

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

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

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

²The modeling strategy of choosing a multiplicative 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 conclusions that follow, but the cost in terms of algebra is quite large.

In the case of homogeneous jurisdictions, $\bar{\lambda}_j(t) = \bar{\lambda}_i(t) \forall i, j$, reduces to

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

It is not too difficult to show that

$$\frac{\prod_{j \neq i}^n [1 - \lambda_j(t)(1-p)] p(1-q) + q}{\prod_{j=1}^n [1 - \lambda_i(t)(1-p)] (1-q) + q} \rightarrow 1 \text{ as } n \rightarrow \infty \quad (17)$$

hence $\bar{\lambda}_i(t+1 | s_i(t) = 0 \forall i) \rightarrow \lambda_i(t)$ from below as $n \rightarrow \infty$. 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 $\bar{\lambda}_i$.

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

2.3 Decentralization and the Level of Corruption.

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

$$E[C(t)] = (1-p)(1-q) \sum_{i=1}^n \binom{m}{n} (1 - \lambda_i(t)). \quad (18)$$

This is the sum over all the jurisdictions of the probabilities that positive production is realized then is 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

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

holds, will choose to separate and take a new draw from the local population of potential politician-bureaucrats. Since separations only take place if the populations (correctly) believe that taking a new draw from the distribution of potential politician-bureaucrats 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 separations 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 then 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 information state the population in a jurisdiction i will update their beliefs appropriately using Bayes rule. In determining the relationship between the expected level of corruption and decentralization we need to analyze both the effects of a change in the number of jurisdictions, n , on the updating process in each information state, and its effect 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 revealing information state a politician-bureaucrat is seen to be honest $s_i(t) = 1$, here 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 partially-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 separation will occur. While the updating of $\lambda_i(t)$ is independent of n , the probability that this information state will occur and hence that this update occurs 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 partially-revealing information state occurring is increasing in n , therefore the likelihood of a separation occurring is increasing in n , ergo expected corruption is decreasing in n .
3. In the non-revealing information state an increase in n increases the probability the population in i attach 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 the proposition tells us that if direct cross-jurisdictional comparisons are required for politician-bureaucrats to experience separation, that is if they occur only in the partially-revealing information state, then corruption is non-decreasing 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 a jurisdiction i are just indifferent between separation from a politician-bureaucrat in both the partially-revealing and non-revealing information states, and separation in the partially-revealing information state only. At n^* expected corruption can (but is not guaranteed to) discontinuously increase in decentralization (see the appendix).

3 Optimal Decentralization and Corruption.

One natural assumption is that the service is provided by the local government because it is subject to decreasing costs, and hence if supplied by the private sector would be a natural monopoly. If we assume that the cost of producing of a unit of the service is increasing in the number of jurisdictions then there can be a trade off between corruption and production efficiency and hence an optimal

level of decentralization. For the purposes of analyzing this issue we shall assume all jurisdictions are identical. Suppose now that the tax cost per person of providing each with $z_i(t) = \{0, 1\}$ units of the service is such that

$$\tau_i(t, n) = \frac{n}{N}. \quad (20)$$

Then the marginal increment to per person taxes as a consequence of adding a jurisdiction is

$$\Delta\tau_i = \frac{1}{N} \quad \forall i, n. \quad (21)$$

Where Δ is the one step ahead difference operator and N is a constant.

The expected marginal increment to per person service supply is given by

$$\begin{aligned} & \lambda_i(t)(1-p)(1-q) \Delta \left[1 - \left(\prod_{j \neq i}^n [1 - \lambda_j(t)(1-p)] (1-q) + q \right) \right] [1 - (1-p)(1-q)\lambda_i(t)] \\ = & \lambda_i(t)(1-p)(1-q) [1 - (1-p)(1-q)\lambda_i(t)] (1-q) \left[[1 - \lambda_j(t)(1-p)]^n - [1 - \lambda_j(t)(1-p)]^{n+1} \right] \\ = & \lambda_i(t)(1-p)(1-q) [1 - (1-p)(1-q)\lambda_i(t)] (1-q) [\lambda_j(t)(1-p)] [1 - \lambda_j(t)(1-p)]^n \end{aligned} \quad (22)$$

The first term in this expression $\lambda_i(t)(1-p)(1-q)$ is the expected service in the partially-revealing information state, the second term $[1 - (1-p)(1-q)\lambda_i(t)] (1-q) [\lambda_j(t)(1-p)] [1 - \lambda_j(t)(1-p)]^n$ is the increment to the probability that this information state will occur as the number of jurisdictions increases by one. It follows that the optimal level of decentralization $n^\#$ satisfies

$$\begin{aligned} & \lambda_i(t)(1-p)(1-q) [1 - (1-p)(1-q)\lambda_i(t)] (1-q) [\lambda_j(t)(1-p)] [1 - \lambda_j(t)(1-p)]^{n^\#} \geq \frac{1}{N} \\ & > \lambda_i(t)(1-p)(1-q) [1 - (1-p)(1-q)\lambda_i(t)] (1-q) [\lambda_j(t)(1-p)] [1 - \lambda_j(t)(1-p)]^{n^\#+1} \end{aligned} \quad (23)$$

clearly there exist parameterizations of the model for which an $0 < n^\# < \infty$ exists.

4 Proximity.

Our analysis can be reinterpreted to give insight into the relationship between proximity and corruption. Consider the economy to be divided into informationally separate islands. Suppose that each island consists of n jurisdictions each of which shares the common productivity shock $\mu(t)$. These common shocks are uncorrelated across informational islands. The smaller are such islands, that is the smaller is n , the more proximate they are in an informational sense. We know from our previous analysis that expected corruption is decreasing in n , it follows immediately that proximity and corruption are positively related. Smaller islands are more likely to experience corruption in any jurisdiction hence the correlation of corruption across nearby jurisdictions.

5 Empirical Analysis

In this section we provide some empirical evidence that supports the theoretical relationship between decentralization and corruption derived above. Although several cross-country studies such as de Mello and Barenstein (2001), and Fisman and Gatti (2002a) find a negative relationship between decentralization and corruption, our empirical analysis advances the existing literature in a couple

of different ways. First, we avoid comparing different countries by examining cross-state corruption in the United States. Data on corruption, decentralization, and on such control variables as public sector wages across states are more comparable than those across countries. Second, we employ a 5-year panel of corruption to control for unobserved state characteristics.

5.1 Data

As our measure of corruption we use data from the Justice Department's "Report to Congress on the Activities and Operations of the Public Integrity Section." The report provides information on the number of federal, state, and local public officials convicted for crimes related to corruption by state. Since there are lags between the time of crime and the time of conviction, we use 5-year averages for the period 1987-2001 for the 48 contiguous U.S. states. Following Fisman and Gatti (2002b) we deflate the data on convictions by state population. We measure decentralization as the share of non-central government expenditure in total government expenditure, which is equal to total expenditure of local government divided by the total expenditure by all levels of government (state and local). This is one of the most widely used measures of decentralization in the literature. We use data from U.S Census Bureau's "Government Finances" for the years 1991, 1996, and 2001.

We include a number of control variables in our empirical analysis to minimize omitted variable bias. According to Fisman and Gatti (2002b) less developed states are likely to be more corrupt. Hence to control for the level of development across states we include 5-year averages of the level of real per capita gross state product. For the level of real per capita gross state product we use data from the Bureau of Economic Analysis and for population we use data from the U.S Census Bureau. Following Glaeser and Saks (2004) we also control for ethnic fractionalization. As Glaeser and Saks (2004) argue ethnic fractionalization is likely to increase corruption by decreasing the popular will to oppose corruption. We use data from the U.S Census Bureau's "Census 2000" to calculate the ethnic fractionalization index; defined as $1 - \sum s_i^2$, where s_i is population share of group i . Following "Census 2000" the ethnic categories included in the calculation of the ethnic fractionalization index are Hispanic, White, Black, Asian, American Indian and Eskimo, Hawaiian and Pacific Islander and Others. Finally, we control for government wages deflated by per capita gross state product. Low government wages are likely to provide an incentive for government officials to engage in corruption. 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. We use data from the Bureau of Labor Statistics for the years 1991, 1996, and 2001. The summary statistics for all variables are given in Table 1.

Table 1: **Summary Statistics**

Variable	Mean	Standard Deviation	Min	Max	N
Corruption	0.313	0.187	0.000	0.865	144
Decentralization	0.519	0.079	0.323	0.654	144
Neighbor Corruption	0.306	0.107	0.053	0.653	144
GSP (in 000)	27.086	5.153	17.216	43.071	144
Ethnic Fractionalization	0.354	0.155	0.068	0.661	144
Government Wage	1.250	0.110	0.956	1.486	144
Population (in 000000)	5.525	5.908	0.457	34.607	144

5.2 Results

Our empirical specification is as follows:

$$Corruption_{st} = \beta_0 + \beta_1 Decentralization_{st} + \beta_2 Neighbor\ Corruption_{st} + \beta_3 X_{st} + \epsilon_{it} \quad (24)$$

where $Corruption_{st}$ represents the corruption in state s during year t , $Neighbor\ Corruption_{st}$ represents the average corruption in the neighboring states, $Decentralization_{st}$ represents decentralization, X_{st} represents the set of other state characteristics that affect corruption (government wages $Government\ Wage_{st}$, per capita gross state product GSP_{st} , ethnic fractionalization $Ethnic\ Fractionalization_{st}$, and population $Population_{st}$) and ϵ_{it} represents the error term.

Table 2: Pooled OLS and RE Estimates

	Pooled OLS	Random Effects
Constant	1.128* (0.396)	0.802** (0.397)
Decentralization	-0.623* (0.202)	-0.516*** (0.291)
Neighbor Corruption	-0.451* (0.147)	0.344** (0.144)
GSP	-0.012* (0.004)	0.009** (0.004)
Ethnic Fractionalization	0.375* (0.133)	0.308*** (0.179)
Government Wage	-0.375** (0.191)	-0.161 (0.213)
Population	0.007** (0.003)	0.006 (0.005)
N	144	144
R ²	0.22	

Dependent Variable: Corruption

*, **, and *** denote significance at 1%, 5%, and 10%, respectively.

Standard errors are given in parentheses.

The results of the Pooled Ordinary Least Squares (OLS) Estimation are given in the first column of Table 1. The estimated coefficient of $Decentralization_{st}$ is negative and highly significant, indicating that higher decentralization reduces the level of corruption. The estimated coefficient of $Neighbor\ Corruption_{st}$ deserves particular attention. None of the cross country studies examine the effects of corruption in neighboring countries. Fisman and Gatti (2002b) uses corruption in the neighboring states in their cross-state study as a control variable, but their results are not statistically significant. The estimated coefficient of $Neighbor\ Corruption_{st}$ in the Pooled OLS Estimation is positive and highly significant indicating the presence of "yardstick competition". In other words, the level of corruption in a state decreases as it decreases in the neighboring states. As far as the estimated coefficients of the control variables are concerned, all of them are highly significant and have the expected signs. Parallel to the existing literature, estimated

coefficients of GSP_{st} and $Government\ Wage_{st}$ are negative and the estimated coefficients of $Ethnic\ Fractionalization_{st}$ and $Population_{st}$ are positive.

The second column of Table 1 gives the results of the Random Effects (RE) Estimation³. Although the estimated coefficients of some of the control variables such as $Government\ Wage_{st}$ and $Population_{st}$ lose their significance, the other control variables are highly significant, and the estimated coefficients of all of the control variables have the expected signs. The estimated coefficients of $Decentralization_{st}$ and $Neighbor\ Corruption_{st}$, are both highly significant have the expected signs.

5.3 Robustness of the Results

The main robustness issue is whether the results are due to endogeneity. Corrupt governments are likely to oppose fiscal decentralization as it decreases their ability to steal. Following Arikan (2004), to correct for possible endogeneity, we instrument for decentralization with the surface area of the state in square miles. A good instrument is a variable that is supposed to be uncorrelated with the error term but correlated with the endogenous variable $Decentralization_{st}$. As Arikan (2004) argues the surface area is not likely to explain the level of corruption and it is a variable that is often used as an explanatory variable for decentralization.

Table 3: **2SLS and 2SRE Estimates**

	Pooled OLS	Random Effects
Constant	2.377** (1.102)	1.553 (0.948)
Decentralization	-2.161*** (1.286)	-1.999 (1.478)
Neighbor Corruption	0.354*** (0.182)	0.273*** (0.159)
GSP	-0.018* (0.006)	0.013*** (0.007)
Ethnic Fractionalization	0.624** (0.248)	0.481*** (0.293)
Government Wage	-0.693** (0.329)	-0.143 (0.251)
Population	0.016*** (0.008)	0.014 (0.009)
N	144	144

Dependent Variable: Corruption

*, **, and *** denote significance at 1%, 5%, and 10%, respectively.

Standard errors are given in parentheses.

The results of the 2 Stage Least Squares (2SLS) Estimation are given in the first column of Table 2. Although the magnitude of the estimated coefficient of $Decentralization_{st}$ is bigger, it remains negative and highly significant. However, estimated coefficient of $Decentralization_{st}$ loses its significance when it is estimated by 2 Stage Random Effects (2SRE). Nevertheless, it too is negative. The results of the 2SRE Estimation are given in the second column of Table 2. The

³Hausman Test suggests Random Effects Estimation.

estimated coefficient of $Neighbor\ Corruption_{st}$ is positive and highly significant in both the 2SLS and 2SRE estimations. The magnitude of the estimated coefficient, on the other hand, is smaller. As far as the estimated coefficients of the control variables are concerned, $Government\ Wage_{st}$ and $Population_{st}$ lose their significance when they are estimated by 2SRE. All of the estimated coefficients of the control variables are highly significant and have the expected signs when they are estimated by 2SLS.

6 Conclusion

We present a model which uses the yardstick competition approach proposed by 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. If the public service is supplied under conditions of decreasing costs, then we are able to define an optimal level of decentralization, and hence an optimal level of corruption. Further, the model can easily be reinterpreted to predict that jurisdictions close together will tend to experience similar levels of corruption. The empirical evidence that we provide supports the results that come out of the model. Using data for the United States, for the period 1987-2001, we find the level of corruption in a state is negatively related with decentralization and positively related to the level of corruption in the neighboring states.

7 Appendix

7.1 Derivation of (13)

We have

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

since

$$\begin{aligned} & \Pr [i^h \mid s_i(t) = 0 \cap \mu(t) = 1] \Pr [s_i(t) = 0 \mid \mu(t) = 1] \Pr [\mu(t) = 1] \\ = & \Pr [s_i(t) = 0 \mid i^h \cap \mu(t) = 1] \Pr [i^h \mid \mu(t) = 1] \Pr [\mu(t) = 1] \end{aligned}$$

it follows that

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

7.2 Derivation of (15).

We have

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

so

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

and

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

So

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

also

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

So finally

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

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

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

7.3 Proof of Proposition 1

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

Lemma 2 *The frequencies at which the three information states arise are given by*

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

Lemma 3 *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 4 *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

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

as required. ■

Lemma 5 *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

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

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

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

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 arrives. 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

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

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 $\frac{\prod_{j \neq i}^n [1 - \lambda_j(t)(1-p)] p(1-q) + q}{\prod_{j=1}^n [1 - \lambda_j(t)(1-p)] (1-q) + q}$ 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. ■

References

- Arikan, Gulsun G. (2004). "Fiscal Decentralization: A Remedy for Corruption?" *International Tax and Public Finance* 11: 175-195.
- Besley, Timothy and Anne Case (1995). "Incumbent Behavior: Vote Seeking, Tax Setting, and Yardstick Competition." *American Economic Review* 85: 25-45
- Fisman Raymond and Roberta Gatti (2002a). "Decentralization and Corruption: Evidence Across Countries." *Journal of Public Economics* 83: 325-345.
- Fisman Raymond and Roberta Gatti (2002b). "Decentralization and Corruption: Evidence from U.S. Federal Transfer Programs." *Public Choice* 113: 25-35.
- Glaeser, Edward L. Rave E. Saks (2004). "Corruption in America." NBER Working Paper 10821.
- Goel, Rajeev K. and Daniel P. Rich (1989). "On the Economic Incentives for Taking Bribes." *Public Choice* 61: 269-275.
- de Mello, Luiz (2000). "Can Fiscal Decentralization Strengthen Social Capital?" IMF Working Paper WP/00/129.
- de Mello Luiz and Matias Barenstein (2001). "Fiscal Decentralization and Governance: A Cross Country Analysis." IMF Working Paper WP/01/71.
- Treisman, Daniel (2000). "Decentralization and the Quality of Government." *Mimeo*.
- van Rijckeghem, Caroline and Beatrice Weder (2001). "Bueaucratic Corruption and te Rate of Temptation: Do Wages in the Civil Service Affect Corruption, and by How Much?" *Journal of Development Economics* 65: 307-331.