Essays in Political Economy and Public Economics

by

Ryan Booth

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

New York University

January, 2012

Dr. Alessandro Lizzeri

Acknowledgments

I am extremely grateful to my advisor, Alessandro Lizzeri, for all of the guidance and advice that he shared with me while I was working on this dissertation. I have learned a lot from Alessandro over the past several years, and his encouragement and input were instrumental in completing this research. This dissertation has benefited from helpful comments and suggestions from the following professors, whom I would like to thank: Oeindrila Dube, Guillaume Fréchette, Alessandro Gavazza, Sanford Gordon, William Greene, Nazgul Jenish, Nicola Persico and especially Shing-Yi Wang. I am also grateful for the help, feedback and friendship that I have received from my classmates Saki Bigio, Sean Flynn, Bernardo Santos da Silveira, Emanuel Vespa and Eduardo Zilberman. The C.V. Starr Center for Applied Economics provided financial assistance to acquire some of the data used in this project, and I am very appreciative of the Center's generosity.

The first chapter of this dissertation was written in collaboration with Emanuel Vespa, and the second chapter was written in collaboration with Deon Filmer and Jamele Rigolini. I am particularly thankful for having had the opportunity to work with Emanuel during our time at NYU. Both our research paper, as well as my other projects, have benefited greatly from the many conversations that Emanuel and I

have shared over these years.

Finally, I would like to thank my parents, William and Nellie, and my fiancée, Jordana, for their constant support during my pursuit of this degree.

Abstract

This dissertation investigates the political economy of government resource allocation and hiring decisions. Chapter one examines the causes of public expenditure persistence and documents such persistence for U.S. municipalities. We investigate four possible causes of this pattern and find the most support for the theory that an interest group induces persistence by influencing local politicians. A dynamic theoretical model is developed to derive quantitative and qualitative predictions about the magnitude of persistence. These conditions are tested using a panel of 595 U.S. municipalities, and the empirical analysis broadly supports the theoretical predictions. Estimates indicate that the median large municipality annually spends about $134 (4.8 percent) more per household due to expenditure adjustment frictions.

Chapter two investigates the countervailing incentives that electoral competition provides elected officials: while competitive elections may mitigate private rent seeking activities, competition may also promote opportunistic vote buying. We test for this trade-off using data from a decentralized public program in a developing country in which households were selected as eligible to receive social assistance benefits intended for the poor. Estimates indicate that while the overall amount of misallocation is only weakly related to electoral competition, the composition of the misallocation

varies: as electoral competition increases, officials' private rent seeking (i.e., allocating benefits to non-poor family members) decreases relative to vote buying (i.e., allocating benefits to non-poor, non-family members).

Chapter three studies the performance and selection of appointed bureaucrats. The chapter utilizes data from a developing country that include information about appointees' job performance and ability, as well as information about candidate appointees that were not selected. These data are used to examine the types of individuals that effectively administer the policy task, and the types of candidates that public officials actually appoint. While candidates with relatively more human capital are selected in general, public officials exhibit a tendency to appoint their associates, and candidates that are "like them." Some evidence suggests that this pattern is due, in part, to favoritism for friends. The chapter also estimates the degree to which suboptimal hiring reduces program performance.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

1.1 Introduction

This paper documents public expenditure persistence in US municipalities and investigates four possible causes of this persistence. We utilize a panel of 595 US municipalities from 1970 to 2000 and present evidence that spending on local public goods is slow to adjust to changes in public demand, with the median large municipality spending about $134 (4.8%) more per household due to this persistence. We find the most support for the theory that persistence is driven by an interest group (local public workers) that offers campaign support to politicians to increase public spending. A dynamic theoretical model is then developed to derive quantitative and qualitative predictions, which are supported by the data.

Previous research has documented persistence for several types of public policies: classic examples include the resilience of farm subsidies in developed economies (Anderson, 2010) and protective trade barriers for previously infant industries in developing countries (Krueger, 1993). With respect to public good provision, several mechanisms may potentially explain policy persistence, such as politically influential special interests, adjustment costs and evolving political preferences of the median

voter. Research that attempts to understand the relative importance of these types of persistence mechanisms, however, is relatively scarce.

There are several reasons why US municipalities are a good candidate to study the causes of policy persistence. First, US municipalities exhibit considerable variability of demographic and spending dynamics, both in the cross-section and time series, while sharing a common national system of laws and, to some extent, culture. Second, the composition of local public goods is relatively homogeneous across medium-to-large municipalities. And third, data are available for a large number of observations.

We use such data to show that municipalities that experience negative (positive) shocks to population spend significantly more (less) on public goods, after controlling for contemporaneous factors that determine constituent demand, such as crime and demographic characteristics. Given this pattern, we then empirically investigate four theoretical mechanisms that might drive this persistence. First, we consider the role that labor adjustment costs, as reviewed by Hamermesh and Pfann (1996), have on changes to municipal spending. Such a mechanism would imply that municipalities are slow to adjust their factor inputs because of costs to laying off workers.

Second, we consider how changing political preferences of voters might influence spending. For instance, Bassetto and McGranahan (2009) consider a model in which public investment hinges crucially on the mobility of constituents, as a forward-looking voter may have different preferences over spending depending on the probability that she remains living in the jurisdiction. Third, we consider the effect that residential sorting, as first studied by Tiebout (1956), might have on public expenditures. Households with a low marginal benefit for public goods, for instance, may endogenously migrate to suburbs over time, whereas public good-loving households may opt to reside in large, declining municipalities that provide more public goods.

Finally, we study the theory that a well-organized interest group might act to influence spending, as studied in Coate and Morris (1999). In particular, we investigate the role that public workers have in influencing policy through political support of local politicians.[1] We utilize information on the political activity of municipal workers that is available for a small subset of the cities in our sample to analyze whether this political activity explains the aforementioned persistence. While tests for the first three mechanisms suggest that these theories do not account for much of the persistence, empirical analysis of the fourth mechanism suggests that politically active workers are associated with higher spending in declining cities.

Because the number of respondents to the political activity survey was relatively small, however, the empirical conclusion regarding this mechanism is not definitive. In order to further explore this issue, we thus develop a theoretical model to draw subsidiary conclusions about the dynamic interaction between local politicians and an interest group that we can then test using our full data set. The model is similar in spirit to Coate and Morris (1999) and the static model of Grossman and Helpman (1994), in which a politician sets a policy variable (public spending) each period that affects the current welfare of both constituents and the special interest group. The politician, however, cares about both public welfare and campaign support from the interest group, which allows for the interest group to offer campaign support to the politician in exchange for policy concessions. Crucially, the *current policy* also augments the interest group's *future capacity* to influence the politician, introducing an endogenous link through time.

While the model is designed to deliver the straightforward relationship between

[1]The Hatch Act of 1939 restricted federal employees' political activities but only applies to municipal employees that work on federally financed projects. Moreover, the Act does not restrict public employees from electioneering, contributing personal money to campaigns or issuing endorsements.

population shocks and public spending, it also generates several ancillary empirical predictions that observables should satisfy if public workers are a primary contributor to spending persistence. We also calibrate the model to simulate the effects of demographic shocks on public spending, which suggests that persistence should be on the order of 2.5-5% higher in "declining" cities, relative to "growing" cities, depending on the specific type of shock.

Armed with these predictions, we then revisit our empirical analysis to test the quantitative and qualitative theoretical results. Quantitatively, our simulated interest group model is roughly consistent with the spending persistence implied by our coefficient estimates. We also disaggregate total municipal spending to separately study five different types of public goods. Our political activity data indicate that police officers and fire fighters are considerably more politically active than some types of local public workers, such as park maintenance personnel.[2] Consequently, the interest group mechanism would predict that spending persistence should be particularly strong for law enforcement and fire protection, and relatively less for public goods produced by workers that are less prone to engage in political activity. The empirical analysis strongly supports this intuition, as well as the other qualitative theoretical predictions.

Finally, we conclude the paper by discussing policy implications and estimating the amount of additional spending in the 25 largest US cities that is due to the persistence mechanism, using data from the 2010 Census. Estimates indicate that of these large municipalities, the median spends around 4.8% more on the public sector

[2]Teachers are also often thought of as a strong interest group. We do not study public education, however, for two reasons. First, in most municipalities, the school budget is set by the local board of education, which is distinct from the city government. Second, and more practically, school district boundaries often do not match up directly with municipal boundaries. Consequently, the relevant demographic information for school districts and municipalities often do not coincide.

wage bill due to policy persistence, with a fifth of the cities spending more than 7.5% more. In particular, we estimate that municipal spending in Detroit is about $400 higher per household due to policy persistence. We also devote particular attention to Vallejo, CA, which declared bankruptcy in 2008, and Harrisburg, PA, which received bailout funds from the state in 2010 to avert a default. For Harrisburg, our estimates indicate that the value of the additional municipal spending due to policy persistence is about 56% of the value of the bailout that was transferred to the city.

Section 1.2 presents our data and several stylized facts regarding spending persistence, and Section 1.3 contains our theoretical analysis and simulation exercise. Section 1.4 then revisits our empirical analysis in light of the theory, and Section 1.5 concludes.

1.2 Data and empirical evidence

1.2.1 Data

To investigate persistence in municipal spending, we collect data from the US Census Bureau's Census of Governments (CoG) and County and City Data Book (CCDB). Appendix A includes a detailed description of our data; a short summary of key points is presented here.

The CoG is conducted every five years and contains detailed financial information for all US government bodies, including cites. We extract information on operational expenditures, the annual wage bill paid to full time employees and capital expenditures, and adjust for inflation by converting the spending data to 2005 dollars. We collect CoG data from 1972 to 2002, where 1972 was the first year where the Census Bureau adopted a uniform accounting standard for constructing spending measures.

Our second main data source, the CCDB, includes demographic information for municipalities with population 25,000 or more. The CCDB compiles its data from the decennial US Census of Population and Housing, and demographic information is thus observed every 10 years. From these data, we extract a number of variables that we use as covariates, such as each city's population and median household income (adjusted to 2005 dollars). We collect CCDB information corresponding to decennial Census years from 1960 through 2000, and then merge each census year with the corresponding CoG reporting year (1970 CCDB is matched with 1972 CoG data, and so on). Data from 1960 is collected to compute demographic growth rates for the 1970 observations, as we discuss below.

Our data thus consists of a panel of observations with time dimension four, corresponding to the Census years 1970, 1980, 1990 and 2000. From 1960 to 2000, a number of municipalities fluctuated above and below the CCDB's reporting threshold of 25,000 residents.[3] To avoid potential sample selection bias, we thus define our population of interest as the set of cities that had 25,000 or more residents from 1960 to 2000.[4] For technical reasons discussed below, we also drop municipalities located in Hawaii and Alaska, which results in a final sample of 595 municipalities. Table 1.1 presents summary statistics for our sample.

1.2.2 Testing for policy persistence

We test for policy persistence by estimating the correlation between municipal spending and shocks to municipal population, which is a strong determining factor in de-

[3]For instance, the population of Key West, Florida, was 33,956 in 1960, 29,312 in 1970, 24,382 in 1980, 24,832 in 1990 and 25,478 in 2000.

[4]As a robustness exercise, we've conducted our analysis when defining our population of interest as cities with 25,000 or more residents across the entire 1970-2000 period. This expanded sample yields results that are consistent with the analysis presented below.

Table 1.1: *Summary statistics, pooled across time*

	Mean	Median	10th	90th
Per cap operational exp's ($)	1,027.64	798.74	432.19	1,911.52
Per cap full time wage bill ($)	618.12	464.17	274.15	1,242.14
Per cap capital exp's ($)	221.69	165.92	39.95	448.13
Population	140,087	63,364	32,992	256,050
Gross population growth rate	1.08	1.03	0.91	1.29
Median household income ($)	53,564	50,705	39,942	69,536
Gross median income growth rate	1.09	1.07	0.92	1.27
Unemployment rate (%)	5.72	5.22	2.82	9.30
Poverty rate (%)	12.96	12.36	5.30	21.55
College graduates (%)	18.39	15.70	7.65	32.28
School-age, 5-17 (%)	19.55	18.90	14.60	25.60
Elderly, 65+ (%)	12.32	12.20	7.60	16.84
Black (%)	14.40	7.60	0.46	39.64
White (%)	78.41	83.14	50.35	98.09
Inequality	1.58	1.52	1.30	1.92
Serious crimes / 100,000	6,181	5,810	2,333	10,572
Land area (1,000 sq ft per cap)	0.37	0.30	0.11	0.69
Change in boundary (%)	3.82	0.00	0.00	0.00
Vote share Democrat (%)	48.80	48.74	35.33	63.30
Stayers (%)	76.78	77.98	69.35	82.60

Notes. All variables are defined for the 595 sample municipalities, except per capita operational expenditures (566 municipalities) and per capita capital expenditures (584 municipalities).

mand for public good provision. Building on previous literature,[5] the linear model

$$\log(g_{it}) = \theta_1 \log(z_{it}) + \theta_2 \log\left(\frac{z_{it}}{z_{i,t-10}}\right) + X_{it}\beta + \delta_t + \alpha_i + \epsilon_{it} \qquad (1.1)$$

[5]Within the context of state-level spending, Bassetto and McGranahan (2009) examine the effect of population change on capital investment by estimating a pooled cross-section. We describe this paper in more detail below. When performing sensitivity analysis, Baqir (2002) also includes the population growth rate as a covariate and reports a negative relationship with per capita public spending, while Ladd (1994) examines county-level spending and estimates a positive relationship.

captures this correlation, where g_{it} represents a measure of expenditures in municipality i and year t, z_{it} represents municipal population and $z_{it}/z_{i,t-10}$ represent the (gross) population growth rates over the past decade. The vector X_{it} contains other covariates for municipality i, while α_i is a city-specific fixed effect and δ_t is a time-specific fixed effect. Finally, ϵ_{it} is an unobserved error term.[6]

Parameter θ_2 measures the degree of persistence in municipal spending, after controlling for contemporaneous factors that influence the demand of public good provision, $(z_{it}, X_{it}, \delta_t, \alpha_i)$. In principle, the level of public good provision should only depend directly on a municipality's *current characteristics*, and not on factors that influenced demand for public goods a decade prior. A negative parameter estimate ($\theta_2 < 0$) is suggestive of policy persistence: *holding current demographic characteristics fixed*, population decline is associated with greater municipal spending (and conversely for growing municipalities). An estimate of $\theta_2 = 0$, however, would indicate no inter-temporal relationship between current spending and past demand for public goods, whereas $\theta_2 > 0$ would point towards the opposite of persistence.

We incorporate the following variables in X_{it}, as is common in the literature: the log of median household income ($\log(y)$), the unemployment rate (*unem*), the poverty rate (*poverty*), the percentage of the adult population with a bachelor's degree (*college*), the percentage of the population between the ages of 5 and 17, inclusive (*sch_age*), the percentage of the population that is age 65 or older (*old*), the percentage

[6]Note that (1.1) is equivalent to

$$\log\left(\frac{g_{it}}{z_{it}}\right) = (\theta_1 + \theta_2 - 1)\log(z_{it}) - \theta_2 \log(z_{i,t-10}) + X_{it}\beta + \delta_t + \alpha_i + \epsilon_{it},$$

where g_{it}/z_{it} is a measure of municipality i's per capita expenditures. The coefficient $-\theta_2$ thus measures the elasticity of public spending with respect to lagged population. We adopt the specification in (1.1) for comparability with our theoretical analysis.

of white and black residents (*white* and *black*), and the log of the ratio of per capita income to median income, which serves as a proxy for income inequality ($\log(ineq)$).

Other time-varying characteristics that are often not considered in the literature may prove to be particularly important given our interest in capturing persistence. One obvious potential omitted variable is crime: a municipality's crime rate may influence spending on public goods, such as law enforcement and fire protection, as well as population change. For instance, an increase in crime may cause constituents to move from the municipality and prompt the local government to hire more police officers. Since several covariates, such as the unemployment rate and poverty rate, are associated with crime, model (1.1) will already capture some of this effect. Nevertheless, in order to minimize potential omitted variable bias, we deal with this issue directly by including the crime rate in X_{it}. Including a contemporaneous crime measure, however, would introduce a simultaneity problem. In fact, the literature on crime, such as Levitt (2002), has focused on estimating the impact of police expenditures on crime while controlling for the simultaneity bias. We follow an approach in the spirit of Corman and Mocan (2000). Because it takes time to significantly increase police expenditures in response to crime, our measure of crime is not a contemporaneous measure, but one that is lagged two years from t. The specific variable that we incorporate is (log) the number of serious crimes committed per 100,000 residents.[7]

A second important determinant of public spending is the population density within the municipality. Throughout our sample, some cities grow in land area by annexing or merging with neighboring communities. Moreover, a city that is shrinking

[7]The CCDB compiles crime information from the FBI's Uniform Crime Reports, which provides information on criminal offenses that are known to law enforcement personnel. The definition of serious crimes include both violent crimes (murder and non-negligent manslaughter, forcible rape, robbery and aggravated assault) and property crimes (burglary, larceny-theft and motor vehicle theft).

in size may require greater public expenditures to provide services to constituents that are more dispersed. To adjust for such an omitted variable, we incorporate two measures in X_{it}: a measure of population density (*land area* pc_{it}, municipal land (in square miles) divided by population), as well as an indicator that equals 1 if the city's land area expanded by more than 50% in the previous 10 years ($boundary_{it}$).

Finally, city-specific characteristics that are constant across time may simultaneously influence public spending and city demographics, but are often unobserved by the econometrician. In addition to city amenities, such as proximity to local beaches or national parks, municipal-specific institutions such as the size of the city's legislative body and electoral system are quite constant across the period that we study.[8] These type of unobserved characteristics will be controlled for by the city-specific fixed effect α_i.[9]

Table 1.2 reports coefficients estimates in model (1.1) for three alternative measures of municipal expenditures: total operational expenditures, the total full time wage bill, and total capital investment.[10] In the operational expenditure (Column (1)) and wage bill (Column (2)) specifications, the correlation between public spending and the rate of population change is negative and significant at the 1% and 5% level, respectively, and of similar magnitude. The population change coefficient in Column (1), for example, suggests that a one percent decrease in the (gross) popu-

[8]Baqir (2002) presents evidence that city political structures and institutions, such as the size of the city's legislative body and electoral system, are difficult to change and tend to persist through time. Evidence presented in Coate and Knight (2009) suggests that a municipality's specific form of government (mayor-council versus council-manager) tends to be quite stable. Between 1987 and 2002, for instance, less than 1% of cities in the authors' sample changed their form of government.

[9]Coate and Knight (2009) and other authors, such as MacDonald (2008), have utilized similar approaches when examining municipal spending using a panel of cities. Cross-section regressions for each of the four years indeed suggest the presence of time-constant unobserved variables. See Appendix B for cross-section estimation results, which vary considerably across years.

[10]Sample size various across different spending measures due to municipal financial reporting.

Table 1.2: *Fixed-effects estimates for municipal spending*

	[1] Operational	[2] Wage Bill	[3] Capital
$\log(z/z_{-10})$	-0.178***	-0.158**	0.134
	(0.050)	(0.050)	(0.175)
$\log(z)$	0.963***	0.900***	0.875***
	(0.050)	(0.048)	(0.162)
$\log(y)$	0.483***	0.450***	1.532***
	(0.127)	(0.119)	(0.439)
unem	-0.261	-0.937**	-7.011***
	(0.350)	(0.336)	(1.308)
poverty	0.262	-0.090	1.239
	(0.322)	(0.327)	(1.154)
college	-0.074	0.037	-0.383
	(0.232)	(0.215)	(0.728)
schage	-0.301	-0.264	3.735
	(0.579)	(0.520)	(1.921)
old	0.665	0.650	2.698
	(0.498)	(0.426)	(1.461)
black	0.323	0.188	-1.378
	(0.185)	(0.171)	(0.769)
white	-0.250	-0.329*	-1.432**
	(0.166)	(0.132)	(0.539)
log(ineq)	0.343*	0.229	-0.281
	(0.162)	(0.162)	(0.485)
log(crime)	0.021	0.007	0.112
	(0.018)	(0.020)	(0.092)
log(land area pc)	0.110**	0.082*	0.215
	(0.039)	(0.038)	(0.133)
boundary	0.019	0.006	-0.041
	(0.034)	(0.036)	(0.097)
Number of Cities	566	595	584

Notes. The dependent variable in each regression is the natural log of the variable listed at the top of the column. The estimation technique is the fixed effects within estimator. All regressions include time fixed effects, which are unreported. Heteroskedastic-robust standard errors are reported in parentheses. Star levels ***, ** and * correspond to statistical significance at the 1, 5 and 10 percent levels, respectively.

lation growth rate is associated with 0.18% more spending on public goods, holding contemporaneous municipal characteristics fixed. Column (3) in Table 1.2, on the other hand, suggests that capital investment does not exhibit policy persistence, as the coefficient for population change is highly insignificant.

In addition to population change, Columns (1) and (2) also show that population level and median household income are the two strongest correlates of public spending, as these coefficients are positive and significant at the 1%. It is worth noting that the coefficient for the (lagged) crime rate is statistically equal to zero across all models. As we will see below, however, the effect of crime will vary across different types of public goods when we disaggregate the spending measures. In particular, the crime rate will be significant and positively related to expenditures on law enforcement and fire protection services, while insignificance across other types of public goods.

Estimates in Columns (1) and (2) of Table 1.2 thus suggest that policy persistence does occur for municipal operational expenditures and, in particular, for wage outlays, but not for capital investment. For parsimony, we focus exclusively on studying persistence in the municipal wage bill for the remainder of the paper. We choose to focus on the wage bill because the sample is larger than municipalities reporting operational expenditures (595 versus 566). Moreover, the following results hold when examining operational expenditures.

1.2.3 Persistence mechanisms

Several theories may explain persistence in municipal spending. One appealing property of our estimation procedure, however, is that we can rule out explanations that are based on variables that do not change in time. In particular, theories of status quo bias such as the presence of veto players in the political process and superma-

jority requirements should be controlled for, given that municipal institutions and legislative procedures tend to persist in time. Consequently, we focus on four possible persistence mechanisms that rely on time-changing characteristics.

Labor adjustment costs

Adjustment costs are often cited as a reason why some firms are slow to modify factor inputs in response to shocks, as reviewed by Hamermesh and Pfann (1996). A similar mechanism may explain the persistence in the wage bill that is highlighted above. Arguably the most prominent sourse of adjustment costs that some municipalities face are legal barriers to laying off public workers. In the US, labor laws governing public sector employment are established at the state level, and the strength of these laws vary considerably across time and states. In particular, states vary in whether or not public workers have the right to bargain collectively to negotiate terms of employment, including wages and the municipality's layoff policy (such as "last in, first out"). In the absence of such barriers, municipal governments would have an easier time reducing labor input in response to evolving demand for public goods.

A large literature, as surveyed by Freeman (1986), has investigated the effect of public sector collective bargaining on municipal finance. In particular, Freeman et al. (1989) provide evidence that police officers benefit from having the state-mandated *right* to bargain collectively, regardless of whether or not the workers actually create a formal union. To test the effect of legal adjustment costs on policy persistence, we follow Freeman et al. by constructing the binary variable

$$B_{it} = \begin{cases} 1 & \text{if public workers are guaranteed the right to bargain} \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

using the NBER Public Sector Collective Bargaining Law Data Set.[11]

Column (1) of Table 1.3 presents the estimation results of model (1.1) after incorporating B_{it} into X_{it}. The coefficient for B_{it} is positive and significant, with a point estimate of 0.11, indicating that municipalities spend about 11% more when public workers are given the right to bargain collectively, which is consistent with previous studies.

Preferences of the median voter

Changing political preferences of the median voter may also induce status quo bias and spending persistence through at least two channels. First, Democrats with a proclivity for public spending may reside in higher proportions in declining regions, whereas fiscally conservative Republicans may be more prone to move to growing cities in the South and West of the US. Consequently, voters' movement to new regions could endogenously drive the negative correlation between population growth and public expenditures.

Second, Bassetto and McGranahan (2009) investigate a mechanism in which voter mobility influences capital investment in US states. A key insight of their theoretical and empirical analysis is that states with high resident turnover may over-invest in capital. If investment benefits are front-loaded and the cost of servicing bonds is pushed into the future, then a voter's willingness to support public borrowing should be increasing in the probability that the voter will move to a new state in the near

[11]The NBER data set disaggregates the legal information by type of public good: police officers, fire fighters, teachers and other local public workers. This disaggregation is performed because some states extend collective bargaining rights to only some types of workers. In general, however, all local public workers enjoy symmetric rights within the same state. In this section, we use the definition of police bargaining power in our analysis. In Section 1.4, we estimate our model using different types of public goods and match the appropriate B_{it} for each good.

Table 1.3: *Testing for persistence mechanisms*

	[1] Adjustment cost	[2] Median voter	[3] Tiebout	[4] All
$\log(z/z_{-10})$	-0.147**	-0.177***	-0.129**	-0.139***
	(0.051)	(0.051)	(0.052)	(0.046)
B	0.113***			0.104***
	(0.024)			(0.052)
Share Dem. Vote		0.171		0.280**
		(0.111)		(0.128)
stayers		0.917***		0.799***
		(0.276)		(0.425)
ρ			0.235***	0.210***
			(0.007)	(0.006)
$\log(z)$	0.916***	0.868***	0.818***	0.814***
	(0.049)	(0.050)	(0.046)	(0.307)
$\log(y)$	0.532***	0.474***	0.232*	0.267**
	(0.118)	(0.119)	(0.128)	(0.197)
unem	-0.857*	-0.935**	-0.877**	-0.878**
	(0.334)	(0.340)	(0.425)	(0.482)
poverty	-0.069	0.082	0.157	0.266
	(0.329)	(0.333)	(0.307)	(0.360)
college	-0.011	0.008	0.328*	0.298
	(0.214)	(0.219)	(0.197)	(0.186)
schage	-0.212	-0.546	-0.244	-0.344
	(0.518)	(0.529)	(0.482)	(0.157)
old	0.781	0.511	0.703*	0.723**
	(0.434)	(0.409)	(0.360)	(0.128)
black	0.144	0.267	0.046	0.069
	(0.167)	(0.172)	(0.186)	(0.018)
white	-0.428**	-0.254	-0.240	-0.242
	(0.133)	(0.134)	(0.157)	(0.036)
log(ineq)	0.282	0.258	0.022	0.055
	(0.167)	(0.163)	(0.128)	(0.030)
log(crime)	0.003	0.006	-0.003	-0.004
	(0.020)	(0.020)	(0.018)	(0.007)
log(land area pc)	0.079*	0.080*	0.047	0.047
	(0.037)	(0.038)	(0.036)	(0.036)
boundary	0.004	0.012	0.019	0.022
	(0.036)	(0.035)	(0.030)	(0.029)
Number of Cities	595	595	595	595

Notes. The dependent variable is the natural log of the wage bill paid to full time municipal employees. All regressions include time fixed effects, which are unreported. Columns [1] and [2] report coefficients from the fixed effects within estimator. Columns [3] and [4] report coefficients from the Spatial Durbin Model, and thus include coefficients for $\omega_i'\mathbf{F_t}$, which are unreported. Heteroskedastic-robust standard errors are reported in parentheses. Star levels ***, ** and * correspond to statistical significance at the 1, 5 and 10 percent levels, respectively.

future.[12]

A similar pattern may also occur for municipal spending: if a voter plans to remain in the city with high probability, then the voter may demand more, or perhaps less, public good provision. For instance, a voter with a high likelihood of remaining in a municipality may have a higher marginal benefit for a safe community and well-functioning local public sector. On the other hand, the voter might have a lower marginal benefit if there are expensive long-term costs associated with hiring public workers, such as expensive labor contracts or future pension liabilities.

To control for such preferences, we create two variables using available county-level data.[13] First, we collect voting data from the US Presidential election that directly preceded the year in which the municipality's budget was spent. We then create the variable dem_share_{it}, which divides the number of votes for the Democratic candidate by the total votes for the Democratic and Republican candidates. Second, we follow Bassetto and McGranahan by constructing a variable $stayers_{it}$, which proxies for the fraction of individual households that remain (or "stay") living in municipality i at time t. The decennial US Census asks respondents whether they changed their place of residency in the past five years and, if so, their county of residence five years prior. The variable $stayers_{it}$ is constructed by computing the number of county i's year t residents that lived in county i in $t-5$, and then dividing this number by the total number of residents that lived in county i in year $t-5$. The variable $stayers_{it}$ thus

[12]It is important to note that voter turnover is distinct from (net) population change. For instance, many cities in Texas have a very high population growth rate and a low rate of resident turnover: once moving to a city, a household is very likely to remain in that city for many years. On the other hand, there are growing cities such as Alexandria, VA that exhibit high resident turnover: households quickly move in and out of Alexandria as the political landscape and job market shifts in nearby Washington, DC following national elections.

[13]See Appendix A for more information about the raw data and construction of the variables. We utilize the Census Bureau's 2002 mapping that matches each city to one county.

offers an estimate of the probability that a randomly selected resident will remain living in municipality i, five years hence.

Column (2) of Table 1.3 presents parameter estimates of model (1.1) when we incorporate $stayers_{it}$ and dem_share_{it} into X_{it}. Of course, given that $stayers_{it}$ and dem_share_{it} are defined at the county level, these variables exhibit some measurement error, so the reader should bear in mind that the coefficient estimates may exhibit attenuation bias. Indeed, the parameter for dem_share_{it} is not significantly different from zero, but $stayers_{it}$ is positive and significant, indicating that the wage bill is higher when residents are more likely to remain living in the municipality.[14] The parameter θ_2, however, remains essentially unchanged after accounting for $stayers_{it}$ and dem_share_{it}.

Tiebout sorting

Tiebout sorting, in which households sort into communities based on preferences over the level of public good provision, may be a third mechanism causing the pattern observed in Table 1.2. Rhode and Strumpf (2003), for instance, document that the cost of moving has declined substantially in the US over the past century, which should lend itself to an increased prevalence of Tiebout sorting.[15] Moreover, continued investment in public transportation and highway systems have made commuting within metropolitan areas much easier for households that reside away from the city center.

The Tiebout hypothesis could drive the observed persistence if, say, households

[14] Bassetto and McGranahan also examine state-level operational expenditures and estimate θ_2 statistically equal to zero, and the coefficient on $stayers_{it}$ strictly negative. That paper, however, estimates a pooled cross-section and omits the population level, $\log(z_{it})$, from the model. We replicate their qualitative results when estimating their model with our data.

[15] Unlike much of the Tiebout literature, however, Rhode and Strumpf find little empirical evidence that Tiebout sorting is a primary contributor to migration patterns in the US over the past 150 years.

that value public goods have remained living in large city centers that heavily invest in the public sector, while households with less preference for public goods have sorted to growing suburbs outside of the main city. Preferences for local public goods may only be weakly related to political leanings in Presidential elections, and thus the controls presented in the electoral test might not fully capture the Tiebout effect described here.

Addressing this mechanism is somewhat more complicated than the two cases discussed above. In particular, there are omitted variables stemming from this Tiebout effect that are correlated with g_{it}, as well as contemporary controls (z_{it}, X_{it}). To address this mechanism, we begin by first identifying these omitted variables, and then modifying (1.1) to account for the omission.

Tiebout sorting is primarily thought of as a "local" phenomenon, in which households sort within a metropolitan area. Indeed, some evidence has been presented that the Tiebout mechanism is not a strong factor in *migration between metropolitan areas*. A demography literature, as surveyed in Greenwood (1997), references private sector employment as the main driving force behind inter-metropolitan migration, while Day and Winer (2006) provide evidence that the impact of public policies on migration between Canadian provinces is small. In short, the evidence suggests that the main determinants of *long distance migration* are not strongly related to the quality and cost of local public goods.

Tiebout sorting more likely occurs to some extent at the local level, *within a metropolitan area*. Following McFadden (1978), consider a two-stage model of residential choice: first, an agent decides which metropolitan area to work in, and then, the agent decides which municipality to reside within the metropolitan area. Within this hierarchical decision model, the agent selects the optimal municipality by com-

paring all municipalities within the metropolitan area across K municipal-specific characteristics. Thus, a household's location decision to reside in municipal i will depend upon both the demographic characteristics of that municipality, as well as the demographic characteristics of the municipalities $i' \neq i$ that are located in the metropolitan area.

In short, the relevant omitted variables in (1.1) involve a vector of attributes of neighboring communities. Consequently, we modify (1.1) to include spatial effects that constitute these omitted variables. The resulting estimation strategy is essentially the fixed effects estimator, but adding in additional covariates to control for characteristics in nearby cities. Let

$$\mathbf{F_t} = \left(\ln(\mathbf{z_t}), \ln\left(\frac{\mathbf{z_t}}{\mathbf{z_{t-10}}}\right), \mathbf{X_t} \right)$$

denote the stacked vectors of covariates for all I cities and let $\ln(\mathbf{g_t})$ denote the vector of public expenditure decisions, at time t. We incorporate $(\mathbf{F_t}, \ln(\mathbf{g_t}))$ into the right hand side of (1.1) and multiply $(\mathbf{F_t}, \ln(\mathbf{g_t}))$ by a weighting matrix Ω, where element $\omega_{i,i'}$ is a measure of geographic "closeness" between cities i and $i' \neq i$. For each i, the vector $\omega_\mathbf{i}$ stacks all weights associated with city i where the weight given to city i is zero. The following model thus incorporates for the omitted variable problem associated with Tiebout sorting:

$$\begin{aligned} \ln(g_{it}) = {} & \theta_1 \ln(z_{it}) + \theta_2 \ln\left(\frac{z_{it}}{z_{it-10}}\right) + X_{it}\beta + \delta_t + \alpha_i \\ & + \omega_\mathbf{i}' \mathbf{F_t} \zeta + \rho \omega_\mathbf{i}' \ln(\mathbf{g_t}) + \epsilon_{it}. \end{aligned} \tag{1.3}$$

As mentioned above, model (1.3) is simply the standard fixed effects model that

incorporates the characteristics of neighboring communities. This class of model is often referred to as a Spatial Durbin Model (SDM) in the spatial econometrics literature. The coefficient ρ captures the correlation between the level of public good provision in city i and the weighted vector of public good provision in all other cities, whereas the vector ζ has an analogous interpretation for the covariates. While we incorporate $(\mathbf{F_t}, \ln(\mathbf{g_t}))$ primarily to control for potential the omitted variables stemming from Tiebout sorting, there are other reasons for $(\rho, \zeta) \neq 0$. For instance, $\rho \neq 0$ could imply that public spending decisions in city i' may directly affect city i via spillovers.

As is extensively discussed in the literature, the estimation of a model with a spatial lag, $\ln(\mathbf{g_t})$, is endogenous by construction: $\ln(g_{i,t})$ depends directly on the error ϵ_{it}, and thus $\ln(\mathbf{g_t})$ is correlated with the vector of errors $\epsilon_{\mathbf{t}}$. While a straightforward maximum likelihood approach has been extensively utilized in the literature to overcome this problem in the cross-section, the technique to extend the SDM to panel data is rather recent. Elhorst (2003) first proposed a technique, but Anselin et al. (2006) demonstrated that this procedure leads to an incorrect estimation of standard errors. Lee and Yu (2010) overcome this problem by proposing a transformation that allows for a consistent estimation of coefficients and the correct computation of standard errors when a spatial lag appears. We utilize an extension of the Lee and Yu methodology that was developed in Beer and Riedl (2009). As far as we know, our paper is the first applied estimation of an SDM in the context of panel data using individual and time fixed effects.[16]

As is standard in the spatial econometrics literature, the spatial weighting matrix

[16]We thank Christian Beer and Aleksandra Riedl for sharing their codes with us.

Ω must be imposed prior to estimation. In this paper, we define

$$\omega_{i,i'} = \begin{cases} 1 & \text{if } i \neq i' \text{ and the distance between } i,\ i' \text{ is less than 50 miles} \\ 0 & \text{otherwise} \end{cases}$$

All results reported below are robust to other specifications of Ω that are typically employed, such as defining $\omega_{i,i'}$ as the inverse of the distance between cities i and i'.

Estimation results for the coefficients in (1.3) are reported in Column (3) of Table 1.3. Do to space considerations, we do not report estimates for (δ_t, ζ) and, of course, the incidental parameters α_i. Comparing across the fixed effects (Column (2) of Table 1.2) and SDM models (Column (3) of Table 1.3), many of the parameter estimates are quite similar. For instance, the effect of population change is negative and significant in both models, with point estimates of -0.158 and -0.129, respectively, which implies that policy persistence is still present through the population channel. In addition, parameter estimates for the effect of the population level (0.900 versus 0.818), the level of median income (0.450 versus 0.232), the unemployment rate (-0.937 versus -0.877), are all significant in the fixed effect and SDM models, respectively. Moreover, the spatial auto-regressive coefficient, ρ, equals 0.235 and is significant at the 1% level, suggesting that spatial effects are non-trivial.

Before moving on, it is important to note that an another appealing consequence of estimating model (1.3) is that the SDM shrinks omitted variable bias, as long as the omitted variable is spatially correlated, as discussed in LeSage and Pace (2009). This is appealing in our model, as many time-varying omitted variable will most likely be correlated across neighboring communities.

Column (4) of Table 1.3 jointly incorporates the variables associated with the

mechanisms.[17] The results are roughly similar to those discussed above. Nevertheless, the coefficient for the population growth rate remains significant and negative, suggesting that there are additional factors that contribute to a statistically significant portion of the persistence.

Politically active public workers

As mentioned above, authors such as Coate and Morris (1999) have investigated how politically active interest groups can cause policy persistence. Within the context of municipal public finance, local public workers are often cited as a strong interest group that can influence public spending by delivering campaign support to local politicians, via mechanisms such as door-to-door electioneering, official endorsements and campaign finance.

Substantial anecdotal evidence suggests that public workers do, indeed, wield such influence in local elections, which is then used to secure favorable terms of employment. For instance, an article in *The Economist* noted that public employees constitute, "powerful political machines [that] help pick the people that sit on the other side of the bargaining table... [especially in] local campaigns, where turnout is low" (January 8, 2011).

Such influence is perhaps most apparent in declining cities. An article in *Time* magazine profiled Detroit mayor David Bing shortly after being elected. The article stated,

> Detroit must reduce the size of its 42 government agencies to be proportionate to a city with a shrinking population and smaller coffers... a risky

[17] Given that, for a particular year t, $B_{\cdot,t}$ is constant across all municipalities within a state, we do not include $B_{i,t}$ within the matrix $\mathbf{F_t}$, as this would be redundant in the spatial setup.

> proposition in a region with historically strong unions. It's a strategy that's particularly dangerous in an election year.... "Changes that should have happened 20 years ago are now upon us," Bing (said). "Previous administrations had folks who were so concerned about getting elected, or re-elected. I'm not worried about that" (Gray, 2009).

Generous public employee pay packages have also been referenced in municipalities that have faced financial distress, such as Vallejo, CA. In 2008, Vallejo entered Chapter 9 bankruptcy as it fell behind on bond payments and racked up a $16 million deficit. Some commentators have blamed past municipal administrations for awarding police officers and fire fighters generous pay and benefit packages, which cost city millions of dollars (Jones, 2008).

While a large-scale data set documenting local public workers' electioneering activities is not available, the International City/County Management Association (ICMA) did conduct two small-scale surveys in 1988 and 1999 that collected such information. The ICMA questionnaire asked the Chief Administrative Officer of each municipality to "indicate whether employee unions/associations have engaged in the following political activities" in the prior decade, where the activities included candidate endorsements, candidate financial contributions and time/in-kind campaign contributions.[18]

Of the cities that responded to both ICMA surveys, 136 appear in our panel as well. Given the relatively small sample size, the fact that the ICMA asks a (potentially uninformed) respondent about retroactive information and possible issues regarding endogeneity, we would like to emphasize that the following empirical results are only suggestive. A pattern does emerge, however, when incorporating the political activity information into our analysis.

[18] See Appendix A for more details regarding these data.

Table 1.4: *Per capita wage bill, tabulated by population growth and political activity*

		Political Activity?		
		No	Yes	Total
Declining Population?	No	\$546.02	\$577.82	\$567.87
		(\$379.71)	(\$328.68)	(\$344.69)
		56	123	179
	Yes	\$559.76	\$636.50	\$601.84
		(\$397.74)	(\$638.30)	(\$541.67)
		42	51	93
	Total	\$551.91	\$595.02	\$579.48
		(\$385.57)	(\$441.20)	(\$421.79)
		98	174	272

Notes. This table presents summary statistics for the annual per capita wage bill paid to full time workers, tabulated across subsets of the data. The columns tabulate the sample by whether or not municipal workers were politically active, whereas the rows tabulate whether the municipal declined in population during the previous decade. Each cell reports average spending, the standard deviation (in parenthesis) and sample size within the subset.

Table 1.4 provides summary statistics regarding the per capita wage bill for several subsets of this reduced dataset. Each cell reports the mean per capita wage bill, the standard deviation (in parenthesis) and sample size within each subset. The columns of the table tabulate the sample by whether or not public workers engaged in political activity (endorsements, campaign finance and/or electioneering), as reported by the ICMA. The rows, on the other hand, tabulate the sample by whether or not the municipality's population declined during the previous decade. Several facts are apparent from the table: first, consistent with the previous results, cities with a declining population base spend about \$34 per capita on public goods, relative to growing cities (\$601.84 versus \$567.67). Similarly, a municipality with a politically active public workforce spends about \$43 more per capita, relative to municipalities with a nonactive workforce (\$595.02 versus \$551.91).

The more interesting result, however, is the cross-tabulation: if the public workforce is not politically active, then declining and growing municipalities differ in their per capita spending by only \$14 (\$559.76 versus \$546.02). On the other hand, when workers engage in the political process, then declining cities spend about \$59 more than growing cities (\$636.50 versus \$577.82).

Table B.2 in Appendix B reports fixed effects regression results when incorporating the other covariates. Column (1) incorporates the binary variable $decline_{it}$ in place of $\log(z_{it}/z_{it-10})$, where again $decline_{it}$ equals one if the municipality's population declined. Column (2) also includes $decline_{it}$, as well as the indicator $activity_{it}$ which equals 1 if public workers were politically active, and the interaction between these two variables. In Column (1), the coefficient for $decline_{it}$ equals 0.029, indicating that declining cities spend about 2.9% more on the wage bill than growing cities. In Column (2), however, the $decline_{it}$ coefficient equals -0.008, whereas the interaction equals 0.088. Given the small sample size, all covariates other than $\log(z_{it})$ are insignificant at 10%. Nevertheless, the interaction between $decline_{it}$ and $activity_{it}$ is significant at 12%.

As noted above, these results are only suggestive and could potentially be misleading because of potential endogeneity and measurement error; moreover, the size of the effects are statistically insignificant at the standard levels. Consequently, to further explore this political channel, we now turn to a theoretical model and simulation exercise to derive ancillary predictions that observables should satisfy if the interest group mechanism is, indeed, a strong driver of the persistence. We will then revisit the data and empirics in Section 1.4, in light of the theory.

1.3 A dynamic theory of interest groups

1.3.1 Model

We begin with a brief overview of the model. Time is discrete and indexed by $t = 0, 1, ...$, and there are two types of agents: an infinitely-lived association that represents public sector workers ("the union") and an infinitely-lived political party.

Each period, the political party has full discretion over selecting the amount of money that the municipality allocates to producing a public good, g_t. As is standard in the literature, we assume that the party's preferences are increasing in both the welfare of the city's constituents and the level of political campaign support delivered by the union to capture the key trade-off that politicians face. Constituent welfare depends upon both g_t, as well as a vector of stochastic, exogenous variables s_t that includes the current demographic profile of the municipality (described below).

Political support, on the other hand, is supplied (at a cost) by the union. We interpret this political support in a broad sense: the intensity with which the union engages in door-to-door campaigning; the mobilization of the union's membership to vote for the party; monetary payments to the party's campaign coffers; in short, actions that increase the party's political strength in local, state or national elections, but are costly for the union. Union preferences are strictly increasing in g_t, and the union can attempt to influence the party's selected budget allocation by offering a binding promise of campaign support in exchange for policy concessions. In addition, the union's payoff is dependent upon its status quo budget allocation, g_{t-1}, as larger unions control more resources for political campaigns and must defend a larger membership base from layoffs.

More specifically, the timing each t is as follows:

1. The state (g_{t-1}, s_t) is publicly observed.

2. The union offers the party political support (contingent on g_t).

3. The party selects g_t, and period payoffs are realized.

State variables

The exogenous state $s_t = (w_t, y_t, z_t) \in S \subset \Re^3_{++}$, where S is compact-valued and convex. The sequence $\{s_t\}_t$ evolves according to the AR1 process $s_{t+1} = \gamma s_t + \epsilon_{t+1}$, where $\gamma = (\gamma_w, \gamma_y, \gamma_z) \in \Re^3_+$ and $\epsilon_{t+1} \in \Re^3$ is distributed according to a continuously differentiable cumulative distribution function $\Gamma(\epsilon)$.[19]

The pair (y_t, z_t) is interpreted as the municipality's median constituent income and population, respectively, at time t. We explicitly incorporate median constituent income into our model because the empirics confirmed that income is an important determinant of public spending. The variable w_t denotes the public sector wage at t. A previous version of this paper allowed the public sector wage to be endogenously determined in a slightly different environment; the main insight of policy persistence in that version is consistent with what follows. We employ the exogeneity assumption for technical ease and implicitly assume that public sector wages are determined by external factors, such as public sector collective bargaining rights. Note that as a consequence, the union implicitly attempts to influence the *size of the public workforce*, given that w_t is exogenous.

The previous period's level of public expenditures, g_{t-1}, is also observed at the beginning of period t. The government must maintain a balanced budget each period,

[19] For simplicity, we abstract from cases where $\gamma s_t + \epsilon_{t+1} \notin S$, due to either drift or a large shock.

and thus $g_t \in [0, y_t z_t]$, where $y_t z_t$ denotes the municipality's tax base at t. We denote this constraint set as $G(s_t) := [0, y_t z_t]$, and define $G := [0, \max_{s \in S} yz]$.

The Union

The union discounts future utility at rate $\beta \in (0,1)$ and has period preferences that are represented by the expression $u(g, g_{-1}) - \phi(C)$, where $u : G^2 \to \Re$ is the payoff the union receives from its current and previous budget allocation, $C \in \Re_+$ is the level of political campaign support delivered to the party and $\phi : \Re_+ \to \Re_+$ is a cost function mapping campaign expenditures to a loss in union utility. The function $\phi(\cdot)$ satisfies $\phi' > 0$ and $\phi'' \geq 0$, whereas u is twice differentiable, strictly concave and satisfies $u_g > 0$.[20]

The union's payoff also depends on its status quo budget allocation, g_{-1}, for the following reasons: first, a larger status quo budget is associated with greater union membership and an increase in union dues. Consequently, a union with a larger status quo budget enjoys a higher capacity to engage in electioneering, as well as increased means to engage in social and civic activities. Consequently, we assume that $u_{g_{-1}} > 0$. Second, the status quo budget allocation should also influence the union's marginal benefit for current expenditures, u_g. As the union's membership increases, its members will demand a larger current budget allocation to guard against layoffs, and to ensure that union dues are being put to good use. Consequently, we assume that $u_{gg_{-1}} > 0$: as the union's status quo budget increases, the marginal benefit of current expenditures increases as well.

[20] Alternatively, g_{-1} can be incorporated explicitly into the union's marginal cost of campaigning. In previous versions of the paper, the union's payoff equaled $u(g) - \phi(C, g_{-1})$. In this setup, $\phi_{g_{-1}} < 0$ and $\phi_{g_{-1}C} < 0$, so that the union's total and marginal cost of campaigning were *strictly decreasing* in status quo size. Under restrictions on the evolution of $\{s_t\}_t$, the insights from such a model are consistent with this version.

Let $h_t = (g_{t-1}, s_t, h_{t-1})$ denote the history of play up to period t and $H_t(g_{-1}, s_0)$ denote the set of feasible histories at t, given initial state (g_{-1}, s_0). A strategy for the union is a sequence of campaign mappings $\hat{C}_t : H_t(g_{-1}, s_0) \to \Re_+$ and public good thresholds $\hat{g}_t : H_t(g_{-1}, s_0) \to G$ whereby the union promises to deliver $\hat{C}_t(h_t)$ political support to the party if $g_t \geq \hat{g}_t(h_t)$.

The political party

The political party also discounts future utility at rate $\beta \in (0, 1)$ and, as mentioned above, has preferences over public welfare and the level of political support delivered by the union. This reduced form approach is employed as a parsimonious way to capture the trade-off that politicians face between maximizing constituent welfare and catering to influential interest groups.

The city's representative citizen (or median voter) has preferences over private consumption, c, and the per capita production of the local public good, represented by the number of public workers per capita, l. This citizen's preferences are CES and represented by

$$\frac{\eta l^\sigma + (1-\eta)c^\sigma}{\sigma},$$

where $\eta \in (0, 1)$ is the relative weight that the representative citizen places on the public good. A large empirical literature in public finance has robustly documented that the price elasticity of demand for a wide variety of local public goods is inelastic. We incorporate this fact by restricting $\sigma < 0$, which is necessary and sufficient for constituent demand for the public good to be inelastic.

We normalize the price of the private consumption good to unity and assume

that the government generates revenue from a lump-sum tax. Consequently, the representative citizen faces a budget constraint of $c + wl \leq y$, which is satisfied with equality. By definition, aggregate public expenditures are $g = wlz$. We can thus write the representative citizen's indirect utility as

$$\frac{z^{-\sigma}}{\sigma}\left[\eta\left(\frac{g}{w}\right)^{\sigma} + (1-\eta)\Big(yz - g\Big)^{\sigma}\right]. \tag{1.4}$$

Total constituent welfare is then derived by multiplying (1.4) by population:

$$v(g,s) \quad = \quad \frac{z^{1-\sigma}}{\sigma}\left[\eta\left(\frac{g}{w}\right)^{\sigma} + (1-\eta)\Big(yz - g\Big)^{\sigma}\right].$$

For the remainder of the paper, we deal directly with $v(g,s)$. It is straightforward to show that v is strictly concave and single-peaked in g.

For future reference, it will be useful to highlight several properties of v. First, $v_{gw} > 0$ (by $\sigma < 0$): demand for the public good is inelastic in the public sector wage. Second, $v_{gy} > 0$: as the representative citizen's income rises, the marginal benefit of the public good increases as well. Finally, $v_{gz} > 0$: as city population grows, the marginal benefit of the public good increases as more people share the cost of the public good while simultaneously reducing per capita provision.

Given constituent welfare, we define the political party's period payoff as $\lambda v(g,s) + C$, where $\lambda > 0$ is the weight that the party places on public welfare relative to the political support delivered from the union, C. A strategy for the political party is a sequence of policy decisions $\hat{\hat{g}}_t : H_t(g_{-1}, s_0) \times G \times \Re_+ \to G$ mapping history h_t and union offer $(\hat{g}_t, \hat{C}_t)$ to a policy decision.

Equilibrium concept

We restrict attention to Markov equilibria. Markov strategies are defined in the usual manner, in which the union's strategy $(\hat{g}, \hat{C})$ depends only upon the current payoff-relevant state (g_{-1}, s), while the party's function $\hat{\hat{g}}$ depends only upon the current state (g_{-1}, s) and union offer (g', C'). In Appendix C, we show existence of a pair of value functions $U(g, s')$ (for the union) and $P(g, s')$ (for the political party) that are continuous in g. Consequently, given a state (g_{-1}, s), we can define the party's participation constraint, which requires that the union's offer (g^*, C^*) satisfy

$$\lambda v(g^*, s) \; + \; C^* \; + \; \beta \int P\Big(g^*, \gamma s + \epsilon\Big)\Gamma(\epsilon) \quad \geq$$
$$\underbrace{\max_{g \in G(s)} \left\{ \lambda v(g, s) \; + \; \beta \int P(g, \gamma s + \epsilon) d\Gamma(\epsilon) \right\}}_{\text{The Party's Outside Option}} \tag{1.5}$$

in order to incentivize the party to select g^*. We now define our equilibrium concept.

Definition. *A **Markov Perfect Equilibrium (MPE)** is a Markov strategy for the union $(g^*(g_{-1}, s), C^*(g_{-1}, s))$ and the political party $g^{**}(g_{-1}, s, g, C)$ such that:*

1. *For every (g_{-1}, s) and (g', C'), the party's strategy $g^{**}(\cdot)$ solves*

$$\max_{g \in G(s)} \left\{ \lambda v(g, s) \; + \; C' \cdot \mathbf{1}_{g \geq g'} \; + \; \beta \int P(g, \gamma s + \epsilon) d\Gamma(\epsilon) \right\}.$$

2. *For every (g_{-1}, s), there does not exist another strategy $(\hat{g}(\cdot), \hat{C}(\cdot))$ such that*

$$u(\hat{g}^{**}, g_{-1}) \; - \; \phi(\hat{C} \cdot \mathbf{1}_{\hat{g}^{**} \geq \hat{g}}) \; + \; \beta \int U(\hat{g}^{**}, \gamma s + \epsilon) d\Gamma(\epsilon) \quad >$$
$$u(g^{**}, g_{-1}) \; - \; \phi(C^* \cdot \mathbf{1}_{g^{**} \geq g^*}) \; + \; \beta \int U(g^{**}, \gamma s + \epsilon) d\Gamma(\epsilon)$$

where $\hat{g}^{**}$ and g^{**} are the party's best responses to $(\hat{g}(\cdot), \hat{C}(\cdot))$ and $(g^*(\cdot), C^*(\cdot))$, respectively.

This is the standard definition of an MPE: for any state and union offer, the political party's strategy must be optimal, while the union's offer must be optimal after accounting for the political party's equilibrium strategy.

1.3.2 The social planner: No politics

If public workers cannot exert influence on the political party, then the party solves the social planner's problem and maximizes public welfare. In this instance, the party chooses $\tilde{g}(s) = \arg\max_{g \in G(s)} v(g, s)$ each period, yielding $\tilde{v}(s) = v(\tilde{g}(s), s)$ welfare to constituents. Note that $\tilde{g}(s)$ does not depend on, g_{-1}, which only enters via the union's preferences. We mention two comments to compare with the political equilibrium:

Comment 1. *In the solution to the planner's problem, the level of public expenditures $\tilde{g}(s)$ does not depend on previous realizations of median income y_{-1} or population z_{-1}:*

$$\frac{d\tilde{g}}{dy_{-1}}(s) = 0 \qquad \textit{and} \qquad \frac{d\tilde{g}}{dz_{-1}}(s) = 0.$$

Conditional on the current state s, public spending has no relationship with previous demographic characteristics in the city, as expenditures fully adjust each period to maximize constituent preferences. The second comment addresses the relationship between public spending and the public sector wage:

Comment 2. *In the solution to the planner's problem, $\tilde{g}(s)$ is strictly increasing in*

the public sector wage:

$$\frac{\partial \tilde{g}}{\partial w}(s) \quad > \quad 0.$$

The inequality reflects the fact that citizen preferences are inelastic in the public good, so that higher public sector wages correspond to greater public expenditures.

1.3.3 Political equilibrium

We begin by establishing existence of an MPE in which the party always satisfies the union's demand for public spending, so that $g^{**}(g_{-1}, s, g^*, C) = g^*$.[21]

Lemma 1. *An MPE exists in which the party always satisfies the union demand.*

Our first proposition establishes that the equilibrium budget allocation $g^*(g_{-1}, s)$ is always higher than the constituents' optimal level of public spending, $\tilde{g}(s)$. Moreover, equilibrium expenditures are positively correlated with previous realizations of median constituent income and population.

Proposition 1. *In the political equilibrium,*

1. *Expenditures are always higher than the constituents' optimum:* $g^*(g_{-1}, s) > \tilde{g}(s)$.

2. *Expenditures are* increasing *in the previous period's (a) median constituent income and (b) population, so that*

$$\frac{dg^*}{dy_{-1}}(g_{-1}, s) > 0 \qquad \textit{and} \qquad \frac{dg^*}{dz_{-1}}(g_{-1}, s) > 0.$$

[21] All proofs are in Appendix C.

The second component of Proposition 1 indeed captures the empirical result in Section 1.2 that public expenditures are positively correlated with past population realizations, or equivalently negatively related to the population growth rate z/z_{-1}. In addition, the proposition also indicates that a similar pattern should also hold for the growth rate in median household income: consider two municipalities, i and j, that are identical in their current demographic characteristics $s_i = s_j$. Since demographics are symmetric, then constituents in both cities have identical preferences for public spending. Nevertheless, if $y_{-1,i} > y_{-1,j}$ ($z_{-1,i} > z_{-1,j}$) so that past median income (population) in city i exceeded city j, then public spending will be strictly higher in city i, ceteris paribus. In other words, conditional on s, public spending is *decreasing in a municipality's median income growth and population growth*, y/y_{-1} and z/z_{-1}. Thus, declining cities exhibit more public spending then growing cities.

The intuition behind Proposition 1 is straightforward. Regarding the first component, the union always finds it beneficial to incentivize the party to set expenditures higher than $\tilde{g}(s)$, as the union's current and future payoffs are strictly increasing in its current budget allocation, whereas the marginal cost of incentivizing the party is zero at $\tilde{g}(s)$. Consequently, spending is always too high if public workers can influence the municipality's policymakers.

The second component of Proposition 1 follows from two facts. The first fact is that public spending is strictly increasing in both population and the median citizen's income: constituents' marginal benefit of the public good is increasing in population and income, which implies that the union finds it marginally cheaper to incentivize the politician to increase expenditures when these variables increase.

The second fact is that current spending is also strictly increasing in the union's status quo budget allocation: a larger status quo budget implies that the union has

a higher marginal benefit for current spending. Consequently, $g^*(g_{-1}, s)$ is strictly increasing in g_{-1}, as the union pledges greater campaign contributions in exchange for additional policy concessions. Proposition 1 then follows from combining these two facts.

One caveat worth mentioning is the interpretation of the phrase "policy persistence." In cities experiencing (population or income) decline, public expenditures persist in the traditional sense, as the union exerts increasing influence on the politician so that total spending remains too high. Expenditures in growing cities, on the other hand, are smaller than expenditures in cities with stable population not because investment is lagging, but because *the union is relatively less influential.* In the political equilibrium, expenditures always exceed the public's optimum; in growing cities, the excess is simply smaller in magnitude.

Proposition 1 captures the persistence that was observed in Section 1.2. The following analysis is employed to derive several other testable predictions that we can check in the data. While Proposition 1 indicates that population and income shocks generate policy persistence, the result is silent regarding which type of shock generates a stronger response. Proposition 2 indicates that income shocks generate a stronger response: a 1% increase in *past constituent income* increases current expenditures by relatively more than a 1% increase in the municipality's past population level. This proposition, however, requires the unrealistic restriction that population shocks are iid. This condition makes the intuition behind the mechanism straightforward to understand. Our model simulation, however, verifies the insights from Proposition 2 when the population shock process is calibrated to our data.

Proposition 2. *If population shocks are iid ($\gamma_z = 0$), then public expenditures are*

more sensitive to previous realizations of constituent income, relative to previous population realizations:

$$\frac{y_{-1}}{g^*(g_{-1},s)} \cdot \frac{dg^*}{dy_{-1}}(g_{-1},s) \quad > \quad \frac{z_{-1}}{g^*(g_{-1},s)} \cdot \frac{dg^*}{dz_{-1}}(g_{-1},s).$$

The intuition behind Proposition 2 is as follows. Recall dg^*/dy_{-1} and dg^*/dz_{-1} can be decomposed into two components: first, $g^*(g_{-1},s)$ in increasing in y and z (in the previous period), and second, $g^*(g_{-1},s)$ is increasing in g_{-1} (in the current period). Consequently, Proposition 2 is equivalent to showing that public expenditures are more sensitive to income shocks.

These derivatives depend on y and z via two channels: (i) through the municipality's tax base, yz, and (ii) directly through the number of individuals residing in the city, z. The elasticity of the tax base with respect to both income and population is 1: A 1% increase in either median income or population results in a 1% increase in the tax base. Consequently, a percentage increase in either variable has an identical effect on constituent preferences for the public good through the tax base. This implies that, via channel (i), the union's marginal cost of influencing the politician responds identically to percent changes in y and z.

Channel (ii), however, depends only on the municipality's population. In Proposition 1, we established that equilibrium expenditures are always higher than the constituents' optimum. As population rises, more individuals suffer from the excessive costs of the public budget. Consequently, a 1% increase in population makes it *more costly* for the union to influence the politician, via channel (ii). Therefore, the elasticity with respect to median income is larger in magnitude.

Analytically, the restriction that population is iid is sufficient because of the dy-

namic component of $dg^*(g_{-1}, s)/dz$, which incorporates how an increase in population today influences the future evolution of population. For instance, if the expected growth rate of population is sufficiently larger than for income, then a small increase in current population might have a drastic impact on expected future population and influence current expenditures by more than the static effect for an increase in income. The insight from Proposition 2 also holds if we replace the iid restriction with bounds on the expected growth rates and relative magnitude of median income and population.[22] As mentioned above, however, our simulation results in Section 1.3.4 robustly support this proposition when (γ_y, γ_z) are calibrated to the data.

The next proposition analyzes the responsiveness of public expenditures to changes in the public sector wage. The two insights are that public spending is increasing in the public sector wage, and that the price elasticity of public spending is *higher* in the absence of political activity (the social planner's policy function, $\tilde{g}(s)$). We have only shown this second insight analytically, however, for the restrictions that wage shocks are iid and the union's campaign cost is linear. While these conditions are sufficient, they are by no means necessary: in our calibration, we relax these assumptions and find that Proposition 3 holds across the entire state space.

Proposition 3. *In the political equilibrium,*

1. *Expenditures are strictly increasing in the public sector wage, w:*

$$\frac{\partial g^*}{\partial w}(g_{-1}, s) \quad > \quad 0.$$

2. *If (a) wage shocks are iid and (b) the union's campaign cost is linear, then*

[22] Specifically, alternative sufficient conditions are (a) in expectation, median income grows at least as much as city population ($\gamma_y \geq \gamma_z$) and (b) the median constituent's income is smaller in magnitude than the city's population ($\max_{s \in S} y \leq \min_{s \in S} z$). Both conditions are supported in the data.

the price elasticity of public expenditures is higher *if the union lacks political influence:*

$$\frac{w}{\tilde{g}(s)} \cdot \frac{\partial \tilde{g}}{\partial w}(s) \quad > \quad \frac{w}{g^*(g_{-1}, s)} \cdot \frac{\partial g^*}{\partial w}(g_{-1}, s). \tag{1.6}$$

The first part of Proposition 3 is in accordance with Comment 2: as w increases, constituents' marginal benefit of the public good increases because demand is inelastic with respect to public good provision. Consequently, it is marginally "cheaper" for the union to influence the party, and thus $g^*_w(g_{-1}, s) > 0$.

Part two of Proposition 3, on the other hand, states that a particular public good's budget allocation is *less sensitive* to the public sector wage when public workers are *politically influential*. In other words, if the public sector wage increases by 1%, then politically active unions will enjoy a smaller budget increase relative to workers that are not influential in local elections.

This result follows from two facts. First, in the political equilibrium, the concavity of the union's objective function is greater than the political party's optimization problem: the union takes into account both the political party's utility (via the incentive constraint), as well as its own utility. This increased concavity implies that the $g^*(g_{-1}, s)$ responds less to wage changes than if the politician unilaterally sets public spending. Second, as the level of public expenditures increases, constituents' price elasticity of demand for the public good becomes smaller in magnitude, due to overspending. Combining these facts, it follows that the price elasticity of public good expenditure is lower when public workers are politically influential.

1.3.4 Simulation

We now present a simple simulation exercise to investigate whether the magnitude of the policy persistence, as implied by our model, is similar in magnitude to what is observed in the data. In addition, we use the simulation results to verify that Propositions 2 and 3 are robust to relaxing the restrictions discussed above.

Model calibration and simulation algorithm

Recall that the union's period payoff is $u(g, g_{-1}) - \phi(C)$. We adopt the standard quadratic cost function $\phi(C) = C^2$ and the Cobb-Douglas functional form to represent the union's preferences over the budget allocation, $u(g, g_{-1}) = g^{\alpha} g_{-1}^{1-\alpha}$. The "true" value of α is unclear, so we present results below for $\alpha \in \{0.5, 0.667, 0.8\}$ and adopt $\alpha = 0.667$ as our baseline parameter value. We consider values of $\alpha \geq 0.5$ because α is the weight that the union places on the current budget allocation, relative to the previous budget allocation.

The two parameters from the public welfare function, (σ, η), are set using estimates from previous research and the constituent demand function $\tilde{g}(s)$. Previous papers have estimated that the price elasticity of demand for local public goods is around -0.45.[23] Given the demand function $\tilde{g}(s)$, this price elasticity corresponds to $\sigma = -1.33$.[24] Given $\sigma = -1.33$, we also compute the corresponding value for η using $\tilde{g}(s)$.[25]

[23] See, for instance, as discussed in Oates (2006) and, for the case of education, Bergstrom et al. (1982).

[24] The constituents' price elasticity of demand for public expenditures, $[w/\tilde{g}(s)] \cdot [\partial \tilde{g}(s)/\partial w]$, depends on σ and $\tilde{g}(s)/yz$, which is the percentage of the tax base that is allocated to local public goods. We estimate $\tilde{g}(s)/yz$ in our data using operational expenditures and selecting the median value of the distribution (3.69%). While the median cutoff is ad hoc, σ is not very sensitive to alternatives.

[25] Given $\tilde{g}(s)/yz$ (see Footnote 24), σ and w, it is straightforward to calculate η. We set $w = 43.7$, which corresponds to the median value of the *average municipal employee salary* in our data (in

We adopt the standard annual discount rate of 0.95. Since our data is observed at 10-year intervals, we set the model's discount rate $\beta = 0.95^{10} = 0.6$. The weight that the political party places on citizen welfare, λ, is set to 0.984/0.016. This value is taken from Goldberg and Maggi (1999), which estimates λ when performing a structural estimation of the Grossman and Helpman (1994) model.

The expected (gross) growth rates for median constituent income and population, $(\gamma_y, \gamma_z) = (1.071, 1.034)$, are computed by taking the median values of the respective growth rate distributions in our data, as shown in Table 1.1. We also calibrate $(\epsilon_y, \epsilon_z) \sim N(0, \Sigma)$ using our data,[26] where

$$\Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yz} \\ \Sigma_{yz} & \Sigma_z \end{pmatrix} = \begin{pmatrix} 2.715 & 5.906 \\ 5.906 & 189.562 \end{pmatrix}.$$

The dimensions of the state space S are all divided by 1,000, so, for instance, $\Sigma_y = 2.715$ corresponds to a standard deviation of y shocks of \$1,648. Given that $B_{.t}$, the indicator for strong state-level public sector collective bargaining laws, is used to proxy for w_t in our empirical analysis, we assume that w_t is constant ($\gamma_w = 1$, $\epsilon_w = 0$), so that changes in the public sector wage are unexpected.

The optimal policy $g^*(g_{-1}, s)$ is computed numerically via value function iteration, using a tolerance of 10^{-6}. Given $g^*(g_{-1}, s)$, we then simulate the effect of income and population shocks on public expenditures. As a benchmark, we simulate sample paths $\{g_t^n\}_{n,t}$ for 10,000 municipalities ($n = 1, ..., 10{,}000$ and $t = 1, ..., 5$) *without* imposing shocks:

thousands). It follows that $\eta = 3.26 \times 10^{-6}$. η is relatively small because $\tilde{v}(s)$ exhibits considerable curvature near 0.

[26] The distribution of (ϵ_y, ϵ_z) appears approximately normal in our data.

1. At $t = 3$, we randomly select an exogenous state s_3^n for each municipality, drawing from a uniform distribution. Given s_3^n, we compute $s_t^n = \gamma s_{t-1}^n + \epsilon_t^n$ for $t = 4, 5$ and $s_t^n = (s_{t+1}^n - \epsilon_{t+1}^n)./\gamma$ for $t = 1, 2$, where $\epsilon_t^n \sim N(0, \Sigma)$.

2. Initial expenditures $\{g_1^n\}_n$ are computed using the model

$$\ln(g_1^n) = \hat{\xi}_0^g + \hat{\xi}_1^g \ln(y_1^n) + \hat{\xi}_2^g \ln(z_1^n),$$

   where $\hat{\xi}^g$ is estimated from our data via OLS.[27]

3. Expenditures $\{g_t^n\}_{n,t=2}^5$ are computed using $g^*(g_{-1}, s)$, $\{s_t^n\}_{n,t=2}^5$ and g_1^n.

We also simulate four expenditure paths that each impose a shock between $t = 2$ and $t = 3$: (i) a positive y shock of magnitude $\overline{\gamma_y}$; (ii) a positive z shock of magnitude $\overline{\gamma_z}$; (iii) a negative y shock of magnitude $\underline{\gamma_y}$; and (iv) a negative z shock of magnitude $\underline{\gamma_z}$. The values of these shocks are computed using the growth rate distribution for income and population in our data. Positive (negative) shocks are defined at the 90th (10th) percentile of the respective distributions. Our baseline parameter values and $(\overline{\gamma_y}, \overline{\gamma_z}, \underline{\gamma_y}, \underline{\gamma_z})$ are reported in Table 1.5.

The construction of the "shock" expenditure paths differ from the benchmark case in one respect; we explain the difference using the positive y shock as an example. Instead of computing the simulation state at $t = 2$ using the formula $s_2^n = (s_3^n - \epsilon_3^n)./\gamma$, we impose the positive y shock by computing $y_2^n = y_3^n/\overline{\gamma_y}$, and then compute $z_2^n = (z_3^n - \overline{\epsilon}_{z,3}^n)/\gamma_z$, where $\overline{\epsilon}_{z,3}^n$ is drawn conditional on $\epsilon_{y,3}^n = y_3^n - \gamma_y y_2^n$. Note that we impose this shock retroactively so that municipalities have the same distribution of demographic characteristics at $t = 3$ and differ only by income (or population) in the

[27] We estimate $\hat{\xi}^g = (-1.062, 0.380, 0.863)$.

Table 1.5: *Baseline parameter values*

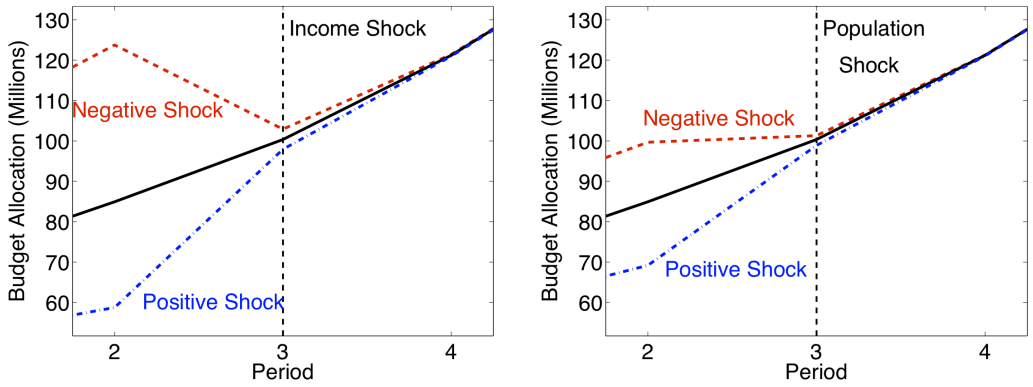
Parameter	Value	Source
α	0.667	Union weights g twice as much as g_{-1}
β	0.6	Annual discount rate of 0.95
λ	0.984/0.016	Goldberg and Maggi (1999)
σ	-1.33	Literature, computations using $\tilde{g}(s)$
(γ_y, γ_z)	$(1.071, 1.034)$	Data: Median growth rate
$(\Sigma_y, \Sigma_z, \Sigma_{yz})$	$(2.715, 189.562, 5.906)$	Data: Computed error variance
$(\overline{\gamma_y}, \overline{\gamma_z})$	$(1.265, 1.292)$	Data: 90th percentile of growth rate dist.
$(\underline{\gamma_y}, \underline{\gamma_z})$	$(0.916, 0.905)$	Data: 10th percentile of growth rate dist.

previous periods.

In addition to this simulation exercise, we also check that inequality (1.6) in Proposition 3 holds across the state space. While it is straightforward to compute $\partial\tilde{g}(s)/\partial w$ analytically, $\partial g^*(g_{-1}, s)/\partial w$ must be computed numerically. We use our numerical solution for $g^*(g_{-1}, s)$ and a quadratic approximation technique to estimate this elasticity: for each point $(g_{-1}, y, z) \in G \times Y \times Z$ on the grid, we estimate the model $g^*(g_{-1}, (w_i, y, z)) = \xi_0^w + \xi_1^w w_i + \xi_2^w (w_i)^2 + u_i$ across the grid points $\{w_i\}_i$, where u_i is an error term. Using the estimates for ξ^w, we can then estimate $[w/g^*(g_{-1}, s)] \cdot [\partial g^*(g_{-1}, s)/\partial w]$ for each $\{w_i\}_i$, given the triple $(g_{-1}, y, z) \in G \times Y \times Z$. For all parameterizations that we consider, inequality (1.6) holds at over 99.5% of the 62,500 grid points.

Numerical results

Figure 1.1 displays municipal spending for the parameter values listed in Table 1.5, with time plotted on the horizontal axis and aggregate expenditures plotted on the vertical axis. In both panels, the solid line represents the expenditure level in the



(a) *Income shock* **(b)** *Population shock*

Figure 1.1: *The path of public expenditures following a shock*

benchmark (no shock) scenario, averaged across the 10,000 simulated municipalities. Figure 1.1a displays aggregate spending when an unexpected income shock occurs in the municipality at $t = 3$, whereas Figure 1.1b displays spending following an unexpected population shock at $t = 3$. In both panels, the red, dashed (blue, dot-dasheded) line denotes aggregate spending for the negative (positive) shock scenario.

As discussed in Proposition 1, municipalities that experience a negative (positive) shock to either income or population spend more (less) than the benchmark city, which can be seen in the gaps between the solid, benchmark line and the other expenditure paths at $t = 3$. These gaps dissipate quickly, however, as public expenditures quickly converge to the benchmark path in the decades following the initial shock.

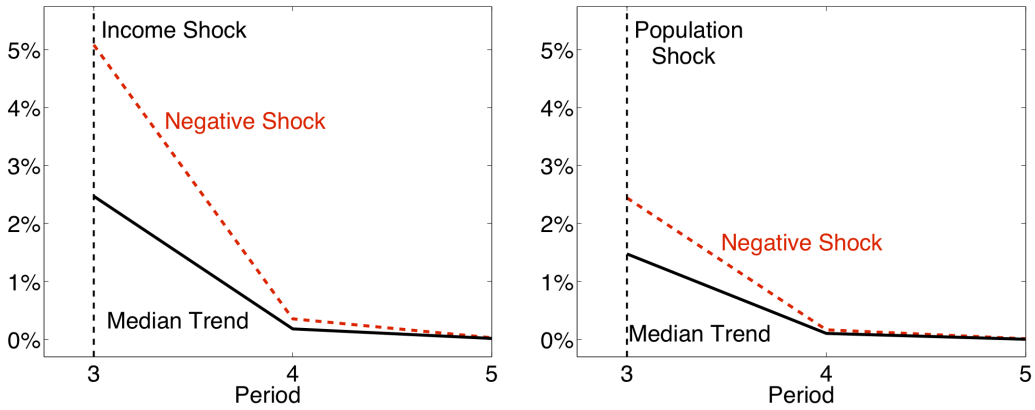
Of the three types of expenditure paths (benchmark, negative shock and positive shock), the path that imposes the positive shock is closest in magnitude to the constituents' preferred expenditure level at $t = 3$, as additional spending is minimized in this case. Consequently, Figure 1.2 plots the percentage of additional spending that occurs in the benchmark and negative shock paths, *relative to the path that imposes*

the positive shock. This percentage is a lower bound on the share of excess spending in the benchmark and negative shock scenarios and will be referenced in our discussion below. Note that the additional spending is driven solely by the union's political activity: the distribution of demographic characteristics is identical across all sample paths for $t = 3, 4, 5$.

Figure 1.2a shows that a municipality spends about 5.1% more on the public good when constituent income experiences a negative shock of magnitude $\underline{\gamma}_y$, relative to if the city had experienced a positive shock of magnitude $\overline{\gamma}_y$ (holding current income fixed). In addition, Figure 1.2a also shows that the benchmark, no-shock municipality spends about 2.5% more, relative to the positive income shock scenario. Figure 1.2b displays the analogous results for the population shock. In this case, the municipality spends about 2.4% and 1.5% more in the negative and no-shock scenarios, relative to the positive shock scenario. Finally, note that Figure 1.2 robustly supports Proposition 2, as income shocks generate about twice as much persistence as the population shocks.[28]

Table 1.6 reports the percentage of additional spending for $\alpha \in \{0.5, 0.667, 0.8\}$. For each value of α, the table reports four values corresponding to the four percentages of additional spending at $t = 3$ that were mentioned in the description of Figure 1.2: the percentage of additional spending in a municipality that experienced the negative income shock $\underline{\gamma}_y$, relative to an identical city that experienced the positive income shock $\overline{\gamma}^y$, and so on. Persistence is always about twice as large in the case of the income shock, relative to the population shock: the range of estimated persistence for the negative income shock is about 3.7% to 5.7%, whereas the analogous population

[28]Note that this is the case, even though the difference in magnitude of the population shocks is larger: $\overline{\gamma_y} - \underline{\gamma_y} < \overline{\gamma_z} - \underline{\gamma_z}$.



(a) *Income shock* **(b)** *Population shock*

Figure 1.2: *Percentage of additional spending, relative to the growing city*

shock ranges from about 1.8% to 2.7%, depending on the value of α. We will examine the comparability of these results with the implications from our empirical estimates in the next section.

Table 1.6: *Simulation results across α values: Percentage of additional spending, relative to the positive shock scenario*

Shock Type	**0.50**	**0.667**	**0.80**
Negative y shock	5.73%	**5.08%**	3.72%
Median y trend	2.69%	**2.47%**	1.85%
Negative z shock	2.74%	**2.42%**	1.80%
Median z trend	1.61%	**1.48%**	1.13%

1.4 Empirics revisited

We now revisit the data to further investigate whether the theoretical mechanism outlined above may be contributing to a large share of the observed spending persis-

tence. First, we reestimate the SDM in equation (1.3) after incorporating the log of the median income growth rate, $\log(y_{it}/y_{i,t-10})$, and then compare the implied level of spending persistence with the simulation results in Table 1.6. Second, we test the other theoretical implications of the theory by disaggregating our measure of spending to individual types of public goods. And finally, we conclude the section by estimating the magnitude of additional spending that occurs in individual US municipalities.

1.4.1 Comparison of model simulation and empirics

Proposition 1 established that shocks to median household income should also cause policy persistence. We thus reestimate model (1.3), incorporating $\log(y_{it}/y_{i,t-10})$ as a covariate and controlling for the other possible persistence mechanisms. The results are reported below, in Column (1) of Table 1.10. The coefficient estimates for population and income change are both negative and significant at the 5% level. As Proposition 2 predicted, the elasticity of public spending with respect to income growth (-0.229) is larger in magnitude than the elasticity with respect to population growth (-0.107). In fact, the magnitude is twice as large.

Given these two parameter estimates, we now create the analog to the predicted additional spending shown in Figure 1.2 of the model simulation. As in Section 1.3.4, we quantify the lower bound estimate for additional spending by taking a "growing city" (at the 90th percentile of the population or income growth distribution) as our benchmark and using the parameter estimates to calculate how much more the city would spend if its growth rate was at the 10th and 50th percentile of the distributions. To begin, we use our data to compute the typical full time municipal wage bill per household in a growing city, for cities that the 90th percentile of the income growth and population growth distributions, respectively. We then use the growth rates at

Table 1.7: *Additional annual spending on the wage bill per household*

Income Shock			Population Shock		
Growing	Median	Declining	Growing	Median	Declining
\$1,794.07	+\$71.63	+\$142.05	\$1,822.94	+\$42.76	+\$69.93
	(+3.99%)	(+7.92%)		(+2.35%)	(+3.84%)

the 10th, 50th and 90th percentiles of these distributions (reported in Tables 1.1 or 1.5) and the parameter estimates (-0.229 for income and -0.107 for population) to estimate the additional expenditures that occur in the "median" and "declining" cities, respectively.

Table 1.7 reports the estimates of additional spending per household, both in terms of dollars and as a percentage of expenditures in the "growing city." In the case of a negative income shock (10th percentile of the distribution), a municipality spends about \$142, or 7.9%, more per household on the full time wage bill, relative to what the same municipality would spend if it had experienced a shock at the 90th percentile in the distribution (\$1,794.07). A municipality experiencing the median income shocks, on the other hand, spends about \$72, or 4.0%, more per household relative to the growing city.

These results are even more pronounced than the simulation results presented in Table 1.6: additional spending in the "negative shock" ("median shock") city was about 5.1% (2.5%) higher than the "positive shock" benchmark. For population shocks, the additional spending was estimated to be around 2.4% for the "declining shock" and 1.5% for the "median shock" scenarios, respectively.

Table 1.8: *Description of the local public goods in our data*

Public Good	Description
Administration	Handling of government-wide planning and legal matters
Fire protection	Fire protection, prevention and rescue services
Road maintenance	Maintenance of roads, bridges and tunnels
Parks and recreation	Maintenance of public parks and recreational services
Police protection	Enforcement of law and order

1.4.2 Comparing persistence across public goods

Up to this point, we have investigated public spending at an aggregate level and treated public workers as one uniform body. In reality, however, municipalities produce a variety of public goods and allocate distinct budgets to each group. Moreover, it is commonly believed that political influence varies considerably across these groups. Some types of public workers, such as police officers, are historically very active in local elections and thus should fit the mold of the "union" in the theory presented above. On the other hand, other types of public workers, such as park maintenance personnel, are not historically known for strongly influencing the local electorate, and thus spending on park maintenance should more closely mirror the social planner's problem discussed above.

To further investigate such issues and test Proposition 3, we analyze spending persistence for the five public goods that are listed in Table 1.8. These five public goods are chosen because the vast majority of large cities directly produce the goods by hiring labor and investing in capital.[29] As with the total wage bill, we extract infor-

[29] In our panel of 595 cities, all directly provide administrative services; 94.6% directly provide fire protection services; 96.8% maintain roads directly; 91.9% maintain public parks; and 90.0% provide law enforcement services. Cities that do not directly provide these services typically outsource production to the private sector or another government body, such as the county or state. For instance, many Californian cities receive fire protection services directly from the California Department of

Table 1.9: *Per capita expenditures, by public good*

	Mean	Median	10th	90th
Per cap administrative wage bill ($)	47.37	40.57	20.91	79.74
Per cap fire wage bill ($)	91.78	84.40	53.31	140.67
Per cap parks wage bill ($)	25.06	21.95	8.07	45.30
Per cap police wage bill ($)	121.68	108.54	68.37	185.81
Per cap roads wage bill ($)	31.85	28.43	13.51	54.73

Notes. The number of sample municipalities reporting wage bill information across the public goods is as follows: administration (595); fire (563); parks (547); police (589); and roads (576). All variables are annualized and reported in 2005 dollars.

mation on the annual wage bill paid to full time employees from the CoG and adjust for inflation by converting the wage bill data to 2005 dollars. Summary statistics for the disaggregated spending data are reported in Table 1.9.

According to the ICMA political activity data, of these five services, fire fighters and police officers are relatively more politically powerful than the employees that produce the other three types of public goods.[30] Moreover, anecdotal evidence also affirms that these two services are the most politically active (see, for instance, Najita and Stern (2001)).

To investigate possible persistence across public goods, we estimate the SDM regression separately for each of the five services by replacing the dependent variable, $\ln(g_{it})$, with the time-t wage bill for each public good.[31] The "partial equilibrium" approach of estimating separate regressions for each type of public good is quite

Forestry and Fire Protection.

[30] In cities with more than 25,000 residents, about 55% (59%) of fire protection (law enforcement) worker associations issue official endorsements for local officials, while 38% (35%) offer financial contributions to fund political campaigns and 29% (27%) actively campaign for candidates. Conversely, a category aggregating "other" local public services reports that workers engage in political activity in 26% (endorsements), 20% (financial contributions) and 17% (active campaigning) of municipalities.

[31] As mentioned above, the NBER Public Sector Collective Bargaining Law Data Set reports bargaining laws separately for police officers, fire fighters and other local public employees. We include the appropriate B_{it} for each regression.

common in the empirical local public finance literature (see, for instance, Alesina et al. (1999)). The assumption of independence across public goods is equivalent to posing that expenditure adjustments for one service do not impact other services' budget allocations. Given that most municipalities face soft budget constraints, the assumption does not too restrictive: the average city spends approximately 60% of its operational budget on wages and salaries for all full-time workers.

Results of the five regressions are reported in Columns (2)-(6) in Table 1.10. We observe that police and fire services exhibit considerable persistence, even after controlling for the other persistence mechanisms: a 1% decrease in the income (population) growth rate leads to an increase in expenditures of 0.33% (0.16%) and 0.25% (0.20%) for police and fire, respectively. Moreover, with coefficients for the other three services are all, in most cases, negative in sign but statistically insignificant. This pattern is largely consistent with the theoretical analysis presented above, in that relatively more politically active services should exhibit more policy persistence, relative to services with less (or no) political sway. Moreover, Proposition 2 is also supported: the estimated elasticity of income growth is larger in magnitude for four of the five services.

While we don't directly observe public sector wages in our data, we do observe one variable that is highly correlated with w_t: public workers' access to collective bargaining rights. As mentioned above, a number of studies such as Hoxby (1996) and as reviewed by Freeman (1986) have examined the effect of collective bargaining on public sector spending and wages, and substantial evidence indicates that collective bargaining leads to higher wages and spending. The bargaining power indicator $B_{\cdot t}$ in our empirical analysis thus serves as a variable that's highly correlated with w_t.

Table 1.10 provide some support for Proposition 3. Across all regressions, the

Table 1.10: *Full time municipal wage bill, by public good*

	[1] Total	[2] Police	[3] Fire	[4] Roads	[5] Parks	[6] Admin
$\log(z/z_{-10})$	-0.107**	-0.158***	-0.204***	0.027	-0.008	-0.037
	(0.054)	(0.047)	(0.061)	(0.114)	(0.117)	(0.073)
$\log(y/y_{-10})$	-0.229**	-0.327***	-0.246**	-0.282	-0.139	0.022
	(0.093)	(0.082)	(0.104)	(0.196)	(0.207)	(0.126)
B	0.109**	0.058***	0.084***	0.136***	0.143***	0.132***
	(0.022)	(0.019)	(0.024)	(0.048)	(0.052)	(0.030)
Share Dem. Vote	0.248*	-0.149	-0.166	0.308	-0.593**	0.337*
	(0.141)	(0.123)	(0.154)	(0.293)	(0.298)	(0.189)
stayers	0.833*	0.034	0.039	0.192	0.365	0.252
	(0.277)	(0.241)	(0.302)	(0.581)	(0.598)	(0.373)
ρ	0.217**	0.149***	0.111***	-0.011	0.065**	0.055**
	(0.007)	(0.005)	(0.005)	(0.051)	(0.026)	(0.023)
$\log(z)$	0.787**	0.799***	0.763***	0.867***	0.827***	0.850***
	(0.047)	(0.042)	(0.052)	(0.099)	(0.102)	(0.064)
$\log(y)$	0.414**	0.655***	0.601***	0.438	0.636**	0.450**
	(0.140)	(0.122)	(0.155)	(0.294)	(0.308)	(0.188)
unem	-0.962**	-0.852**	-0.172	-1.197	-1.39	-1.682***
	(0.423)	(0.368)	(0.469)	(0.887)	(0.948)	(0.569)
poverty	0.257	-0.565**	-0.502	-0.021	-0.060	-0.470
	(0.307)	(0.267)	(0.336)	(0.646)	(0.667)	(0.413)
college	0.208	-0.127	-0.037	0.388	-0.051	0.090
	(0.199)	(0.174)	(0.220)	(0.418)	(0.439)	(0.268)
schage	-0.432	0.221	-0.192	2.836***	-2.11**	0.516
	(0.480)	(0.422)	(0.556)	(1.011)	(1.041)	(0.646)
old	0.608*	-0.238	0.370	2.344***	1.654**	0.920*
	(0.361)	(0.315)	(0.401)	(0.758)	(0.786)	(0.485)
black	0.047	0.412**	0.276	-0.154	-0.199	0.189
	(0.184)	(0.161)	(0.213)	(0.404)	(0.404)	(0.247)
white	-0.298*	-0.306**	-0.250	0.083	-0.191	-0.306
	(0.157)	(0.138)	(0.180)	(0.335)	(0.340)	(0.212)
log(ineq)	0.056	0.098	0.127	0.024	-0.084	0.489***
	(0.127)	(0.111)	(0.141)	(0.267)	(0.274)	(0.171)
log(crime)	-0.004	0.053***	0.069***	0.032	0.029	0.009
	(0.017)	(0.015)	(0.019)	(0.037)	(0.039)	(0.024)
log(land area pc)	0.040	0.007	-0.045	0.126*	0.133*	0.230***
	(0.036)	(0.031)	(0.039)	(0.075)	(0.078)	(0.048)
boundary	0.028	0.019	0.038	-0.074	-0.078	-0.065
	(0.030)	(0.026)	(0.033)	(0.063)	(0.066)	(0.041)
Number of Cities	595	589	563	576	547	595

Notes. The dependent variable is the log of the full time wage bill paid to employees that produce the public good specified in the column name. Regressions include time fixed-effects and coefficients for $\omega_i'\mathbf{F_t}$ (unreported). Heteroskedastic-robust standard errors are reported in parentheses. ***, ** and * correspond to significance at the 1, 5 and 10 percent levels.

effect of $B_{\cdot t}$ on spending is positive and significant, which corresponds to the first component of the proposition. Moreover, in Table 1.10, we observe that the estimated marginal effect of $B_{\cdot t}$ for fire and police services are at least two standard deviations smaller than the analogous estimated effect for administration, parks and road maintenance services. Given that fire fighters and police officers are traditionally the most politically active local public workers, this provides some evidence for the second component of Proposition 3: namely, the elasticity of operational expenditures with respect to wage should be *lower* for services with relatively more political power.

In addition, it is also useful to note that other coefficient estimates are in congruence with what one might expect: for instance, the coefficient on the crime rate is positive and significant at 1% for police and fire services, whereas this coefficient is insignificant for the other three services, while population and income level are positively associated with all services at the 1% or 5% significance level (except for income in the case of roads).

1.4.3 Discussion and policy implications

Table 1.7 presented estimated the magnitude of policy persistence with respect to municipal spending per household. Building off of this methodology, Table 1.11 provides specific cost estimates for the 25 largest US cities, as of 2010. Population data are gathered from the 2010 Decennial Census, while median income information are gathered from the 2009 American Community Survey and full-time employee wage bill data are collected from the 2009 Annual Survey of Governments. Given this data, we then compute persistence-induced spending implied by our parameter estimates in Column (1) of Table 1.10 and the demographic growth rates over the previous decade.

The final two columns of Table 1.11 are entitled "Additional Spending." The

Table 1.11: *Estimates of additional spending in the 25 largest US cities (circa 2010)*

City	State	Pop. (2010)	Pop. growth (2000-10)	Median income (2009)	Median income growth (2000-2009)	Additional spending $'000s	Additional spending % of wage bill
New York	NY	8,175,133	2.1	51,009	4.1	702,464	3.48
Los Angeles	CA	3,792,621	2.6	48,199	3.1	126,542	3.64
Chicago	IL	2,695,598	-6.9	47,412	-5.2	112,624	6.45
Houston	TX	2,099,451	7.5	43,069	-9.0	43,408	5.89
Philadelphia	PA	1,526,006	0.6	41,650	-3.9	69,058	5.38
Phoenix	AZ	1,445,632	9.4	49,054	-9.8	43,510	5.90
San Antonio	TX	1,327,407	16.0	46,412	-4.0	22,657	3.95
San Diego	CA	1,307,402	6.9	67,020	8.0	11,822	2.18
Dallas	TX	1,197,816	0.8	38,856	-18.8	60,675	8.94
San Jose	CA	945,942	5.7	76,689	-12.4	31,597	6.87
San Francisco	CA	805,235	3.7	78,909	6.1	46,639	2.89
Austin	TX	790,390	20.4	56,559	-10.6	27,580	5.13
Columbus	OH	787,033	10.6	46,084	-16.9	27,608	7.54
Fort Worth	TX	741,206	38.6	49,508	-1.5	4,034	1.52
Charlotte	NC	731,424	35.2	55,326	-16.3	14,899	5.38
Detroit	MI	713,777	-25.0	28,225	-28.7	79,099	14.36
El Paso	TX	649,121	15.2	38,600	-6.9	9,517	4.69
Memphis	TN	646,889	-0.5	37,078	-16.1	67,377	8.38
Baltimore	MD	620,961	-4.6	42,916	3.5	49,931	4.30
Boston	MA	617,594	4.8	58,737	13.7	11,933	1.23
Seattle	WA	608,660	8.0	81,319	11.8	7,914	1.30
Washington	DC	601,723	5.2	64,799	19.7	314	0.02
Denver	CO	600,158	8.2	53,320	-5.4	28,503	4.98
Milwaukee	WI	594,833	-0.4	35,603	-19.7	28,772	9.29
Portland	OR	583,776	10.3	56,017	-4.8	13,795	4.65
Las Vegas	NV	583,756	22.0	53,664	-9.1	8,772	4.63
Vallejo	CA	115,942	-0.7	54,971	-17.3	3,760	8.70
Harrisburg	PA	49,528	1.2	27,250	-21.2	2,420	9.53

first of these columns provides the aggregate amount of the extra spending in the city, relative to a city at the 90th percentile in both the population and median income distributions, while the second column displays the percentage of spending that the additional amount represents. All of the 25 sample cities spend more than the benchmark growing city, with the median of these 25 municipalities spending 4.84% more. The median dollar amount of additional spending per household corresponds to about $134 annually and exceeds $400 in Detroit.

These computations are quite meaningful for cities in distress. Vallejo, CA and Harrisburg, PA have both received considerable attention due to their fiscal challenges. As mentioned above, Vallejo entered Chapter 9 bankruptcy in 2008 when facing a $16 million deficit, and some commentators have blamed influential police officers and fire fighters as a contributing factor (Jones, 2008). According to our computations, additional spending associated with policy persistence represents about 24% of this deficit. Harrisburg, on the other hand, fell behind its debt payments in 2010 and received a $4.4 million bailout from the state (Varghese, 2010). The estimate on the value of additional persistence-related spending is about 56% of bailout funds that the city received.

1.5 Conclusion

This paper investigated public spending persistence in US municipalities and several potential causes of this persistence. Empirical analysis of 595 municipalities suggested that operational expenditures are negatively correlated with the growth rate of municipal population and median household income, suggesting that spending does not adjust to changes in public demand. The paper investigated four mechanisms that

might cause this empirical pattern, and found some evidence in support of an interest group mechanism. A dynamic theoretical model was then analyzed to derive several qualitative and quantitative predictions for observables, which were consistent with the data. Cost estimates indicate that additional spending constitutes about $134 per household in the median large city.

Several avenues of future research are as follows: first, public workers' political activity was observed only for a subset of cities. A more robust test of the political mechanism would involve building a political activity data set, such as by direct surveying city governments as in Ferreira and Gyourko (2009). Second, while a large theoretical literature has investigated the political influence that bureaucrats and public workers have on public spending, considerably less attention has been devoted to empirical applications to investigate the additional spending that is caused by these political activities. Future research could shed light on the influence such workers have on spending. Finally, an important local public service that we've omitted from our analysis is public education. Moe (2006) has investigated the political activity of teachers in California, and a broader investigation may yield interesting insights.

Chapter 2

2.1 Introduction

Elections are often utilized as a mechanism by which constituents can hold policymakers accountable. Electoral competition, however, may provide countervailing incentives to elected officials: competitive elections may mitigate private rent seeking, but may also promote opportunistic vote buying. While both channels have been addressed in the literature, empirical research has tended to analyze these incentives in isolation and, to our knowledge, has not addressed the inherent trade-off between these activities.

This paper examines how policymakers adjust their private rent seeking and vote buying activities in response to electoral competition. The analysis utilizes survey data that contain information about a decentralized program in a developing country that selected households eligible for social assistance benefits intended for the poor. The data contain detailed household-level information about program beneficiaries and non-beneficiaries, allowing us to identify the types of households that were inappropriately included and excluded from the program. These data are then linked to census and election data, which are used to examine the trade-offs incumbents face.

Over the last several decades, many countries have decentralized government programs based on the belief that local officials are more accountable to constituents and have better information regarding the needs and preferences of the local population.[32] Some have argued, however, that in developing countries decentralized policies are particularly prone to capture by local elites,[33] a theory that has been supported by a growing literature that documents evidence consistent with such capture.[34] A key question, therefore, is how local elections influence local government policy decisions in low income settings.

Within the theoretical political economy literature, research such as Banks and Sundaram (1993) has examined how incumbents' private incentives can influence policy decisions, and how reelection motives and benefits of retaining power can discipline such incumbents. Empirically, Ferraz and Finan (2011) use an objective measure of corruption to test the theory that electoral incentives discipline politicians' rent seeking behavior. The authors exploit randomized audits of local Brazilian governments to construct direct measures of appropriation and corruption. They find a strong causal impact between mayors' reelection incentives and the extent to which these mayors engage in corrupt activities.

On the other hand, another strand of the political economy literature, including the probabilistic voting models of Lindbeck and Weibull (1987) and Dixit and Londregan (1996), examines political parties' incentives to target (or redistribute) benefits to swing voters to increase vote shares in an election. Empirically, Schady

[32] See Oates (1999), and references therein, for a review of the "fiscal federalism" literature, which examines the relative costs and benefits of decentralizing certain types of government functions. Empirically, Alderman (2002) provides evidence that local officials and villagers possess local information that is helpful for identifying households that are poor.

[33] See, for instance, Bardhan and Mookherjee (2005) and Drèze and Sen (1989).

[34] See, among others, Reinikka and Svensson (2004), Galasso and Ravallion (2005), Olken (2007) and Araujo et al. (2008).

(2000) and Camacho and Conover (2011) provide evidence consistent with the theory that politicians target public resources to potential swing voters ahead of elections in Peru and Colombia, respectively.[35]

These two strands of literatures suggest that while electoral competition may reduce rent extraction, it may at the same time foster opportunistic "vote buying behavior," such that the net impact on overall resources misallocation remains a priori ambiguous. Our analysis merges these literatures by looking at how electoral competition jointly affects rent extraction and vote buying behaviors. To do so, we examine the way in which politicians distribute benefits across constituents as the degree of electoral competition changes.

We utilize survey data from rural villages in a developing country that contain information about an anti-poverty targeting program. This program established procedures for selecting impoverished households that qualify for a poverty identification card, which entitles cardholders to social assistance benefits. These procedures are implemented within each village and call for an important role of the local officials, such as the village Chief, with the goal of taking advantage of local knowledge of which households are poor. The program was designed so that the number of identified beneficiaries approximately equals the overall poverty rates computed by the central government, in an effort to respect capacity constraints inherent to social assistance programs.

Within each village, a committee of local implementers, appointed by the Chief, interviews households in the village using a short questionnaire. This questionnaire, a Proxy Means Test (PMT), gathers information about household livelihood and

[35]Other papers, such as Drazen and Eslava (2010) and Kneebone and McKenzie (2001), provide evidence that the composition of government spending is modified directly ahead of elections.

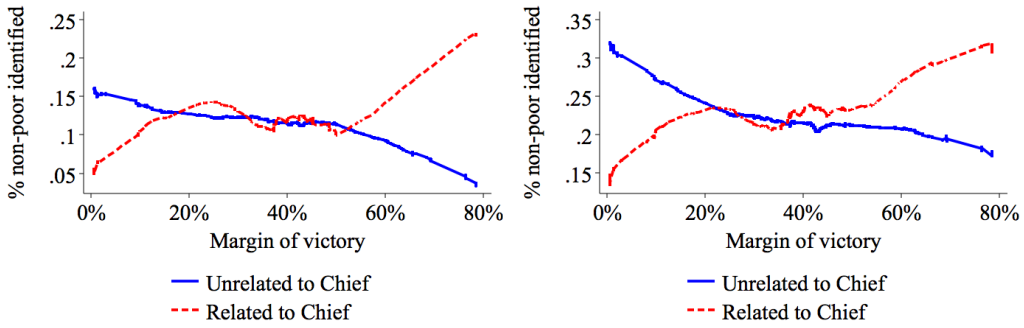
contains an objective decision rule regarding whether or not a household should be selected as a beneficiary. Village authorities, however, have some discretion to adjust the list of beneficiaries at the margin, in order to utilize local knowledge of household circumstances.[36]

As may be expected, these procedures lead to the selection of households that reflect more villagers' perceptions of who are poor, rather than poverty status based upon a consumption indicator, and may lead toward higher satisfaction among beneficiaries (Alatas et al., 2010). In fact, overall targeting accuracy of the program was relatively good, compared to similar programs in other countries. At the same time, however, allowing for discretion of beneficiary selection may open the door for the political leader of the village, the Chief, to influence the selection process and strategically include or exclude households. We study this possibility by looking, *ex post*, at the characteristics of households that were selected.

Our data include several metrics of household livelihood that include per capita daily consumption, villagers' poverty perceptions and the objective PMT score, as well as information regarding households' family ties to the Chief. We draw upon this information to assess the extent to which non-poor relatives of the Chief were included in the list of beneficiaries (a measure we interpret as rent-seeking), and the extent to which non-poor, unrelated households were included (a measure we interpret as opportunistic behavior).

With the targeting information in hand, we then link these data to information regarding the degree of electoral competition in local elections. Periodically, a local Council that represents several villages is elected via proportional voting. Across the

[36]Chapter 3 examines the selection and performance of the appointed local implementers, and provides evidence that the skill level of these implementers matters for determining targeting accuracy.



(a) *Villager perceptions* **(b)** *PMT score*

Figure 2.1: *Share of non-poor households selected as beneficiaries, by incumbent margin of victory*

localities in our data, there is effectively two party competition for these Council seats, and the party that wins majority control of the Council has the ability to appoint the Chiefs in all villages under the Council's jurisdiction. Within the country, it is widely understood that the appointed Chief is typically a strong party representative in the village. Anecdotal evidence suggests that some Chiefs act as an "agent" for the "principal" incumbent party within the village.

As we show below, electoral competition does not seem to influence overall exclusion of poor households and inclusion of non-poor households. While the overall share of "included" non-poor households is relatively low (by international standards), if anything, the share of non-poor beneficiaries is increasing in electoral competition, which is consistent with the evidence presented by Camacho and Conover (2011) for Colombia. An interesting pattern emerges, however, between electoral competition and the *type* of non-poor households identified as beneficiaries. Figure 2.1 shows the share and type of non-poor households that were identified as beneficiaries, by the

incumbent party's margin of victory in the previous Council election.[37] Figure 2.1a identifies households as non-poor using villagers' subjective livelihood perceptions, whereas Figure 2.1b utilizes the targeting program's objective PMT score to classify livelihood status. The dashed (solid) lines depict the probability that a non-poor household related (unrelated) to the Chief was selected as a beneficiary. Figure 2.1a indicates that in villages where the previous Council election was highly competitive (i.e., a margin of victory close to 0), about 5 percent of non-poor households *related to the Chief* were identified as beneficiaries, whereas about 16 percent of non-poor households *unrelated to the Chief* were included as beneficiaries. As the incumbent's margin of victory grows to the point where all seats on the Council were allocated to the incumbent's party, however, around 25 percent of non-poor households related to the Chief were included as beneficiaries, while the proportion of unrelated, non-poor households that were included falls below 5 percent.

This pattern suggests that as electoral competition increases, the Chief trades off rent seeking (i.e., allocating benefits to his or her non-poor relatives) for allocating resources to potential swing voters (i.e., non-poor households that are unrelated to the Chief), while keeping the overall level of resource misallocation constant. To be sure, the pattern in Figure 2.1 is only suggestive; in particular, the analysis does not account for household-level characteristics that might have caused a "borderline non-poor" household to be accidentally included as a beneficiary. Moreover, it does not account for possible omitted variables that could drive the observed pattern.

To examine whether the pattern that emerges in Figure 2.1 is robust, we uti-

[37] Margin of victory is computed as [vote share of winning party] − [vote share of runner-up party]. Thus a margin of victory of 10 percent corresponds to the winner receiving 10 percentage points more votes than the runner-up. The plots in Figure 2.1 were generated using the `lowess` command in Stata.

lize a number of econometric specifications to control for other covariates and the potential endogeneity of electoral competition. We employ four separate specifications, including instrumental variables, (village-specific) fixed-effects and fixed-effects-instrumental variables techniques. We also leverage two metrics for classifying a household as poor or non-poor, and also test whether other household characteristics associated with being selected as a beneficiary, in addition to familial ties to the Chief, are affected by the degree of electoral competition.

Our paper is related to several recent studies that analyze the impact of electoral competition at the sub-national level on reform incentives and program manipulation.[38] Bardhan and Mookherjee (2010) study the impact of electoral competition on land reforms in West Bengal. They find strong evidence of opportunistic behavior by leftist local governments who, in spite their supposed ideology, enact land reforms only if pressured by tight electoral results.[39] Camacho and Conover (2011) document manipulation of Colombia's targeting system (SISBEN) by local authorities, and find that the overall number of households identified as beneficiaries is significantly higher in municipalities where mayoral elections are more competitive. Unlike Bardhan and Mookherjee (2010) and Camacho and Conover (2011), however, we are able to construct a direct measure or rent seeking and political opportunism by leveraging household-level data. This allows us to jointly study rent seeking and opportunistic incentives under a unified framework. Moreover, we go beyond the municipal level of

[38] Another related paper, Niehaus et al. (2011), examines the problem of designing a PMT when the PMT implementers have incentives to manipulate beneficiary selection, and the monitoring of PMT implementers is imperfect. The authors test their theoretical predictions using data from India.

[39] Bardhan and Mookherjee imbed both an electoral (i.e., "Downsian") motive, a moral hazard (e.g., "rent seeking") motive and an ideological motive within their theoretical model. In their empirical analysis, the authors test whether political parties act in a manner consistent with ideology (i.e., adhere to the stated platform) or whether their actions are more consistent with a Downsian (or "quasi-Downsian") framework (i.e., parties are primarily motivated by reelection and rents).

analysis studied in these papers, and document household-level characteristics that reinforce or mitigate the capture effect under electoral competition.

The paper is organized as follows. Section 2.2 presents a simple theoretical model linking electoral competition to both rent seeking and opportunistic vote buying behavior. Section 2.3 describes program context and data. Section 2.4 presents our econometric strategy. Section 2.5 discusses the results, and Section 2.6 concludes.

2.2 Theoretical framework

The purpose of discussing a simple theoretical framework is twofold: first, this framework structures our empirical analysis; and second, the framework illustrates the institutional setting that we study.[40] We consider a two period model ($t = 1, 2$), and three types of players: political parties (of which there are two), local village Chiefs and voters. For simplicity, we assume there is no discounting; period 2 is simply included so that voters have an incentive to vote rationally.

The two political parties share control of a Council that governs the constituency. The parties' share of Council power is based upon a proportional voting rule, in that the number of seats held is equal to the share of votes received from the electorate. At the beginning of the game, one party (the "incumbent") is exogenously endowed with majority control of the Council. The other party will be referred to as the "opposition," but we will primarily focus our attention on the incumbent. The incumbent's utility is strictly increasing in its vote share: the incumbent receives payoff $P(S)$ from receiving a vote share $S \in [0, 1]$. We assume that, over the feasible values of vote shares, $P'(\cdot) > 0$ and $P''(\cdot) < 0$: greater vote share implies more seats on

[40] Some features of the theoretical model also appear in the frameworks of Ferraz and Finan (2011) and Besley (2006).

the Council, more "spoils from office" and more opportunities to exploit power for private gain. The marginal benefit from seats, however, is diminishing: an additional Council seat when all other seats are controlled holds relatively little added value.

The constituency is comprised of an infinite number of villages of measure 1. To be sure, in reality a Council governs around nine villages, but this assumption is made for technical ease. Each village is governed by a Chief that is appointed by the party that controls the Council. For each party, the identities of its Chiefs are fixed within the villages: if the incumbent party receives sufficient votes to maintain majority control of the Council at $t = 2$, then the same Chief will hold power in village v at $t = 1, 2$. This modeling approach corresponds to anecdotal evidence that the Chief is typically the strong party representative in the village.

Each period, the Chief of village v has an "excess" amount of resources from a program that can be allocated to village constituents. Within the context of the targeting program, we view these excess benefits as additional resources that can be allocated to non-poor constituents, in addition to the intended poor constituent beneficiaries.[41] To distribute these excess resources, the Chief of village v can take one of two actions: either allocate the excess benefits to his or her relatives ($g_{vt} = 0$), or allocate these benefits to potential swing voters ($g_{vt} = 1$).

Chiefs are one of two types j: a *rent seeking* type r, or a *party loyalist* type l. The Chief's type is privately known to the Chief. The *ex ante* probability that a Chief (either incumbent or opposition) is type $j = l$ is π. Rent seeking Chiefs receive

[41] As we discuss below, the targeting program does a relatively good job at selecting poor households, suggesting that the Chief and other villagers internalize the benefits the program will bestow upon the impoverished. In practice, however, additional (technically non-eligible) beneficiaries are also selected by the local implementers. Given that program monitoring is only weakly enforced, the Chief and incumbent party have the ability to allocate some additional "excess" benefits to the non-poor.

a payoff of U from allocating the excess benefits to relatives, whereas loyalist Chiefs derive utility from allocating the excess benefits to swing voters. The incumbent party, however, can offer the Chief of village v a costly transfer payment of $T_v \geq 0$ in exchange for the Chief allocating resources to swing voters. These transfers can be interpreted as favors such as promises to share in the Council's "spoils of office" that give the Chief some utility. In equilibrium, to provide incentives to rent seeking Chiefs to influence swing voters to vote for the incumbent, the party will have to offer the same utility to the Chief as allocating excess benefits to relatives, such that $T_v = U$.

Within each village, there is a measure 1 of voters, of which λ are potential (non-poor) swing voters. These voters care only about their expected utility from being allocated the excess benefits. In each period, these voters receive a payoff of $u(g_{vt})$, where $\overline{u} = u(1) > u(0) = 0$.[42] The remaining $(1 - \lambda)$ voters in the village, on the other hand, are not influenced by (g_{v1}, g_{v2}); these voters are thought of as partisans. A fraction $s+\epsilon$ of these voters turn out to vote for the incumbent, where s is a publicly observed signal of how many of the partisans will vote for the incumbent in the village (e.g., the signal could be voters' historical preferences for the party, observed from previous electoral outcomes) and $\epsilon \sim F$ is a noise term that is only observed after the elections occur. Consequently, if swing voters in a fraction $\hat{x}$ of the villages vote for the incumbent party, then the incumbent party's expected payoff from the election is $\int P(s + \hat{x}\lambda + \epsilon)dF(\epsilon)$.

The timing of the game is as follows: in period 1, all Chiefs privately observe

[42] We abstract from the fact that the number of non-poor swing voters λ may exceed the budget of excess benefits. This can be rationalized by assuming that the Chief randomly allocates benefits across the swing voters if $g_{vt} = 1$. Thus, $u(1)$ would incorporate a voter's expected utility from receiving benefits, conditional on $g_{vt} = 1$.

their types. The incumbent party can then privately offer a non-observable contingent transfer payment to some or all of the incumbent Chiefs to provide incentives to these Chiefs to select $g_{v1} = 1$. After observing this private offer, each Chief then selects $g_{v1} \in \{0, 1\}$. Voters observe the policy, vote for the incumbent or opposition party and period 1 payoffs are realized. In period 2, the winning party's Chiefs make the allocation decision $g_{v2} \in \{0, 1\}$, and then voters' period 2 payoffs are then realized.

We solve for the Bayesian Nash Equilibrium via backward induction. In period 2, each Chief will select the policy that maximizes his or her utility, so that $g_{v2}(r) = 0$ and $g_{v2}(l) = 1$. Therefore in period 1, swing voters will vote for the party whose Chief is most likely to be a party loyalist type. Voters can condition their beliefs about the incumbent Chief's type, j_v, by observing the Chief's policy decision g_{v1}. Swing voters will thus vote for the incumbent if and only if $E[j_v = l \mid g_{v1}]\overline{u} \geq \pi\overline{u}$.

In the first period, party loyalist types will always want to select $g_{v1} = 1$, since this decision maximizes their utility. On the other hand, rent seeking type Chiefs can attempt to hide their type from voters and select $g_{v1} = 1$ to pool with the party loyalist types. Given that there are a continuum of villages, voting in any one village does not influence the overall Council election outcome, and thus an individual Chief does not internalize the influence that voting in his village has on the incumbent party's reelection prospects.[43] The incumbent party, however, can provide incentives to the Chief of village v to select $g_{v1} = 1$ by offering this Chief a transfer $T_v = U$ in exchange for this action. Since these transfers are costly, the incumbent party will balance the marginal cost and benefit of influencing the Chiefs. Let $x \in [0, 1]$ denote the measure of villages for which the incumbent party offers the deal of $T_v = U$ to

[43] An alternative model would allow the Chief to internalize some of the benefits that are accrued to his or her party from winning the election.

the Chief. If the voters expect that the incumbent offers the incentive to measure x of the Chiefs, then voters in village v will have the expectation that:[44]

$$prob[j_v = l] \quad = \quad \begin{cases} 0 & \text{if } g_{v1} = 0 \\ \frac{\pi}{\pi+(1-\pi)x} & \text{if } g_{v1} = 1 \end{cases}.$$

Observe that, for any $x \in [0,1]$, swing voters will vote for the incumbent party if $g_{v1} = 1$ because $prob[j_{v1} = l \mid g_{v1} = 1] \geq \pi$. Therefore the incumbent party's problem is

$$\max_{x\in[0,1]} \int P\Big(s + \lambda[\pi + (1-\pi)x] + \epsilon\Big) dF(\epsilon) - xU$$

where again x denotes the share of villages for which the incumbent party incentivizes the Chief to select $g_{v1} = 1$, and $\pi + (1-\pi)x$ denotes the measure of villages for which $g_{v1} = 1$. Assuming an interior solution, it follows that

$$\lambda(1-\pi) \int P'\Big(s + \lambda[\pi + (1-\pi)x^*] + \epsilon\Big) dF(\epsilon) = U,$$

where the left hand side of the equation denotes the party's marginal benefit from increasing the vote share by making the transfer offer to x^* share of village Chiefs, while the right hand side is the marginal cost of making this offer. Then from the

[44] Note that

$$\begin{aligned} prob[j_v = l \mid g_{v1} = 1] &= \frac{prob[j_v = l \wedge g_{v1} = 1]}{prob[g_{v1} = 1]} \\ &= \frac{prob[j_v = l] \cdot prob[g_{v1} = 1 \mid j_v = l]}{prob[g_{v1} = 1]} \\ &= \frac{\pi}{\pi + (1-\pi)x}. \end{aligned}$$

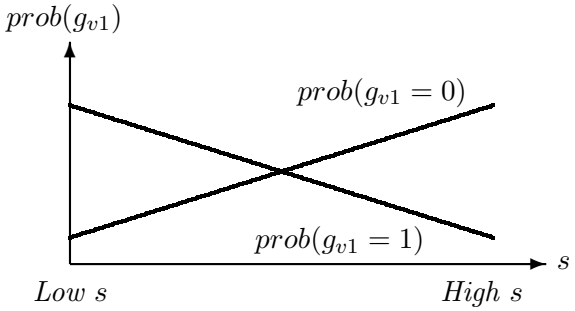


Figure 2.2: *Benefits extracted by Chiefs ($g_{v1} = 0$) and allocated to swing voters ($g_{v1} = 1$)*

Implicit Function Theorem, it follows that

$$\begin{aligned} \frac{\partial x^*}{\partial s} &= -\frac{\lambda(1-\pi)\int P''(\cdot)dF}{[\lambda(1-\pi)]^2\int P''(\cdot)dF} \\ &= -\frac{1}{\lambda(1-\pi)} < 0. \end{aligned}$$

This result states that as the incumbent party observes a higher signal s regarding the number of votes that it will receive in the upcoming election, fewer benefits will be targeted to swing voters (and more benefits will be targeted to the Chiefs' relatives). This follows from the fact that, throughout the constituency, the overall share of swing voters allocated benefits in the first period, $\pi + (1-\pi)x^*$, is decreasing in the expected share of votes the incumbent will receive from the partisan types. Figure 2.2 depicts this relationship graphically: as the incumbent party expects a higher vote share s, then more Chiefs extract the private benefits ($g_{v1} = 0$) because the incumbent party offers fewer benefits to its agents, because the swing voters' support is less valuable.

2.3 Program context and data

This section reviews the data we utilize to perform our empirical analysis, and the setting in which these data were collected. We begin by describing the targeting program in detail, and then proceed to elaborate on the government structure. We then describe our dataset and review the livelihood measures that we construct to partition our sample into poor and non-poor households.

2.3.1 The targeting program

The targeting program that we study aims at identifying impoverished households that reside within the country we consider. Households that are deemed eligible for assistance are allocated an identification card. Across the country, households that received the card ("beneficiaries") have access to free health care at participating medical facilities, which amounts to free health insurance. Recently, these identification cards have increasingly been used by the development community to provide other assistance.

Implementation of the targeting procedures is decentralized and calls for an important role of local officials, with the goal of leveraging local knowledge of households' livelihood conditions. In each village, a committee of around seven villagers (the "local implementers") is tasked with identifying the households that are poor in the village. In theory, the committee of local implementers is selected in a two-stage process: first, the local Chief proposes a list of candidate implementers, and second, villagers vote for their preferred implementers during a village meeting.[45] In prac-

[45] In the village meeting, villagers can propose additional candidates. However, this occurred in fewer than 15 percent of villages, according to the Chief.

tice, however, voting occurs in only around 60 percent of villages, and when voting does occur, the total number of candidates exceeds the total number of implementers by only one or two in the average village. Therefore, the Chief wields considerable influence over the selection of the local implementers.

In each village, the local implementers determine which households are eligible for assistance in three steps: in a first step, the local implementers interview households in the community by administering a short questionnaire (a Proxy Means Test, or PMT) that was developed centrally by the targeting program. As with similar programs around the world, the PMT questions are intended to gather information such as the quality of the household's dwelling and asset ownership. For each question, the household is assigned a point value based on the household's response, and households with a total point score below a certain range are deemed as an eligible beneficiary.

In a second step, local implementers can use their local knowledge of a household's circumstances (e.g., the household recently suffered a crop failure) to adjust the livelihood category indicated by the PMT. Finally, in a third step, the implementers are supposed to display the draft list of beneficiaries in public, and the implementers and Chief are supposed to organize a village-level meeting to receive feedback regarding the list of beneficiaries. According to villagers, however, this third step in the process is often not implemented: only 21 percent of households reported that a draft list of beneficiaries was publicly displayed (36 percent did not know), and only around 29 percent of households reported that a village-level meeting occurred to present and discuss the list of beneficiaries (37 percent did not know).[46]

[46] As mentioned in the Introduction, targeting procedures may lead to the selection of households that *villagers perceive as poor*, as opposed to households that are *technically below a (consumption- or income-based) poverty line*. In fact, the targeting program's objective was to create a PMT that proxies for villagers' poverty perceptions, instead of benchmarking the PMT to a consumption-based poverty metric. This is noteworthy, because villagers' perceptions of who is poor can differ

The decentralized nature of the targeting procedures raises the possibility that local leaders may capture the process and manipulate which households are selected as beneficiaries. Capture can happen through several channels: local implementers are allocated some freedom to subjectively modify the list of beneficiaries; the Chief is involved in overseeing and guiding implementation; and while the national program staff monitor the overall percentages of households that are categorized as poor (by comparing the shares of households selected as beneficiaries with other poverty rates implied by other surveys), there is little-to-no external monitoring of village-level program implementation and whether households selected as beneficiaries are actually poor. Anecdotal evidence and informal accounts of some villagers suggest that nepotism and, in particular, non-poor relatives of the Chief are sometimes included as program beneficiaries.

2.3.2 Electoral competition and government structure

In the country under study, there are a number of political parties that compete for public office. Two parties, however, have emerged as serious contenders for power in the states where the data were collected. We refer to these parties as the Left party (LP) and Right party (RP). The LP is the stronger party; at the time of the data collection, LP controlled the national and several sub-national government bodies. The LP is a socialist-leaning, populist party. The opposition RP, on the other hand, promotes a capitalist, liberal platform and maintains different views on foreign policy.

The elected government body that is most decentralized is the local Council. This Council contains around seven members (depending upon the size of the constituency)

significantly from households that fall below an objective poverty line, as discussed in Alatas et al. (2010).

and governs around nine villages. The Council members are elected according to a closed list proportional voting system, implying that constituents vote for their preferred party and seats on the Council are allocated to parties based upon the overall proportion of votes received in the election.

In the most recent Council election prior to the collection of our data, LP and RP were allocated over 93 percent of all Council seats in our randomly selected sample of villages (described below). The LP received around 60 percent of Council votes in the median village, whereas the RP received around 25 percent of votes (with the remaining votes split between several less popular parties). There is substantial heterogeneity, however, in the degree of electoral competition across Council elections, and in fact, the RP received the majority of votes in some jurisdictions. Figure 2.3 displays the distribution of the margin of victory for the winning party in this election in our sample of jurisdictions.[47] The degree of electoral competition varies considerably: in some jurisdictions, the first-place party won by a margin of less than 1 percent, whereas in other jurisdictions, the margin of victory for LP exceeded 60 percent of votes. We utilize margin of victory as our measure of electoral competition, as opposed to simply the vote share, because the percentage of votes received by the "fringe" of smaller parties varies across Council jurisdictions. This measure allows us to compare the level of competition between the two main parties.

Party control of the Council is important not only for setting jurisdiction-level policies, but also because the Council appoints all local Chiefs under its jurisdiction. As mentioned above, it is generally understood that the party with majority control

[47]Margin of victory is defined as the vote share of the first-place party, minus the vote share of the second-place party. This variable, or a transformation thereof (e.g., 1 − margin of victory), has been used in the literature to measure electoral competition: see, for instance, Camacho and Conover (2011) and Besley and Burgess (2002).

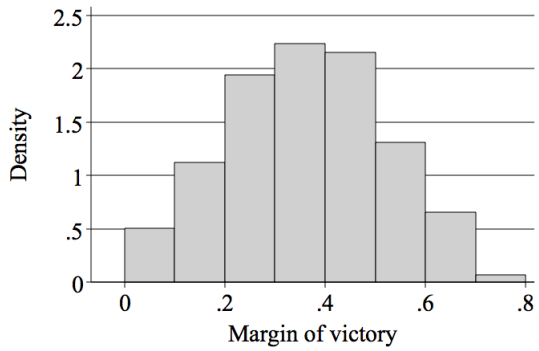


Figure 2.3: *Distribution of margin of victory*

of the Council typically appoints an influential party representative as the Chief in each village. The Chief has then the opportunity to marshal support for his or her party.

2.3.3 Data description

Our analysis draws from a survey that was fielded within several months of the decentralized implementation of the targeting procedures, but before the program's identification cards were allocated to the households. Data were collected in 299 randomly selected villages in four states within the country, where the probability of village selection was proportional to the number of households residing in the village.

Within each village, data were collected from three different populations: households, local implementers and the Chief. Ten households were randomly selected for interview in each village.[48] Household selection was stratified such that five of the

[48]In villages with 250 or more households, the village was partitioned into smaller segments. One village segment was then selected to be the area of interview, again with the probability proportional to the number of households residing in the segment. In one large village, two sets of ten households were selected for interview.

sample households had been selected as beneficiaries, and five of the sample households had not been selected as beneficiaries. Stratification was conducted using the official list of beneficiaries that was created by the local implementers. In total, 2,943 household were interviewed, of which 1,467 (1,476) were (not) beneficiaries.[49] The Chief in each village and local implementers that administered the targeting procedures were also interviewed.[50]

Overall, three different questionnaires were designed to collect various types of information. The household questionnaire was by far the most detailed of the three survey instruments. Much of this questionnaire was derived from the government's socio-economic survey household questionnaire, including the consumption modules. The questionnaire also includes a replica of the targeting program's PMT, and questions regarding whether or not the local village implementers and Chief followed the procedures indicated by the program.

In the spirit of Alatas et al. (2010), the household questionnaire also included two modules that gathered information from each respondent household regarding the other sample households that were interviewed in the village. The first module elicited respondents' perceptions regarding the livelihood status of the other sample households in the village. The respondent was shown a card with the name of the household head written on it and asked,

[49]In each village, four replacement households (two beneficiaries, and two non-beneficiaries) were selected, in the event that a household was unavailable for interview. In 74 percent of these cases where a household member was not available, the reason was because household members were at work or away from the village. Sampling weights are adjusted to reflect the actual number of beneficiaries and non-beneficiaries in each village, and are also adjusted for non-response.

[50]In one (three) [one] village, data were not collected for the local implementers (Chief) [both implementres and Chief] because respondents were unavailable. Data were collected from over 80 percent of local implementers. In 93 percent of villages where a leader was interviewed, the actual Chief was interviewed. If the Chief was unavailable, then the Deputy Chief or other official local leader was interviewed.

> Households in your village can be classified into one of four livelihood levels: very poor, poor, adequate or rich. In your opinion, is (*name of household head*)'s household very poor, poor, adequate or rich?

This module of the questionnaire was extensively piloted and strategically placed towards the beginning of the questionnaire, before any mention of the targeting program, to mitigate response bias.

The second module elicited information regarding whether the other sample households in the village are related to either the Chief, or any of the local implementers. Respondents were asked about other sample households in order to mitigate the possibility of a household misreporting its own relationship status. Each respondent household in the sample was asked if each of the other sample households was related to either the Chief, or any of the local implementers. The answers provided by the respondent households in the village were aggregated to identify the relationship statuses of each sample household in the village.[51] Below, this variable is utilized to gauge the extent to which non-poor relatives of the Chief (and local implementers) are identified as beneficiaries (a measure we will link to rent-seeking), and the extent to which non-poor, unrelated households are identified as beneficiaries (a measure we will link to opportunistic, vote buying behavior).

The local implementer and Chief questionnaires were designed to elicit information regarding characteristics of these individuals. These survey instruments gathered in-

[51] Respondents were informed that familial relations included siblings, parents and children; grandparents and grandchildren; and aunts/uncles, nieces/nephews and cousins. A household-level indicator variable for being related to the Chief was created; this variable equals 1 if either (i) a member of the household is the Chief; (ii) the household respondent identified itself as relative to the Chief; or (iii) two or more other respondents in the village identified the household as related to the Chief (and 0 otherwise). An analogous variable was created with regards to the local implementers. This definition of familial relationships is similar to the "elite connectedness" variable in Alatas et al. (2010).

formation regarding the individual's demographic background and years of completed schooling, and also included a short nonverbal intelligence test (a shortened version of the Raven's Progressive Matrices). The Chief questionnaire also included questions regarding general characteristics about the village.

2.3.4 Livelihood measures and summary statistics

The subsequent analysis requires that we classify each sample household as either poor or non-poor, in order to determine whether a household was inappropriately included or excluded as a beneficiary. To this end, two distinct methods are utilized to partition the households. One of these methods utilizes the targeting program's formal PMT, which was re-administered by the professional enumerators that collected the data utilized in this paper. This benchmark is objective in nature and follows the targeting program's devised method for classifying households as poor.

The targeting program's intended "poverty benchmark," however, was villagers' perceptions of whether or not a household is poor (see Footnote 46). While the program's PMT was designed to proxy for such perceptions, some local implementers modified the livelihood category indicated by the PMT. Therefore, we also classify households as poor or non-poor using villagers' poverty perceptions. To do so, we draw from the poverty perceptions module of the household survey and identify a sample household as poor if the household was classified as poor by half or more of the other respondent households in the village that classified the sample household.[52]

As may be expected, there are differences between the perceptions-based livelihood

[52]If half of the villager respondents perceived a sample household as poor and the other half as non-poor, then the sample household was classified as poor. If fewer than three respondent households reported a livelihood category for a sample household, then the perceptions variable was coded as "missing."

categorization and the PMT. According to the targeting program's PMT, 17.9 percent of the households are eligible beneficiaries (i.e., 17.9 percent of the households received a PMT score below the targeting program's poor threshold), whereas villagers perceive 36.6 percent of the households as poor. The large difference in the overall poverty rate is driven by the poverty threshold: only 15.9 percent of all "PMT poor" households are perceived as not-poor, and only 4.8 percent of all households perceived as poor are classified as "PMT non-poor." This indicates that households identified as poor by the PMT are, in general, also perceived as poor by villagers; villagers, however, perceive that significantly more households are poor, relative to the PMT. Given these differences and that in practice, local implementers were allowed to use local knowledge about poverty perceptions to include additional households, we take the poverty perceptions as the main "poverty benchmark" when classifying households as poor or non-poor. As a robustness exercise, we then estimate our econometric models by partitioning households using the PMT livelihood categorization.

Table 2.1 presents summary statistics for beneficiary (31.1 percent of the population) and non-beneficiary households (68.9 percent of the population). The first set of variables are explicit livelihood measures, such as the household's PMT score. The second set of variables are additional household characteristics that implementers may have (legitimately) considered when determining whether to identify the household as a beneficiary. Finally, the third set of variables capture the degree of social connectedness of the household.

While some types of variables, such as the livelihood measures, differ for the two groups, other variables, such as those listed under the social connectedness heading, are quite similar. The heads of non-beneficiary households are more educated (3.7 years of schooling versus 2.7 years), but educational attainments remain overall fairly

Table 2.1: *Means of household-level variables*

	Beneficiary	Non-beneficiary	Total
Livelihood measures			
Normalized PMT score (0=lowest, 1=highest)	0.40	0.67	0.58
Villagers perceive as non-poor (%)	0.24	0.82	0.63
Per capita daily consumption, USD	1.72	2.40	2.19
Household characteristics			
Number of household members	4.55	4.88	4.78
Years of schooling: Head	2.66	3.69	3.37
Head is female (%)	0.38	0.26	0.29
Ethnic minority (%)	0.05	0.05	0.05
Suffered crop failure: Loss of USD 25+ (%)	0.32	0.46	0.41
Owns 1+ motorcycle (%)	0.21	0.57	0.46
Owns 1+ plough (%)	0.21	0.39	0.34
Owns 1+ tractor (%)	0.02	0.12	0.09
Owns 1+ tv (%)	0.32	0.68	0.57
Social connectedness			
Related to Chief (or is Chief) (%)	0.18	0.18	0.18
Related to implementer (or is implementer) (%)	0.38	0.39	0.39
Number of social organizations	0.11	0.21	0.18
Percent of sample households	0.31	0.69	1.00

Notes: Column 1 (Column 2) reports sample means for the 1,467 (1,476) sample households that were selected (not selected) as beneficiaries by the program. Column 3 reports sample means for the entire sample of 2,943 households. Household variables cover all households, except the villager poverty perceptions variable, which is defined for 2,715 households. The normalized PMT score divides the number of points the household received on the PMT test (as implemented during the data collection) by the total number of possible points. The number of social organizations variable considers membership in an artisan, farmer's, religious, savings or women's association. All means are sample weighted.

low for both groups. There are more female-headed households among beneficiaries, while ethnic minorities have equal presence in both groups. Non-beneficiary households are also more likely to have suffered from crop failure in the previous year, which could be a reflection of the fact that landlessness is more prevalent among beneficiaries.

Overall, households related to the Chief (and related to a local implementer) are evenly distributed between beneficiary and non beneficiary households. Interestingly, however, a majority of households related to the Chief (69 percent) are deemed as non-poor according to the villagers' poverty perceptions (the proportion raises to 86 percent for the PMT). We interpret these benefits allocated to non-poor relatives of the Chief as our measure of rent extraction. Similarly, we interpret the benefits allocated to unrelated, non-poor households as a measure of vote buying behavior. To be sure, some related households may be swing voters. Nevertheless, given that familial ties are quite strong in the country, it remains an unlikely event. More importantly, non-poor, non-relative voters may also contain partisan voters (in addition to swing voters). The vote buying indicator will thus be measured with error, generating a bias of the parameter estimate towards zero.

The second row of the table indicates that 18 percent of non-beneficiary households are perceived as poor (which is similar to other similar programs internationally), whereas 24 percent of beneficiary households are perceived as non-poor. In particular, the analysis shall focus on these non-poor beneficiaries.

Table 2.1 also reports sample means for estimated per capita household consumption. This livelihood metric follows the methodology that the national government utilizes to construct consumption estimates that are used to analyze socio-economic trends. This measure is essentially constructed by summing daily household consump-

tion across a variety of categories and dividing by the number of household members. The natural log of this measure is included in our analysis to control for the fact that local implementers may have selected households, in part, using information about consumption.

In some econometric specifications, we also make use of village and Council characteristics listed in Table 2.2. Data on Chief, local implementers and some village-level characteristics are derived from our survey data, whereas the other characteristics are derived from census data. Margin of victory is generated using publicly available voting data.

2.4 Econometric specification and identification

2.4.1 Baseline specification

We begin our econometric analysis by relating the beneficiary status of a household to a set of household, village and Council jurisdiction variables. We explicitly allow for a relationship between recipient status and the margin of electoral victory in the Council's jurisdiction, which can be affected by the extent to which a household is connected to the Chief. Our baseline estimation equation is given by:

$$\begin{aligned} B_{ivc} &= X_{ivc}\beta_1 + X_{vc}\beta_2 + X_c\beta_3 \\ &\quad + \gamma_1 margin_c + \Big(\delta_0 + \delta_1 margin_c + \nu_{vc} + \nu_c\Big) R_{ivc} + \lambda_{vc} + \lambda_c + \epsilon_{ivc}, \end{aligned} \tag{2.1}$$

where B_{ivc} equals 1 if household i in village v of Council c was selected as a beneficiary (and 0 otherwise); X_{ivc}, X_{vc} and X_c are vectors of household, village, and Council jurisdiction control variables (as listed in Tables 2.1 and 2.2); $margin_c$ is the margin

Table 2.2: *Means of village- and Council jurisdiction-level variables*

	Mean	N
Village characteristics		
# households in village	197.60	299
# households in village / # households in jurisdiction	0.12	299
% of households: Own $\leq$ 0.5 ha of land	0.41	299
% of households: Own $\geq$ 2.0 ha of land	0.21	299
Distance to state administrative center (km)	31.40	299
Chief characteristics		
Age	56.83	295
Percentile score on the cognitive test	0.52	295
Years of completed schooling	5.29	295
Number of social organizations	0.69	295
Average local implementer characteristics		
Average age	42.66	297
Average percentile score on the cognitive test	0.60	297
Average years of completed schooling	6.80	297
Average number of social organizations	0.49	297
Council jurisdiction characteristics		
Incumbent party's margin of victory	0.41	231
% of households: Own 1+ mobile phone	0.27	231
% of households: Own 1+ tv	0.53	231
% of adults: Completed primary school	0.31	231
% of adults: Literate in the local language	0.65	231

Notes: Local implementer characteristics are averaged over the implementers that responded to the implementer survey. All means are sample weighted. Variables are derived from the survey data and the most recent census. The number of social organizations variables were constructed as discussed in the notes for Table 2.1. The sample size for the Council jurisdiction-level variables is smaller than for the village characteristics because in some Council jurisdictions, multiple villages were randomly selected; these variables are defined for all 299 villages in the sample.

of victory of the 2007 elections for the winning political party in Council c; and R_{ivc} equals 1 if household i is related to the Chief (and 0 otherwise). Observe that we use a linear probability model because we are particularly interested in interpreting

interaction effects, and this is difficult to do in a non-linear estimation setting (see, for instance Ai and Norton (2003)).[53]

The model allows for various types of error terms. First, it allows for village- and Council-specific effects in the overall error structure, $(\lambda_{vc}, \lambda_c)$. Second, it allows for village- and Council-specific effects in the association between relationship to Chief and beneficiary status, (ν_{vc}, ν_c). And third, it allows for a household-specific idiosyncratic error term, ϵ_{ivc}. In our base-case estimation, we assume that these terms are, conditional on the other variables in the model, uncorrelated with the margin of victory. Below, we discuss possible deviations from this assumption, and how we modify the estimation accordingly. Specifying the model in this way allows us to recover several parameters of particular interest, namely:

(i) the additional probability δ_0 that relatives of the Chief are selected as beneficiaries when the margin of victory is equal to zero. For non-poor constituents, it should be expected that $\delta_0 < 0$: when the electoral environment is very competitive, relatives of the Chief should be less likely to be selected as a beneficiary;

(ii) the impact δ_1 of the margin of victory on the additional probability that relatives of the Chief are selected as beneficiaries. For non-poor constituents, our theory predicts that $\delta_1 > 0$: as the incumbent's margin of victory increases, relatives of the Chief should be more likely to be selected as a beneficiary, relative to other non-poor households. Moreover, $\delta_0 + \delta_1 > 0$: when the incumbent expects to retain all Council seats, then the Chief's relatives should be strictly more likely to be selected; and

(iii) the impact γ_1 of the margin of victory on the probability that a household unrelated to the Chief is selected as beneficiary. For non-poor constituents, our theory

[53] When we estimate the model using a Probit specification, however, most estimates remain qualitatively unaffected (results are available upon request).

predicts that $\gamma_1 < 0$: swing voters are less likely to be selected as a beneficiary when the incumbent party is confident of reelection.

Observe that these predicted signs for $(\delta_0, \delta_1, \gamma_1)$ should primarily occur for non-poor households. The total number of beneficiaries selected is a "soft" constraint, and the fact that the targeting program under study performs relatively well in including poor households as beneficiaries suggests that local implementers and the Chief internalized the social value of the program for the poorest households in the village. Moreover, opportunistic vote buying behavior and inappropriate allocation of benefits should primarily occur for non-poor households, that is, households that are "not supposed" to receive benefits. Therefore, we shall investigate the signs of these parameters separately for poor and non-poor households.

2.4.2 Augmented specifications

The baseline specification (2.1) assumes that the error terms are uncorrelated with the margin of victory. This assumption may be violated, however, which would lead to inconsistent parameter estimates. We address this problem of identification using instrumental variables and fixed-effects estimators.

Instrumental variables

The first approach accounts for a potential correlation between $(\lambda_{vc}, \lambda_c)$ and $margin_c$, which could come about in the presence of unobserved household preferences in the Council's jurisdiction that are correlated with both the electoral outcome in the local election, and the extent to which program benefits are valued by constituents. In particular, the party that wins the Council election, and the resulting margin of victory, could be related to constituents' preferences over the *local policies* that these parties

support. As previously discussed, the two strong parties that compete in the local elections have different platforms: the dominant party (Left party) holds a socialist/populist platform, whereas the weaker opposition party (Right party) promotes a liberal/capitalist platform.

Given the parties' traditional platforms, we could expect that in areas with strong support for LP's platform, which are the majority of areas, there would be a desire for greater local government intervention to extend benefits to more households, i.e. a positive correlation between $(\lambda_{vc}, \lambda_c)$ and B_{ivc}. Given the nature of electoral competition, this would also imply a positive correlation between $(\lambda_{vc}, \lambda_c)$ and $margin_c$, which would bias γ_1 upwards (i.e., towards zero) since the non-poor should be more likely to receive benefits; the bias imparted to δ_1 remains less clear, as we would not expect there to be any "additional" probability of receiving benefits associated with being related to the Chief. Similarly, if the dominant party in a particular locality is particularly good at delivering benefits to swing voters, we would also expect a positive correlation between $(\lambda_{vc}, \lambda_c)$ and B_{ivc}. Note that in both cases, the bias imparted under these conditions is away from the direction predicted by our theory, suggesting that in the basic model we are likely to underestimate the impacts of electoral competition on benefits.

Nevertheless, we address this potential bias by building on Bardhan and Mookherjee (2010), who examine local elections in India and base their instrumental variables approach on the fact that voters have inherent biases for certain political parties, based for instance on historical traditions, ideology and policy positions that are unrelated to local policy (i.e., national-level foreign policy). Specifically, when analyzing the impact of local elections, Bardhan and Mookherjee use as instruments outcomes of national elections and previous local elections to account for these underlying, un-

observed factors. These instruments correlate with national-level policies and the innate preferences of voters, but should be orthogonal to the particular local policies that will affect the upcoming local election. In full similarity, we use two instruments for our primary measure of electoral competition:

(i) The lagged margin of victory in the Council elections. The Council election cycle that we utilize to construct the margin of victory $margin_c$ represents the first election that followed major decentralization efforts to empower Councils to determine local policies. Prior to this election, the local Councils had very limited power, essentially no budget and no ability to officially allocate benefits or assistance to constituents. As such, prior Council elections were more symbolic, as the Council was not relevant for local policy. The lagged margin of victory in the Council elections is therefore likely to reflect the heterogeneity of constituents' preferences for national-level parties and policies, as opposed to local policies (which were not yet well established) or outcomes of the local party machine.

(ii) The margin of victory in the most recent national elections. The most recent national elections were also conducted via a closed list, proportional election system. We compute the average vote shares for each party, where the average is taken over the district where Council c is located. Since preferences for political parties systematically vary geographically, district-averaged margin of victory should capture the dispersion of regional voters' preferences for political parties' national-level policies.

Based on these two variables, we define four variables that we use to instrument for the margin of victory $margin_c$ and its interaction $margin_c \cdot R_{ivc}$:[54]

- $margin_c^{C_{-1}}$: The lagged margin of victory that occurred in the previous Council

[54]See Wooldridge (2002), which explains why $[margin_c^{C_{-1}}, margin_c^{A}] \cdot R_{ivc}$ are valid instruments for $margin_c \cdot R_{ivc}$ (conditional on $[margin_c^{C_{-1}}, margin_c^{A}]$ being valid instruments for $margin_v$).

Table 2.3: *Margin of victory regressions*

	[1]	[2]	[3]
$margin_c^{C-1}$	0.401***		0.357***
	(0.071)		(0.069)
$margin_c^A$		0.609***	0.506***
		(0.123)	(0.111)
F-statistic	31.982	24.709	22.889
R^2	0.386	0.347	0.444
N	231	231	231

Notes: Each column reports results for the OLS specification that regresses $margin_c$ on the instruments listed on the table and Council jurisdiction-level controls X_c listed in Table 2.2 and state fixed-effects. Robust standard errors are reported in parenthesis. The star level *** corresponds to statistical significance at 1%.

election.

- $margin_c^A$: The margin of victory in the most recent national election, averaged at the district-level.
- $R_{ivc} \cdot margin_c^{C-1}$: The interaction between $margin_c^{C-1}$ and R_{ivc}.
- $R_{ivc} \cdot margin_c^A$: The interaction between $margin_c^A$ and R_{ivc}.

Table 2.3 reports estimates from an OLS regression of the incumbent party's most recent margin of victory in the Council election on the instruments discussed above and Council controls. Estimates indicate that the instruments are significant in predicting the degree of electoral competition in the most recent Council election.

Village-level fixed-effects

The second approach we use to account for potential omitted variables biases is fixed-effects. We estimate a specification of equation (2.1) that includes a full set of village dummy variables, which will capture all village- and Council-level variables that affect the *level* of benefits distributed. We therefore estimate the model

$$B_{ivc} = X_{ivc}\beta_1 + \Big(\delta_0 + \delta_1 margin_c + \nu_{vc} + \nu_c\Big)R_{ivc} + \tau_{vc} + \epsilon_{ivc} \qquad (2.2)$$

where τ_{vc} are the village fixed-effects.[55] The validity of this approach rests on the assumption that (ν_{vc}, ν_c) are mean-zero errors that are uncorrelated with $margin_c$ (conditional on the other variables in the model). While this assumption seems plausible to us, it is possible that it could be violated. A violation of this assumption could occur, for instance, if more benefits are distributed to the Chief's relations and fewer benefits to swing voters, while concurrently the Chief "delivers" more votes to swing the election to the incumbent party. Given that each village contains only about 12 percent of constituents that vote in the Council election, this seems somewhat unlikely.

Note that in this specification, we cannot identify the direct effect of village- or Council-level variables. We cannot therefore estimate the effect γ_1 of $margin_c$ on swing voters' likelihood of receiving benefits, but only the additional probability that Chiefs' relatives are awarded benefits, compared to non-relatives.

[55] Results remain very similar if Council fixed-effects are used, and village-level controls are retained in the model.

Fixed-effects instrumental variables

The third approach combines these two previous methods and estimates a fixed-effects instrumental variables model. While this increases the set of assumptions required for the approach to be valid, it allows us to assess the robustness of our results to controlling for the alternative sources of bias. In this specification, we use $R_{ivc} \cdot [margin_c^{C-1}, margin_c^A]$ to instrument for $R_{ivc} \cdot margin_c$.

2.5 Empirical results

We begin by assessing whether the overall prevalence of mistargeting of poor and non-poor voters varies systematically with the degree of electoral competition. To investigate this possibility, we partition our sample into poor and non-poor household subsamples, and then estimate the following modified version of (2.1)

$$B_{ivc} = X_{ivc}\beta_1 + X_{vc}\beta_2 + X_c\beta_3 + \gamma margin_c + \delta R_{ivc} + \lambda_{vc} + \lambda_c + \epsilon_{ivc} \quad (2.3)$$

separately for these two subsamples. Equation (2.3) is identical to our baseline equation (2.1), except the random parameter for R_{ivc} has been replaced with a fixed parameter δ. In this specification, the parameter γ captures how the degree of electoral competition influences the overall probability that a household was selected as a beneficiary (regardless of whether or not the household is related to the Chief).

Results for this model are reported in Table 2.4. Columns 1-3 (Columns 4-6) report parameter estimates for the poor (non-poor) subsample of households, where households were partitioned using villagers' perceptions of which households are poor.[56]

[56]Qualitative results are the same if households are partitioned according to the PMT. The coeffi-

Table 2.4: *Overall probability of being selected as a beneficiary*

	Poor household sample			Non-poor household sample		
	[1]	[2]	[3]	[4]	[5]	[6]
	Base	IV (1)	IV (2)	Base	IV (1)	IV (2)
R_{ivc}	0.043	0.044	0.047	0.009	0.009	0.010
	(0.045)	(0.044)	(0.045)	(0.018)	(0.017)	(0.017)
$margin_c$	0.027	0.098	0.264	-0.108**	-0.173*	-0.062
	(0.105)	(0.191)	(0.220)	(0.055)	(0.095)	(0.120)
X_{ivc}	Yes	Yes	Yes	Yes	Yes	Yes
(X_{vc}, X_c)	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	No	No	Yes	No	No	Yes
R^2	0.153	0.153	0.170	0.152	0.151	0.159
N	1,277	1,277	1,277	1,391	1,391	1,391

Notes: Each column reports parameter estimates of a linear probability model in which the dependent variable equals 1 if the household was selected as a beneficiary and 0 otherwise. Columns 1-3 (Columns 4-6) report parameter estimates for the subsample of households that are perceived as poor (non-poor) by other households in the village. Column titles correspond to the respective econometric specification that was used to estimate (2.3). Household-level controls, X_{ivc}, are listed in Table 2.1. Village- and Council jurisdiction-level controls, (X_{vc}, X_c), are listed in Table 2.2. Standard errors are clustered by village and reported in parenthesis. Observations are sample weighted. The star levels ***, ** and * correspond to statistical significance at 1%, 5% and 10%, respectively.

Columns 1 and 4 report estimates of (2.3) using the OLS estimator, while Columns 2, 3, 5 and 6 report estimates for the instrumental variables model in which $margin_c$ is instrumented for by $(margin_c^{C_{-1}}, margin_c^A)$.

In our data, the degree of electoral competition varies considerably across the states within our sample, but varies to a lesser extent within states. Consequently, we estimate the instrumental variables specification both omitting state fixed-effects (Columns 2 and 5) and including them (Columns 3 and 6). The specification that

cient for $margin_c$, however, is negative and insignificant in Column 5.

omits the state fixed-effects utilizes the full variability of the electoral outcomes across all Councils, but may also omit other state-specific characteristics that influence targeting. When the state fixed-effects are included, the estimates utilize only the within-state variation of competition to identify the parameters of interest, which remains smaller.

Columns 1-3 in Table 2.4 suggest that the incumbent's margin of victory does not influence the probability that a poor household is selected as a beneficiary; the γ coefficient is insignificant across all three specifications. On the other hand, Columns 4-6 provide mixed evidence regarding the relationship between electoral competition and the share of non-poor households (inappropriately) included as beneficiaries. Columns 4 and 5 indicate that the share of non-poor beneficiaries is decreasing in the incumbent's margin of victory, which is consistent with the findings of Camacho and Conover (2011). Nevertheless, when state fixed-effects are incorporated into the instrumental variables model, the parameter estimate γ remains negative but becomes insignificant. Observe that the parameter estimates for δ suggest that relatives of the Chief have the same likelihood of being selected as a beneficiary, when compared to poor non-relatives.

Next, we investigate whether the incumbent party's margin of victory influences the *composition* of the types of poor and non-poor households selected as beneficiaries, in a manner consistent with rent seeking and opportunistic vote buying. Table 2.5 reports the key parameter estimates $(\gamma_1, \delta_0, \delta_1)$ for each of the specifications discussed in Section 2.4, separately for both the poor and non-poor household subsamples (again, partitioned according to respondents' perceptions of households' poverty status). The instrumental variables specification is again estimated both without (IV (1)) and with (IV (2)) state fixed-effects.

Table 2.5: *Testing the rent seeking and vote buying mechanisms*

	Poor household sample					Non-poor household sample				
	[1] Base	[2] IV (1)	[3] IV (2)	[4] FE	[5] FE-IV	[6] Base	[7] IV (1)	[8] IV (2)	[9] FE	[10] FE-IV
R_{ivc}	-0.133	0.011	0.065	-0.144	-0.032	-0.082*	-0.183**	-0.178**	-0.104*	-0.275*
	(0.128)	(0.208)	(0.222)	(0.170)	(0.166)	(0.045)	(0.085)	(0.088)	(0.057)	(0.149)
$R_{ivc} \cdot margin_c$	0.447	0.081	-0.046	0.378	0.167	0.228**	0.482**	0.473**	0.301**	0.765**
	(0.298)	(0.515)	(0.562)	(0.423)	(0.401)	(0.105)	(0.205)	(0.212)	(0.135)	(0.381)
$margin_c$	-0.036	0.071	0.258			-0.151**	-0.274**	-0.166		
	(0.115)	(0.222)	(0.248)			(0.061)	(0.107)	(0.129)		
X_{ivc}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(X_{vc}, X_c)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
State FEs	No	No	Yes	No	No	No	No	Yes	No	No
Village FEs	No	No	No	Yes	Yes	No	No	No	Yes	Yes
prob$(\delta_0 + \delta_1 = 0)$	0.089	0.768	0.956	0.385	0.579	0.028	0.015	0.020	0.022	0.040
R^2	0.156	0.154	0.170	0.399	0.101	0.154	0.150	0.159	0.313	0.228
N	1,277	1,277	1,277	1,297	1,297	1,391	1,391	1,391	1,418	1,418

Notes: Each column reports parameter estimates of a linear probability model in which the dependent variable equals 1 if the household was selected as a beneficiary and 0 otherwise. Columns 1-5 (Columns 6-10) report parameter estimates for the subsample of households that are perceived as poor (non-poor) by other households in the village. Column titles correspond to the respective econometric specifications discussed in Section 2.4. Household-level controls, X_{ivc}, are listed in Table 2.1. Village and Council controls, (X_{vc}, X_c), are listed in Table 2.2. In Columns 1-4 and 6-9, standard errors are clustered by village and observations are sample weighted. The star levels ***, ** and * correspond to statistical significance at 1%, 5% and 10%, respectively.

Overall, results are consistent with the theoretical predictions. The reported estimates provide strong support to strategic rent seeking behavior for the non-poor subsample (Columns 6-10), whereas family ties to the Chief does not seem to affect beneficiaries' selection among the poor, which confirms the hypothesis that the "soft" budget constraint allows for covering most poor households. Results also have strong economic meaning. The negative coefficients for R_{ivc} in the regressions for non-poor households indicate that, with a zero margin of victory, non-poor households with family ties to the Chief are around 10 to 20 percentage points *less* likely to be selected as beneficiaries, relative to other non-poor households. The positive coefficients associated with the interaction $R_{ivc} \cdot margin_c$, and the fact that $\delta_0 + \delta_1 > 0$, also indicate that as the incumbent party's margin of victory increases, the Chief's non-poor family members become relatively more likely to be identified as beneficiaries. In fact, when the margin of victory is large (close to one), the situation almost reverses: non-poor households with family ties to the Chief are 15 to 30 percentage points *more* likely to be selected as beneficiaries, relative to other non-poor households.

On the other hand, results provide somewhat mixed support to opportunistic vote buying behavior: the negative coefficient on $margin_c$ suggests that a 10 percentage point increase in the winning party's margin of victory is associated with about a 1.5 to 2.5 percentage point decline in the probability that a non-poor, unrelated household is selected as a beneficiary. When including the state fixed-effects in Column 8, however, the γ_1 parameter becomes insignificant, suggesting that the data may not have enough power to capture the within-state variability in the effect of $margin_c$ on the probability that non-poor, unrelated households are selected as beneficiaries.[57]

[57] If state fixed-effects are incorporated into the baseline specification in Column 6, the parameters $(\delta_0, \delta_1, \gamma_1)$ are all statistically significant at the 10 percent or 5 percent level.

2.5.1 Robustness

We now turn to investigate the robustness of the results presented in Table 2.5. We begin by delving deeper into the possibility that the statistical relationships observed in Table 2.5 are spurious in nature. Up until this point, we have imposed the assumption that $margin_c$ only affects the parameter of one household-level variable: R_{ivc}. Under the theoretical framework presented in Section 2.2, this assumption may seem natural: the incumbent's margin of victory shouldn't necessarily influence the probability that, say, female-headed households, households that suffered a crop crisis or television-owning households are included as beneficiaries. If the results in Table 2.5 are indeed spurious, and the hypothesized political economy mechanisms are not active, then it may be the case that $margin_c$ also influences the marginal effects of other household-level variables included in X_{ivc}.

To test this possibility, we augment the baseline model (2.1) and estimate

$$\begin{aligned} B_{ivc} &= X_{ivc}\overbrace{\Big(\eta_0 + \eta_1 margin_c + \zeta_{vc} + \zeta_c\Big)}^{\text{Replaces } \beta_1} + X_{vc}\beta_2 + X_c\beta_3 \\ &\quad + \gamma_1 margin_c + \Big(\delta_0 + \delta_1 margin_c + \nu_{vc} + \nu_c\Big)R_{ivc} + \lambda_{vc} + \lambda_c + \epsilon_{ivc}, \end{aligned}$$

where the elements $\{\eta_0, \eta_1, \zeta_{vc}, \zeta_c\}$ are vectors of dimension $\dim(X'_{ivc})$ and have an analogous interpretation as $(\delta_0, \delta_1, \nu_{vc}, \nu_c)$. The key question is whether $\eta_1 \neq 0$, i.e. whether, for some household-level controls, $margin_c$ is influential in determining the variable's parameter value. In addition to estimating this augmented baseline specification, we also augment the other specifications that were estimated in the results reported in in Table 2.5.[58]

[58] For the instrumental variables specifications, we instrument for $margin_c \cdot [1, R_{ivc}, X_{ivc}]$ using the

Estimates for (δ_1, η_1) are reported in Table 2.6. Observe that Table 2.6 reports estimates from the model that incorporates all interaction terms *jointly*. Columns 1-5 display estimates for the poor household subsample, and Columns 6-10 for the non-poor subsample. For non-poor households, the main qualitative conclusions regarding rent-seeking still hold across specifications: after including all of the interaction terms, the interaction of the $margin_c$ and R_{ivc} remains significant under all specifications. In contrast, *all other terms are insignificant*: only two of the 55 parameters are significant, and only at the 10 percent level. This suggests that, for non-poor households, $margin_c$ primarily affects the household-level parameters through familial ties with the Chief, R_{ivc}. Observe, in particular, that the interaction of the margin of victory with households' relations with the local implementers is not significant. This confirms the role of the village Chief as an influential and visible party representative in the village.

For poor households, the only interaction term that is significant across multiple (four) specifications is the one concerning ethnic minorities: a 10 percentage point increase in margin of victory for the incumbent implies that poor, ethnic minority households are about 10 to 20 percentage points less likely to be included as a beneficiary. This suggests that a narrow margin of victory may benefit against discrimination of ethnic minorities.

Next, we consider if results are sensitive to the way in which the poor and non-poor households subsamples are divided. Tables 2.4-2.6 utilized data that partitioned households using an inherently *subjective* welfare metric, villager perceptions. Conceptually, such perceptions may vary systematically across villages. To check for robustness, we partition households into poor and non-poor subsamples using the

vectors $margin_c^{C-1} \cdot [1, R_{ivc}, X_{ivc}]$ and $margin_c^{A} \cdot [1, R_{ivc}, X_{ivc}]$.

Table 2.6: *Interaction between $margin_c$ and other household-level controls*

	Poor households				
	[1] Base	[2] IV (1)	[3] IV (2)	[4] FE	[5] FE-IV
$R_{ivc} \cdot margin_c$	0.414	0.044	-0.119	0.298	0.069
	(0.286)	(0.501)	(0.537)	(0.418)	(0.418)
# household members$_{ivc} \cdot margin_c$	-0.073*	-0.043	0.000	-0.054	-0.020
	(0.044)	(0.097)	(0.097)	(0.055)	(0.092)
Years of schooling$_{ivc} \cdot margin_c$	0.000	-0.093	-0.105	-0.006	-0.021
	(0.036)	(0.080)	(0.078)	(0.044)	(0.057)
Head is female$_{ivc} \cdot margin_c$	0.128	-0.532	-0.589	-0.159	-0.136
	(0.192)	(0.437)	(0.412)	(0.226)	(0.350)
Ethnic minority$_{ivc} \cdot margin_c$	-1.079***	-2.154**	-1.817**	-1.173*	-0.441
	(0.370)	(0.980)	(0.898)	(0.657)	(1.494)
Suffered crop failure$_{ivc} \cdot margin_c$	0.007	0.180	0.144	0.288	0.332
	(0.227)	(0.478)	(0.457)	(0.276)	(0.373)
Owns 1+ motorcycle$_{ivc} \cdot margin_c$	0.450	0.261	0.403	0.349	0.279
	(0.298)	(0.657)	(0.664)	(0.405)	(0.472)
Owns 1+ plough$_{ivc} \cdot margin_c$	0.010	0.231	0.032	-0.115	-0.048
	(0.232)	(0.396)	(0.408)	(0.300)	(0.372)
Owns 1+ tractor$_{ivc} \cdot margin_c$	1.092*	0.830	0.683	0.385	0.105
	(0.645)	(1.022)	(0.993)	(0.727)	(0.745)
Owns 1+ tv$_{ivc} \cdot margin_c$	-0.181	0.418	0.264	-0.173	0.470
	(0.234)	(0.519)	(0.527)	(0.274)	(0.333)
Related to implementer$_{ivc} \cdot margin_c$	0.070	0.895*	0.795	0.158	0.350
	(0.213)	(0.528)	(0.508)	(0.262)	(0.376)
# social organizations$_{ivc} \cdot margin_c$	0.048	-0.453	-0.516	0.394	0.291
	(0.283)	(0.677)	(0.681)	(0.369)	(0.540)
X_{ivc}	Yes	Yes	Yes	Yes	Yes
(X_{vc}, X_c)	Yes	Yes	Yes	No	No
State FEs	No	No	Yes	No	No
Village FEs	No	No	No	Yes	Yes
R^2	0.168	0.119	0.138	0.405	0.092
N	1,277	1,277	1,277	1,297	1,297

Note: Table continued on the following page.

Table 2.6: *Continued*

	Non-poor households				
	[6] Base	[7] IV (1)	[8] IV (2)	[9] FE	[10] FE-IV
$R_{ivc} \cdot margin_c$	0.220*	0.627**	0.626**	0.276*	0.805**
	(0.124)	(0.247)	(0.253)	(0.160)	(0.399)
# household members$_{ivc} \cdot margin_c$	-0.036	-0.001	-0.004	-0.054*	-0.016
	(0.025)	(0.038)	(0.039)	(0.029)	(0.068)
Years of schooling$_{ivc} \cdot margin_c$	-0.009	-0.008	-0.011	-0.007	-0.011
	(0.014)	(0.022)	(0.022)	(0.018)	(0.039)
Head is female$_{ivc} \cdot margin_c$	-0.179	-0.187	-0.208	-0.243	-0.562
	(0.115)	(0.199)	(0.197)	(0.157)	(0.344)
Ethnic minority$_{ivc} \cdot margin_c$	0.314	0.186	0.197	0.171	0.632
	(0.204)	(0.310)	(0.318)	(0.353)	(0.690)
Suffered crop failure$_{ivc} \cdot margin_c$	0.109	-0.033	-0.060	0.043	-0.231
	(0.103)	(0.188)	(0.190)	(0.143)	(0.329)
Owns 1+ motorcycle$_{ivc} \cdot margin_c$	0.034	0.066	0.077	0.058	0.039
	(0.093)	(0.199)	(0.200)	(0.117)	(0.294)
Owns 1+ plough$_{ivc} \cdot margin_c$	-0.117	-0.271	-0.310*	-0.110	-0.449
	(0.112)	(0.184)	(0.180)	(0.188)	(0.335)
Owns 1+ tractor$_{ivc} \cdot margin_c$	0.092	0.063	0.076	0.185	0.136
	(0.105)	(0.186)	(0.189)	(0.197)	(0.439)
Owns 1+ tv$_{ivc} \cdot margin_c$	0.036	0.129	0.119	0.009	0.108
	(0.094)	(0.171)	(0.169)	(0.114)	(0.309)
Related to implementer$_{ivc} \cdot margin_c$	-0.015	-0.070	-0.075	-0.073	0.076
	(0.111)	(0.186)	(0.188)	(0.154)	(0.326)
# social organizations$_{ivc} \cdot margin_c$	-0.105	-0.139	-0.109	-0.068	-0.367
	(0.091)	(0.170)	(0.172)	(0.116)	(0.295)
X_{ivc}	Yes	Yes	Yes	Yes	Yes
(X_{vc}, X_c)	Yes	Yes	Yes	No	No
State FEs	No	No	Yes	No	No
Village FEs	No	No	No	Yes	Yes
R^2	0.159	0.149	0.156	0.318	0.225
N	1,391	1,391	1,391	1,418	1,418

Note: See the notes for Table 2.5.

objective PMT point score computed by the professional enumerators that executed the data collection. A household is classified as poor (or eligible) if the household's PMT score is below the targeting program's point threshold.

Table 2.7 presents these estimates. In Columns 1-5 we show results for technically eligible households, and in Columns 6-10 we show results for technically non-eligible households.[59] While the estimates, for non-eligible households, have the correct sign, their significance drops. In particular, the interaction between margin of victory and households' relation with the Chief loses significance under the IV estimations, and the estimates for vote buying (γ_1) are only significant in the baseline specification (Column 6). In full similarity with the main analysis, estimates for eligible households are all insignificant.

The drop in significance does not necessarily undermine our findings, but rather suggests that *where the eligibility threshold is set* in splitting the sample is of high importance. The PMT only identified 17.9 percent of households as poor, against 36.6 percent for the perception measure. Recall that 31.1 percent of the population in our sample villages is a beneficiary. Thus, the PMT splits the sample at a threshold around which most households are beneficiaries, and includes in the sample of non-eligible households many households that, independently from their relation with the Chief, are deemed poor by everybody in the village and have been fairly included among the beneficiaries. Such an inclusion obviously affects the ability to capture rent seeking and vote buying effects.

To investigate this issue, we recalibrated the PMT beneficiary eligibility cutoff to reflect the overall percentage of households that were identified as poor by lo-

[59] In this specification, X_{iv} includes an indicator for whether or not villagers perceive household i as poor.

cal implementers (31.1 percent). The household sample was then partitioned such that households receiving a PMT score below the recalibrated cutoff were classified as eligible. Parameter estimates for these specifications are reported in Table 2.8. For the non-poor subsample, the baseline, fixed-effects and fixed-effects instrumental variables specifications (Columns 6, 9 and 10) provide strong support for our theoretical predictions. In the two instrumental variables specifications (Columns 7 and 8), however, the key parameter estimates generally have signs that correspond to the theoretical predictions, but are statistically insignificant, thus providing mixed evidence in support of the countervailing incentives.

2.6 Conclusion

This paper tests the degree to which policymakers' private rent seeking and vote buying behaviors respond to electoral competition. The analysis employs survey data that contain information regarding a decentralized program in a developing country that selected households eligible for social assistance benefits intended for the poor. We find that the overall level of resource misallocation is invariant or, if anything, increasing in the degree of political competition. The composition of *who benefits* from misallocation, however, depends upon the level of competition: as electoral competition increases, local officials trade off extracting private benefits for allocating resources to potential swing voters. Non-poor relatives of the local Chief are about 10 to 20 percentage points *less* likely to be allocated benefits, relative to other non-poor households, when the previous election outcome was competitive. When the incumbent party won the previous election with a large margin of victory, however, the situation reverses: non-poor relatives of the local Chief are about 15

to 30 percentage points *more* likely to be allocated benefits. Some evidence, though weaker, also suggests that opportunistic vote buying behavior increases in response to competition: non-poor voters that are unrelated to the Chief are about 15 to 20 percentage points more likely to be identified as beneficiaries if expected electoral competition is high. We also find that other household characteristics associated with receiving benefits are unaffected by electoral competition, suggesting that favoritism works mainly through family ties.

Table 2.7: *Identifying poor and non-poor households using the PMT score*

	Poor household sample					Non-poor household sample				
	[1] Base	[2] IV (1)	[3] IV (2)	[4] FE	[5] FE-IV	[6] Base	[7] IV (1)	[8] IV (2)	[9] FE	[10] FE-IV
R_{ivc}	-0.032	-0.203	-0.220	-0.215	-0.149	-0.079	-0.065	-0.055	-0.109*	-0.176
	(0.139)	(0.288)	(0.278)	(0.219)	(0.327)	(0.052)	(0.092)	(0.097)	(0.065)	(0.114)
$R_{ivc} \cdot margin_c$	0.181	0.607	0.622	0.491	0.321	0.245**	0.210	0.189	0.345**	0.514*
	(0.302)	(0.681)	(0.646)	(0.488)	(0.758)	(0.124)	(0.233)	(0.246)	(0.161)	(0.294)
$margin_c$	0.118	-0.239	-0.185			-0.128**	-0.121	0.062		
	(0.151)	(0.292)	(0.280)			(0.065)	(0.116)	(0.144)		
X_{ivc}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(X_{vc}, X_c)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
State FEs	No	No	Yes	No	No	No	No	Yes	No	No
Village FEs	No	No	No	Yes	Yes	No	No	No	Yes	Yes
prob$(\delta_0 + \delta_1 = 0)$	0.401	0.311	0.283	0.338	0.697	0.035	0.319	0.380	0.024	0.067
R^2	0.200	0.189	0.218	0.583	0.141	0.303	0.303	0.310	0.403	0.387
N	660	660	660	675	675	2,008	2,008	2,008	2,040	2,040

Notes: See the notes for Table 2.5. Columns 1-5 (Columns 6-10) of this table report parameter estimates for the subsample of households that received a PMT score indicating that the household is poor (non-poor).

Table 2.8: *Estimates after recalibrating the PMT eligibility cutoff*

	Poor household sample					Non-poor household sample				
	[1] Base	[2] IV (1)	[3] IV (2)	[4] FE	[5] FE-IV	[6] Base	[7] IV (1)	[8] IV (2)	[9] FE	[10] FE-IV
R_{ivc}	0.033	-0.036	0.002	-0.015	-0.008	-0.110*	-0.078	-0.074	-0.168**	-0.234*
	(0.108)	(0.244)	(0.241)	(0.142)	(0.209)	(0.057)	(0.093)	(0.094)	(0.072)	(0.124)
$R_{ivc} \cdot margin_c$	-0.019	0.157	0.076	0.117	0.055	0.329**	0.250	0.232	0.445**	0.630**
	(0.245)	(0.595)	(0.587)	(0.334)	(0.515)	(0.137)	(0.237)	(0.243)	(0.178)	(0.313)
$margin_c$	0.031	0.039	0.082			-0.132**	-0.161	0.023		
	(0.122)	(0.265)	(0.276)			(0.062)	(0.109)	(0.129)		
X_{ivc}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(X_{vc}, X_c)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
State FEs	No	No	Yes	No	No	No	No	Yes	No	No
Village FEs	No	No	No	Yes	Yes	No	No	No	Yes	Yes
prob$(\delta_0 + \delta_1 = 0)$	0.929	0.735	0.825	0.624	0.883	0.012	0.244	0.296	0.016	0.042
R^2	0.194	0.194	0.215	0.478	0.159	0.256	0.256	0.266	0.377	0.343
N	1,117	1,117	1,117	1,141	1,141	1,551	1,551	1,551	1,574	1,574

Notes: See the notes for Table 2.5. Columns 1-5 (Columns 6-10) of this table report parameter estimates for the subsample of households that received a PMT score indicating that the household is poor (non-poor), after adjusting the PMT point cutoffs.

Chapter 3

3.1 Introduction

Across the world, politicians and government officials appoint bureaucrats and public servants to assist in administering policies. In many developing countries, for instance, local officials hire villagers to execute tasks such as managing the village budget or building infrastructure. Little is known, however, regarding the extent to which candidates are appointed based on skill and merit, or whether considerations such as nepotism, cronyism or information asymmetries influence hiring decisions. In low income, low human capital settings, suboptimal bureaucrat selection may have a large, adverse effect on program performance, as unqualified appointees may not possess the basic skills that are necessary to effectively implement the policy.

This paper utilizes data from a decentralized public program in a developing country in which a committee of villagers was appointed by the local Chief to identify households eligible for social assistance benefits intended for the poor (an anti-poverty "targeting" program). Data were collected from the appointees, a random sample of candidate public servants that were not appointed and the Chiefs themselves, which allows for the study the types of individuals that the Chiefs tended to hire, and also

the types of individuals that performed best at the task.

Local governments in developing countries hire villagers to implement a variety of programs and execute official administrative government functions. The academic literature has examined the policy outcomes of several such programs. For instance, Olken (2007) studies road construction projects in Indonesia that were coordinated by a local implementation team that had the authority to hire villagers to assist with construction. Besley et al. (2011) and Niehaus et al. (2011), among others, study India's Below the Poverty Line (BPL) program, in which local governments appoint village enumerators that determine households' eligibility for a BPL identification card that grants access to social assistance programs.[60] Camacho and Conover (2011) examine a similar targeting program, Colombia's SISBEN,[61] in which municipal governments hire enumerators and contractors to administer questionnaires to constituents to identify beneficiaries that are eligible for social assistance.

Several previous papers within the empirical political economy literature have specifically studied how politicians' preferences and electoral incentives influence public hiring decisions. A body of research, including Alesina et al. (2000) and Enikolopov (2010), use aggregate-level employment data to investigate whether politicians engage in "redistributive" public employment and patronage, i.e. hiring more public workers than is socially optimal.[62] Other papers, such as Iyer and Mani (forthcoming) and

[60] Niehaus et al. (2011), for instance, examine the problem of a centralized government agency designing a questionnaire that assists enumerators in identifying the poor. If such a questionnaire is implemented by a corruptible agent, the authors show that increasing the number of eligibility criteria may strictly worsen targeting performance, as monitoring becomes increasingly difficult.

[61] Castañeda (2005) provides a detailed summary of the SISBEN implementation procedures.

[62] Both authors utilize public employment data from U.S. municipalities. Alesina et al. (2000) provide evidence that municipal employment is increasing in income inequality and ethnic fractionalization. The authors' theory predicts such an outcome, as politicians strategically "disguise" redistribution via public employment. Enikolopov (2010) provides evidence that municipalities with a directly elected (as opposed to appointed) executive hire more full-time public employees (in particular during election years), which the author argues is consistent with patronage hiring and driven

the aforementioned Olken (2007), utilize individual-level data to investigate public employment decisions. Olken provides evidence that family members of local leaders are more likely to be paid to work on a road building project when the probability of external audit is higher.[63] Iyer and Mani, on the other hand, utilize data on Indian bureaucrats to examine the interaction between career-motivated bureaucrats and politicians that can assign policy tasks to these civil servants, but do not have the ability to hire or fire.

Building upon the studies, this paper utilizes detailed, candidate-level information on human capital, cognitive ability, demographic information and relationships with the local Chief to investigate whether Chiefs select the most qualified candidates, or whether other considerations influence selection. Information regarding selected bureaucrats' technical ability to effectively implement program objectives, and the overall policy outcome of village-level committees, is also utilized to estimate how suboptimal selection impacted overall program performance.

The targeting program established procedures for selecting impoverished households that qualify for a poverty identification card, which, as in India's BPL program, entitles cardholders to social assistance benefits. These procedures are implemented within each village and call for an important role of the local officials, such as the village Chief, with the goal of taking advantage of local knowledge of households' livelihood conditions. The program was designed so that the number of identified beneficiaries would approximately equal the overall poverty rates computed by the central government, in an effort to respect capacity constraints inherent to social

by the incentives of the executive.

[63] Olken suggests that this pattern may reflect local leaders trading off one form of corruption (extracting money from the program budget) for another form that is more difficult to detect (nepotism) as the probability of audit increases.

assistance programs.

Within each village, a committee of around seven villagers (referred to as the "bureaucrats" or "appointees") was appointed by the Chief to implement the identification procedures.[64] The Chief was given considerable freedom to select this committee, within some general program guidelines (e.g., 25 percent of committee seats were supposed to be reserved for women). Appointees received several benefits, including a monetary payment, for participating as a bureaucrat, and as I discuss below, evidence suggests that high capacity villagers were willing to participate as a bureaucrat. Therefore, the Chief had a large pool of candidates to select from.

Each appointee interviewed a subset of households in the village using a short questionnaire. This questionnaire, a Proxy Means Test (PMT), was designed to gather information about household livelihood and contains an objective decision rule regarding whether or not a household should be selected as a beneficiary. Appointees, however, were given some discretion to adjust the list of beneficiaries at the margin, in order to utilize local knowledge of household circumstances. As it may be expected, these procedures lead to the selection of households that reflect more villagers' perceptions of who are poor, rather than poverty status based upon a consumption indicator, and may lead toward higher satisfaction among beneficiaries (Alatas et al., 2010). In fact, relative to other international targeting programs, overall identification accuracy of the program under consideration was relatively good.

In order to effectively execute the targeting task, appointees needed the technical ability to properly administer the PMT (e.g., command of basic algebra and the

[64] In theory, the Chief was supposed to select a list of candidate bureaucrats, and then villagers were supposed to vote for their preferred candidates. As I describe below, however, the public vote did not occur in many villages, and when it did, the number of candidates typically exceeded the number of committee members by only one or two.

ability to think logically); moreover, appointees with additional knowledge of special circumstances influencing households' livelihoods would also have the advantage of being able to incorporate this information into their selection process. It's an empirical question, however, of whether Chiefs appointed individuals based on qualifications such as these, or whether other considerations influenced appointments.

The paper examines the Chiefs' selection criteria by utilizing the aforementioned candidate-level data. Estimation results indicate that, in general, Chiefs tended to appoint individuals with relatively high degrees of educational attainment and intelligence, as well as individuals that participate in social and civic organizations. At the same time, however, Chiefs were also significantly more likely to appoint their associates, as well as individuals that exhibit similar characteristics as the Chief (e.g., individuals around the same age as the Chief), leading to a reduction in targeting accuracy.

The paper then considers whether nepotism, cronyism (i.e., hiring friends and associates) and imperfect information about candidates' skills influences selection.[65] Section 3.3 develops a simple theoretical model that incorporates a private rent seeking motive for the Chief (i.e., nepotism or cronyism), as well as an information asymmetry regarding candidate quality, in the spirit of the "statistical discrimination" literature of Phelps (1972) and, in particular, Aigner and Cain (1977). The model assumes that *ex ante*, the Chief has relatively accurate information about the skill quality of candidates within his or her network (i.e., relatives and social associates), and noisier information about individuals outside of his or her network. Theoretical predictions

[65] Given the small number of appointees in each village, it seems relatively less likely that patronage was an overriding motivator for the Chief, given that so few patrons could be influenced. Chapter 2, however, provided evidence that this targeting program was utilized to strategically allocate benefits to non-poor households in electorally competitive districts.

are derived from the model, and data are shown to be consistent with the cronyism mechanism.

Armed with these estimates, the paper then turns to investigate characteristics that are associated with strong performance in executing the program's procedures. First, the paper examines candidate qualities that are associated with competent implementation of the targeting program's PMT; not surprisingly, intelligent villagers with more educational attainment perform better. The paper also investigates the characteristics of bureaucrat committees associated with including (excluding) poor (non-poor) households from the list of beneficiaries. Committees comprised of socially active bureaucrats perform significantly better at selecting households that are perceived as poor by villagers (and not selecting the non-poor), and more educated committees perform better as well. Interestingly, more intelligent committees allocated significantly more benefits to non-poor households, and committees with a greater share of members within the Chief's network exhibited lower targeting accuracy. Overall, parameter estimates suggest that suboptimal hiring reduces the number of poor households included in the program by 13 percent, and doubles the number of non-poor households included in the program.

In addition to the aforementioned papers, this research is also related to several other literatures. First, the paper is related to a growing literature that studies local capture in rural villages, including papers such as Reinikka and Svensson (2004), Galasso and Ravallion (2005), Olken (2006) and Araujo et al. (2008) (among others). Suboptimal selection of public workers that is motivated by nepotism or cronyism can indeed be viewed as a form of capture. In addition, the paper is tangentially related to a literature that examines bureaucracy, including Alesina and Tabellini (2007),

Alesina and Tabellini (2008) and Maskin and Tirole (2004).[66] These papers study the implications if policy decisions are made by bureaucrats, as opposed to elected politicians. The key mechanism in these papers is the different incentives that the two types of officials face. In addition, Bardhan and Mookherjee (2005) examine the relative trade-off of allocating power to a potentially corruptible central bureaucrat, versus decentralizing a policy and opening the possibility of local capture.

The paper is organized as follows: Section 3.2 describes the data and public program. Section 3.3 contains the analysis for bureaucrat selection. In particular, this section includes the theory that investigates why Chiefs may tend to hire associates, the econometric strategy that's employed to study this topic and the estimation results. Section 3.4 then investigates bureaucrat performance, and also estimates the effects of suboptimal appointments. Finally, Section 3.5 concludes.

3.2 Program context and data

This section describes the program under study (Subsection 3.2.1) and the data (Subsection 3.2.2) in more detail. Subsection 3.2.3 discusses bureaucrat compensation and villagers' willingness to work as a public servant.

3.2.1 The targeting program

The primary objective of the targeting program is to allocate poverty identification cards to impoverished households (and, to *not allocate* cards to non-poor households). Households that are deemed eligible for an identification card ("beneficiaries") have

[66] See also Besley and Coate (2003), which compares elected and appointed regulators. The authors study how incentives differ between these two classes of bureaucrats, and how these incentives influence policy outcomes.

access to free health care at participating medical facilities, which amounts to free health insurance. Recently, other forms of assistance are increasingly being allocated to cardholders.

As mentioned above, implementation of the program procedures is decentralized and primarily headed by a committee of around seven villagers.[67] Guidelines indicate that committee members should be "intelligent" and be representative of ethnic minorities in the village; moreover, the program mandated that at least 25 percent of a village's bureaucrats should be female. In addition, the program suggested that appointees be chosen from current village organizations, such as community associations and religious leaders.

In theory, the committee is selected in a two-stage process: first, the Chief proposes a list of candidates, and second, villagers vote for their preferred bureaucrats during a village meeting.[68] In practice, however, voting occurs in only around 60 percent of villages, and when voting did occur, the total number of candidates exceeded the total number of implementers by only one or two in the average village. This suggests that the Chief wields considerable influence over the selection of the public servants. In fact, analysis in Subsection 3.3.4 tests whether the criteria by which bureaucrats were selected was different in villages where voting occurred. This hypothesis is rejected in the data: estimates indicate that the selection criterion is not different when villagers had the opportunity to vote, providing support that the Chief exhibits significant influence over the appointment process.

In each village, bureaucrats determine which households are eligible for assistance in three steps. In a first step, the bureaucrats interview households in the community

[67] The actual size of the committee, however, varies according to village size.

[68] In the village meeting, villagers are supposed to have the option to propose additional candidates. Additional candidates were nominated in fewer than 15 percent of villages, according to the Chief.

by administering a short questionnaire (a Proxy Means Test, or PMT) that was developed centrally by the targeting program. As with similar programs around the world, the PMT questions are intended to gather information such as the quality of the household's dwelling and asset ownership. For each question, the household is assigned a point value based on the household's response, and households with a total point score below a certain range are deemed as an eligible beneficiary.

In a second step, bureaucrats can use their local knowledge of a household's circumstances (e.g., the household recently suffered a crop failure) to adjust the livelihood category indicated by the PMT.[69] Finally, in a third step, the bureaucrats are supposed to display the draft list of beneficiaries in public, and the bureaucrats and Chief are supposed to organize a village-level meeting to receive feedback regarding the list of beneficiaries. According to villagers, however, this third step in the process is often not implemented: only 21 percent of households reported that a draft list of beneficiaries was publicly displayed (36 percent did not know), and only around 29 percent of households reported that a village-level meeting occurred to present and discuss the list of beneficiaries (37 percent did not know).[70]

The decentralized nature of the program, however, also raises the possibility that the Chief and bureaucrats may manipulate the beneficiary selection process, to some degree, to allocate benefits to non-poor friends and relatives. Such considerations may influence Chiefs' decisions regarding who to appoint as a bureaucrat.

[69] In fact, the program also provided suggestions regarding additional criteria (not included in the PMT) that could merit eligibility.

[70] The Introduction briefly mentioned that these targeting procedures may lead to the selection of households that *villagers perceive as poor*, as opposed to households that are *technically below a (consumption- or income-based) poverty line*. In fact, the targeting program's objective was to create a PMT that proxies for villagers' poverty perceptions, instead of benchmarking the PMT to a consumption-based poverty metric. This is noteworthy, because villagers' perceptions of who is poor can differ significantly from households that fall below an objective poverty line, as discussed in Alatas et al. (2010).

3.2.2 Data description

Analyses in this paper draw from a survey that was fielded within several months of the implementation of the targeting procedures, but before the program's identification cards were allocated to the households. Data were collected in 299 randomly selected villages in four states within the country, where the probability of village selection was proportional to the number of households residing in the village.

Within each village, data were collected from three distinct populations: households, bureaucrats and the Chief. Ten households were randomly selected for interview in each village.[71] Household selection was stratified such that five of the sample households had been selected as beneficiaries, and five of the sample households had not been selected as beneficiaries. Stratification was conducted using the official list of beneficiaries that was created by the bureaucrat committee. In total, 2,943 household were interviewed, of which 1,467 (1,476) were (not) beneficiaries.[72] The Chief and bureaucrats were also administered specially designed survey instruments.[73]

[71] In villages with 250 or more households, the village was partitioned into smaller segments. One village segment was then selected to be the area of interview, again with the probability proportional to the number of households residing in the segment. In one particularly large village, two sets of ten households were selected for interview.

[72] In each village, four replacement households (two beneficiaries, and two non-beneficiaries) were selected, in the event that a household was unavailable for interview. In 74 percent of these cases where a household member was not available, the reason was because household members were at work or away from the village. Sampling weights are adjusted to reflect the actual number of beneficiaries and non-beneficiaries in each village, and are also adjusted for non-response (i.e., weights reflect the population in the four states where data were collected).

[73] In one (three) [one] village, data were not collected for the bureaucrats (Chief) [both bureaucrats and Chief] because respondents were unavailable; these five villages are excluded from the analysis. Overall, data were collected from over 80 percent of bureaucrats that operated in the remaining 294 villages.

Individual-level data

Individual socio-economic data, such as age and years of completed schooling, were collected for the bureaucrats, Chiefs and the villagers that reside in the ten randomly selected households that were interviewed in each village. Bureaucrats and Chief information was collected directly from each subject, while information for the villagers was recorded on a household roster that was completed during the household interview.[74] The information for the bureaucrats, Chiefs and adult members of the randomly selected households are pooled and utilized in the subsequent analysis.[75]

In addition to the socio-economic questions, the questionnaires also included a nonverbal intelligence test: an abridged version of the Raven's Standard Progressive Matrices. The Raven test presents respondents with an abstract picture that exhibits an overall design pattern, but a portion of the picture is blank. The respondent is asked to select one of several pictorial options to fill in the blank space, in order to complete the overall pattern of the picture. This test has been widely administered in developed and developing countries to measure intelligence; according to Brouwers et al. (2009), "the Raven tests are still considered to be measures of intelligence that show less influence of confounding cultural factors on the cross-national differences than any other intelligence test" (p. 330). This test was administered to the Chiefs and bureaucrats, as well as the adult household member with the highest educational

[74]Roster information was collected for all household members, including household members that no longer reside in the village (e.g., individuals that are considered household members but permanently reside in a city or work abroad). Household members that were absent from their household for six or more of the previous 12 months are dropped (1,392 cases); all results are qualitatively unaffected if these observations are retained. In addition, 549 sample villagers for which no schooling information was reported are dropped as well, as the schooling variable is incorporated into all analyses below.

[75]In some villages, the Chief selected him or herself as a bureaucrat. Moreover, in some villages, the Chief's or a bureaucrat's household was among the ten randomly selected households that were interviewed. These duplicate observations were accounted for prior to analyzing the data.

attainment that was available on the day the sample household was interviewed.

Data were also collected to identify which bureaucrats and villagers are related to the Chief. In order to mitigate the possibility of misreporting, each respondent household in the sample was asked to identify whether the Chief was related to each of the bureaucrats, or each of the other sample households that were interviewed in the village. For each bureaucrat and individual, the answers provided by the respondent households in the village were aggregated to identify whether the individual is related to the Chief.[76]

Finally, the bureaucrat questionnaire also included a ten-question PMT exam that tested each bureaucrat's ability to correctly administer the targeting program's PMT. Questions on this exam were modeled off of indicators on the PMT and were meant to simulate possible scenarios that the bureaucrat could have faced while administering the PMT. For instance, one question read, "The floor of a house is 7 meters across the front and 6 meters deep. How many square meters is the floor space of the house?" Results from this exam are utilized to assess the bureaucrats' technical capacity to implement the PMT.

Sample means for the aforementioned variables are displayed in Table 3.1. Column 1 (Column 2) of the table reports means for the sample bureaucrats (Chiefs), whereas Column 3 reports means for the entire pooled sample of bureaucrats, Chiefs and adult villagers that reside in the randomly selected households. On average, bureaucrats have about 1.5 more years of completed schooling relative to Chiefs, and

[76]Respondents were informed that familial relations included siblings, parents and children; grandparents and grandchildren; and aunts/uncles, nieces/nephews and cousins. An individual was defined as related to Chief if either (i) a member of his or her household is the Chief; (ii) the individual's household identified itself as a relative of the Chief; or (iii) two or more other respondents in the village identified the individual's household as related to the Chief. This definition of familial relationships is similar to the "elite connectedness" variable in Alatas et al. (2010).

Table 3.1: *Sample means of individual-level variables*

	Description	Bureau.	Chief	Total
A. Individual characteristics				
$school_{iv}$	Schooling (years)	6.79	5.29	4.46
$association_{iv}$	Number of associations	0.49	0.69	0.20
age_{iv}	Age	42.81	56.82	36.54
$business_{iv}$	Occupation in business†	0.06	0.04	0.15
$male_{iv}$	Male†	0.65	0.96	0.44
$minority_{iv}$	Ethnic minority†	0.03	0.05	0.04
$intel_{iv}$	% correct: Intelligence test	0.60	0.52	0.46
B. Connectedness				
R_{iv}	Related to (or is) Chief†	0.24	0.93	0.18
A_{iv}	Associate of Chief†	0.22	0.50	0.07
C. PMT exam				
PMT_exam_{iv}	% correct: PMT exam	0.56	.	.
Sample size		1,559	294	10,468

Notes. Column 1 (Column 2) [Column 3] reports sample means for the bureaucrats (Chiefs) [total sample, including the random sample of adult villagers]. Variables marked with the † symbol are indicator variables. The PMT exam was completed by 1,543 bureaucrats. The intelligence test was completed by 4,438 individuals in total. All means are sample weighted.

about 2.3 more years of schooling relative to the general village population. Relative to the Chiefs (villagers), bureaucrats answered 8 (14) percentage points more questions correctly on the intelligence test.

To proxy for an individual's degree of social involvement, the variable $association_{iv}$ is created, which sums the number of local associations that members of individual i's household belong to. Greater social involvement and integration in the community may increase a bureaucrat's knowledge of the livelihood status of other villagers (as well as other villagers' perceptions of who is poor), and thus may contribute to a more accurate identification of beneficiaries. Information was gathered for five types of local associations (artisan, farmer's, religious, savings and women's associations). On average, bureaucrats belong to about 0.3 more associations than the general population, but about 0.2 fewer associations than Chiefs.

Interestingly, Chiefs tend to appoint a relatively high proportion of associates that are in the same social associations, whereas a similar pattern does not exist for the Chiefs' relatives. The variable R_{iv} equals 1 if individual i is related to the Chief of village v (and 0 otherwise), whereas the variable A_{iv} equals 1 if a member of individual i's household belongs to the same organization as a member of the Chief's household.[77] Only 7 percent of adult villagers belong to the same association as the Chief, whereas Chiefs' social associates constitute 22 percent of all bureaucrats. On the other hand, 18 percent of villagers are related to the Chief, while 24 percent of bureaucrats are related to the Chief.

One point worth mentioning is that the actual Chief was interviewed in only 93 percent of villages. In the remaining villages, the Chief was unavailable on the interview day, in which case the Deputy Chief (or another official) was interviewed. In

[77] In particular, if the Chief is not a member of an association, then $A_{iv} = 0$.

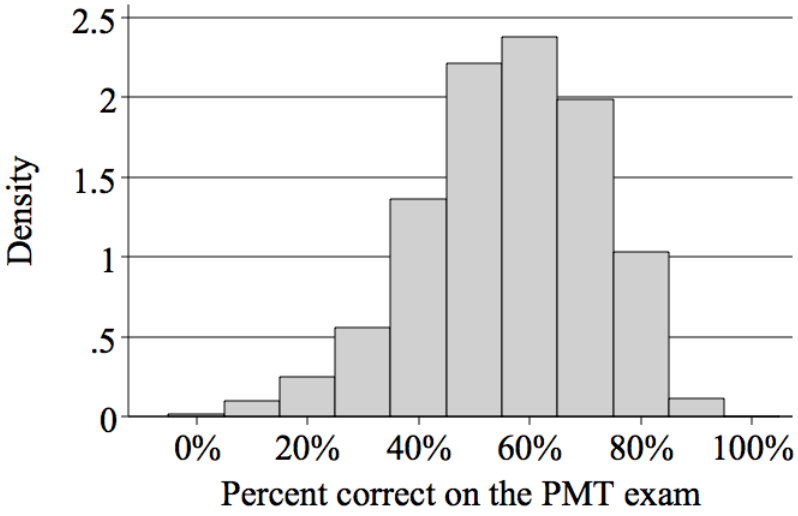


Figure 3.1: *Distribution of PMT_exam_{iv}*

these villages, data from the Deputy Chief are utilized in the following analyses, implicitly treating the Deputy as the official Chief. All qualitative results, however, are robust to dropping the 7 percent of villages in which the actual Chief was unavailable for interview.

Finally, Figure 3.1 displays the distribution of the percentage of questions that the bureaucrats answered correctly on the PMT exam. On average, the bureaucrats answered 56 percent of questions correctly, but the distribution is disperse, suggesting that there is a variability in bureaucrats' technical capacity.

Aggregated variables

Below, aggregated data are utilized to perform one set of analyses. The sample means of these control variables are displayed in Table 3.2; variables listed in Panel A are

constructed at the village-level and were gathered from either the country's national census, or from the interview with the village Chief. Variables listed in Panel B are constructed at the Council jurisdiction-level, where each jurisdiction contains on average about 9 villages. The most decentralized elected government body in the country is this local Council; a description of the Council.

Table 3.2: *Sample means of village- and Council jurisdiction-level variables*

A. Village characteristics	
# households in village	197.90
# households in village / # households in jurisdiction	0.12
% of households: Own $\leq$ 0.5 ha of land	0.41
% of households: Own $\geq$ 2.0 ha of land	0.21
Distance to state administrative center (km)	31.48
B. Council jurisdiction characteristics	
Incumbent party's margin of victory	0.41
% of households: Own 1+ mobile phone	0.27
% of households: Own 1+ tv	0.54
% of adults: Completed primary school	0.32
% of adults: Literate in the local language	0.65
Sample size	294

Note. All means are sample weighted.

Targeting accuracy

Finally, household-level data are utilized to assess bureaucrat performance in the field. The relevant household data are an indicator for whether or not a household was identified as an eligible beneficiary by the bureaucrats, and two distinct measures for classifying households as poor or non-poor. The first measure is derived directly from the targeting program's formal PMT; this PMT was re-administered to each sample household by the professional enumerators that collected the data utilized in

this paper. This poverty measure is objective in nature and follows the targeting program's official method for classifying households as poor.

The targeting program's intended "poverty benchmark," however, was villagers' perceptions of whether or not a household is poor (see Footnote 70). In practice, some bureaucrats modified the livelihood category indicated by the PMT to account for these perceptions. Therefore, households are also classified as poor or non-poor using villagers' perceptions. In the spirit of Alatas et al. (2010), the household questionnaire included a module that gathered information about respondents' perceptions regarding the livelihood status of the other sample households in the village. Each respondent was shown a card with the name of the household head written on it and asked,

> Households in your village can be classified into one of four livelihood levels: very poor, poor, adequate or rich. In your opinion, is (*name of household head*)'s household very poor, poor, adequate or rich?

This module of the questionnaire was extensively piloted and strategically placed towards the beginning of the questionnaire, before any mention of the targeting program, to mitigate response bias. A sample household is identified as poor if the household was classified as poor by half or more of the other respondent households in the village that classified the sample household.[78]

As may be expected, there are differences between the list of eligible beneficiaries, the PMT and villagers' perceptions. Bureaucrats identified 31.1 percent of households as eligible for benefits; villagers, on the other hand, reported that 36.6 percent of

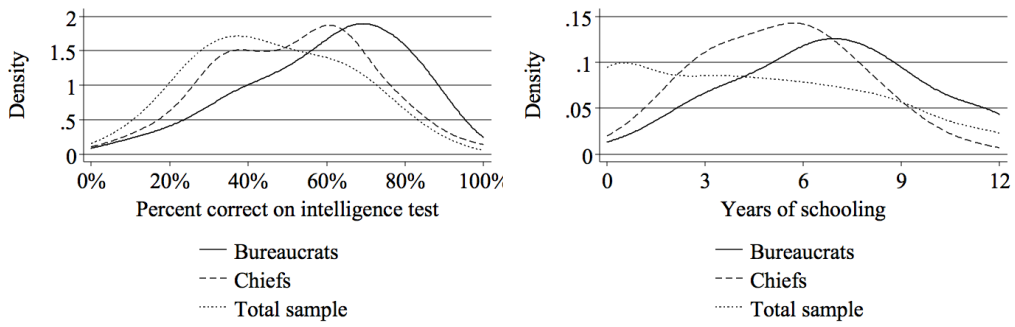
[78] If half of the villager respondents perceived a sample household as poor and the other half as non-poor, then the sample household was classified as poor. If fewer than three respondent households reported a livelihood category for a sample household, then the perceptions variable was coded as "missing."

households are poor, whereas the targeting program's PMT only deemed 17.9 percent of households eligible for benefits.

3.2.3 Willingness to be a bureaucrat

A key question is whether villagers (in particular, high skilled villagers) were willing to participate as bureaucrats. According to the staff of the targeting program, villagers have a sense of pride (i.e., psychological benefit, or prestige) of assisting their village by participating as a bureaucrat. There appears to be some truth to this: on a scale of 1 to 4, with 1 being "not important at all" and 4 being "very important," the average villager rated the importance of the targeting program as 3.7. Survey data suggest that many villagers in the top half of the skill distribution would have accepted the position *without compensation*. Respondents were read a generic description of the tasks that bureaucrats were expected to execute, and then asked if they would have been willing to participate as a bureaucrat (without any mention of monetary payment or private benefit). About 59 percent of villagers with 5 or more years of schooling (the median schooling attainment), and 69 percent of villagers with 10 or more years of schooling, answered that they would have participated.

Of course, a number of villagers may not be willing to work as a bureaucrat for free. Consequently, the central targeting program offered monetary compensation in exchange for working as a bureaucrat. To offset the opportunity cost of high-ability candidates in the village, the targeting program paid an average daily wage to bureaucrats that was around 50 percent greater than the median per capita daily consumption of rural households in the states where data were collected. Moreover, given that the specific timing of interviews and meetings was somewhat flexible, it may not have been necessary for most bureaucrats to disrupt their regular schedule



(a) $intel_{iv}$ **(b)** $school_{iv}$

Figure 3.2: *Distribution of* $intel_{iv}$ *and* $school_{iv}$*, by bureaucrat, Chief and the general population samples*

to a large degree.[79]

The summary statistics in Table 3.1 support the idea that, across villages, relatively higher capacity individuals were willing to participate as a bureaucrat. This intuition is also supported in Figure 3.2, which plots the distribution of $intel_{iv}$ and $school_{iv}$ among bureaucrats, Chiefs and the general village population. Bureaucrats' distributions of these variables second order stochastically dominate the other distributions.

3.3 Bureaucrat selection

3.3.1 Theoretical framework

To begin, a simple theoretical model is developed to articulate predictions about bureaucrat selection. As discussed above, if Chiefs tend to appoint relatives and

[79] Also, some villagers may also have been interested in enjoying non-pecuniary benefits associated with being a bureaucrat, such as selecting friends and relatives as beneficiaries.

associates, then this may be due to favoritism (i.e., nepotism or cronyism). On the other hand, however, such behavior may simply reflect the fact that the Chief has better information about such candidates. Both features are modeled below, and a theoretical prediction is derived in an attempt to distinguish between these two motives.

Consider a Chief that is tasked with selecting a bureaucrat from a set of candidates $\{1, ..., I\}$. In reality, the Chief actually selects several bureaucrats in the village, but the case of one bureaucrat is considered to easily articulate the mechanisms. Each candidate i is endowed with a characteristic vector $(T_i, x_i, \tilde{x}_i)$, where the elements of this vector are interpreted as follows:

- $T_i \in \{0, 1\}$ equals 1 if candidate i is within the Chief's social network (e.g., is a relative or associate), and 0 otherwise.
- $x_i \in \{0, 1\}$ denotes a characteristic that the Chief can observe for all candidates (e.g., ethnicity).
- $\tilde{x}_i \in \Re$ denotes a characteristic that the Chief may or may not be able to easily observe (e.g., intelligence or an individual's educational attainment).

In addition, $x^C \in \{0, 1\}$ is used to denote the Chief's value of the observable characteristic. For simplicity, the observable characteristic x_i is treated as a binary variable, but in practice this observable characteristic could take on multiple values or be multidimensional. For instance, x_i could denote the age (or a general age group) of candidate i, if age is common knowledge in the village.

If candidate i is selected as a bureaucrat, then i produces $f(\tilde{x}_i)$ output, where the production function $f(\cdot)$ is continuous, strictly increasing and weakly concave. There

characteristic $\tilde{x}_i$ represents a key input for producing the good, such as cognitive ability or a dimension of human capital. The Chief has preferences over candidates that are represented by the expression

$$\omega \cdot E\Big[U\big(f(\tilde{x}_i)\big) \mid \tilde{s}_i\Big] \;+\; (1-\omega) \cdot T_i \cdot \mathbf{1}_{x_i = x^C}$$

where U is a function mapping bureaucrat production into social welfare, $\tilde{s}_i$ is a signal (described below), $\mathbf{1}_{x_i = x^C}$ is an indicator function signifying that the Chief and i exhibit the same observable characteristic and $\omega \in [0, 1]$.

The function $E[U(f(\tilde{x}_i)) \mid \tilde{s}_i]$ is thus the expected welfare that i will produce for the public. The function U satisfies $U' > 0$ and $U'' < 0$: greater bureaucrat production (i.e., targeting accuracy) is preferred by the Chief, but there is a diminishing marginal benefit of this production. On the other hand, the expression $T_i \cdot \mathbf{1}_{x_i = x^C}$ is a private benefit that the Chief receives for appointing a network member ($T_i = 1$) that shares the same characteristic ($\mathbf{1}_{x_i = x^C} = 1$). The implicit assumption is that the Chief prefers to favor such individuals: the Chief's "cronies" are those in his or her social circle that are closest in terms of characteristics such as ethnicity, age, etc. Then ω is the relative weight that the Chief places on social welfare.

As alluded above, the Chief may have imperfect information regarding $\tilde{x}_i$. In particular, prior to selecting a bureaucrat, the Chief observes a signal

$$\tilde{s}_i \;=\; \begin{cases} \tilde{x}_i + \epsilon_i & \text{if } T_i = 0 \\ \tilde{x}_i & \text{if } T_i = 1 \end{cases}$$

for each candidate i, where ϵ_i is a zero mean, *ex ante* unobserved noise term. In other words, if candidate i is in the Chief's network ($T_i = 1$), then the Chief can observe $\tilde{x}_i$,

whereas if candidate i is outside of the Chief's network ($T_i = 0$), then the Chief can only observe a noisy signal of the candidate's ability. Of course, in reality, the Chief may not have perfect observability of, say, an associate's cognitive ability. However, the key assumption here is that the Chief has relatively better information about the characteristics of members of his or her network, relative to individuals outside of the network.

The distribution of the noise term ϵ_i is assumed to depend upon whether or not the Chief and candidate i are similar in terms of *observable characteristics*. Intuitively, if the Chief and candidate i are roughly the same age or of the same ethnic group, then presumably the Chief will have better information regarding i's unobserved characteristic, relative to a villager that is, say, 30 years younger than the Chief and outside of the Chief's social network. For instance, if the Chief and candidate i are of similar age, then they may have attended school together.

Formally, the assumption is that $\epsilon_i \sim G^{=}$ if $x_i = x^C$, and $\epsilon_i \sim G^{\neq}$ if $x_i \neq x^C$, where the G distributions are differentiable, have support $(-\infty, \infty)$ and are such that $G^{\neq}$ is a mean-preserving spread of $G^{=}$. In other words, if the Chief and candidate i exhibit the same observable characteristic, then the Chief's signal $\tilde{s}_i$ will be more precise, relative to another candidate that exhibits a different observable characteristic than the Chief. The respective probability density functions are denoted $g^{=}$ and $g^{\neq}$, which are symmetric around 0. Given this information structure, the following proposition can be established:

Proposition 4. *The Chief's optimal policy is such that:*

1. The Chief exhibits a bias for appointing candidates within his or her network:

   - $(T_i, T_j) = (1, 0)$ *and* $f(\tilde{x}_i) = f(\tilde{x}_j) \Rightarrow$ *the Chief is more likely to select* i.

2. *If the Chief maximizes welfare ($\omega = 1$), then* (a) when selecting between network members, the Chief appoints network members based only on merit*, and* (b) when selecting between non-network members, the Chief is biased towards appointing individuals that are like him or herself:

   - $(T_i, T_j) = (1, 1)$ *and* $f(\tilde{x}_i) \geq f(\tilde{x}_j) \Rightarrow$ *the Chief will select* i.
   - $(T_i, T_j) = (0, 0)$, $f(\tilde{x}_i) = f(\tilde{x}_j)$ *and* $x_i = x^C \neq x_j \Rightarrow$ *the Chief is more likely to select* i.

3. *If the Chief maximizes private rents ($\omega = 0$), then the Chief* appoints network members that are like him or herself.

Proof. Each item is proven separately:

1. Consider the case where $\omega = 1$; cases in which $\omega < 1$ will then follow immediately. Let $\tilde{x}_i = \tilde{x}_j$ and $(T_i, T_j) = (1, 0)$. If the Chief receives the signal $s = \tilde{x}_j$, then the Chief's expected utility is such that

$$U\Big(f(\tilde{x}_i)\Big) \quad > \quad \int U\Big(f(s-\epsilon)\Big)g(\epsilon)d\epsilon, \tag{3.1}$$

   where the inequality follows from the strict concavity of U, f strictly increasing and weakly concave and Jensen's inequality. By continuity, define $\hat{s}$ such that[80]

$$U\Big(f(\tilde{x}_i)\Big) \quad = \quad \int U\Big(f(\hat{s}-\epsilon)\Big)g(\epsilon)d\epsilon.$$

   By U and f increasing, it follows that $\hat{s} - s = (\tilde{x}_j + \hat{\epsilon}) - \tilde{x}_j = \hat{\epsilon} > 0$. Therefore, the Chief selects i over j with probability $G(\hat{\epsilon}) > 0.5$.

[80] If such an $\hat{s}$ does not exist, then by continuity, (3.1) is satisfied for all s, which implies that the Chief will select i.

2. Let $\omega = 1$. Part 2 has two subparts:

   (a) Let $T_i = T_j = 1$. The Chief will prefer i over j iff $U(f(\tilde{x}_i)) \geq U(f(\tilde{x}_j))$.

   (b) Let $T_i = T_j = 0$, $\tilde{x}_i = \tilde{x}_j$ and $x_i = x^C \neq x_j$. Let ϵ' be a generic shock generating the signal $s' = \tilde{x}_i + \epsilon'$ received by the Chief. Since $G^{\neq}$ is a mean-preserving spread of $G^{=}$ and $U(f(z))$ is strictly concave, then the Chief's expected utility is such that

$$\int U\Big(f(\tilde{x}_i + \epsilon' - \epsilon)\Big) g^{=}(\epsilon) d\epsilon \quad > \quad \int U\Big(f(\tilde{x}_j + \epsilon' - \epsilon)\Big) g^{\neq}(\epsilon) d\epsilon.$$

   Similar to Item 1, define $\epsilon''(\epsilon')$ such that

$$\int U\Big(f(\tilde{x}_i + \epsilon' - \epsilon)\Big) g^{=}(\epsilon) d\epsilon \quad = \quad \int U\Big(f(\tilde{x}_j + \epsilon''(\epsilon') - \epsilon)\Big) g^{\neq}(\epsilon) d\epsilon,$$

   where again $\epsilon''(\epsilon') > \epsilon'$. Therefore, conditional on ϵ', the probability that i is preferred to j is $G^{\neq}(\epsilon''(\epsilon'))$. Therefore, the overall probability that i is selected is

$$\begin{aligned}
\int_{-\infty}^{\infty} G^{\neq}\Big(\epsilon''(\epsilon)\Big) g^{=}(\epsilon) d\epsilon \quad &= \quad \int_{0}^{\infty} \Big[G^{\neq}\Big(\epsilon''(-\epsilon)\Big) + G^{\neq}\Big(\epsilon''(\epsilon)\Big) \Big] g^{=}(\epsilon) d\epsilon \\
&> \quad \int_{0}^{\infty} \Big[G^{\neq}(-\epsilon) + G^{\neq}(\epsilon) \Big] g^{=}(\epsilon) d\epsilon \\
&= \quad \int_{0}^{\infty} g^{=}(\epsilon) d\epsilon \quad = \quad 0.5,
\end{aligned}$$

   where the first equality follows from $g^{=}(\epsilon) = g^{=}(-\epsilon)$ by $G^{=}(\epsilon)$ symmetric around 0, the inequality follows from $G^{\neq}(\epsilon)$ strictly increasing and $\epsilon''(\epsilon) > \epsilon$ $\forall \epsilon$, and the second equality follows from $G^{\neq}(-\epsilon) + G^{\neq}(\epsilon) = 1$ by symmetry.

3. The result is immediate.

□

Part 1 of Proposition 4 simply states that, given two candidates with identical skills, the Chief is more likely to have appointed the candidate that belongs to his or her network. The rationale is that, *ex ante*, the Chief directly observes the skill level of the candidate of his or her network, whereas the Chief observes a noisy signal about the skill level of the candidate that is not in his or her network. Given that the Chief is risk averse, he or she will tend to appoint a network member, unless the signal of a non-network member is sufficiently high. If the Chief enjoys hiring network members per se ($\omega < 1$), then this will only reinforce the result.

The final two parts of the proposition provide a test for determining whether imperfect information or favoritism are the primary cause of Chiefs' tendencies to hire individuals that are like them. Part 2 of Proposition 4 suggests that if the Chief maximizes welfare but faces the imperfect information structure outlined above, then the Chief will select members of his or her network based solely on skill. When choosing between two identical candidates from outside the network, however, the Chief is more likely to favor the candidate with a similar observable characteristic as the Chief (as the quality signal is relatively more precise). Conversely, Part 3 of Proposition 4 suggests that if the Chief tends to appoint individuals based on favoritism, then the Chief will exhibit a bias towards selecting individuals like him or herself from among the network members.

3.3.2 Econometric strategy

Next, the paper discusses the econometric strategy that is used to investigate the selection criteria that Chiefs utilize when appointing bureaucrats, and whether biases are present in their appointment process. In particular, this subsection explains how the theoretical predictions presented in Proposition 4 are tested, in order to understand whether the favoritism or information asymmetry mechanism is more dominant.

Prior to discussing the Chief's biases, however, the paper first outlines the basic econometric framework that is used to assess the relationship between a villager's observable characteristics and the probability of being selected as a bureaucrat. The empirical setup is similar in spirit to Besley et al. (2005). Let $X_{iv} = (x_{1iv}, ..., x_{Jiv})$ denote a vector of characteristics for individual i in village v; the variables listed in Panel A of Table 3.1 are included in this vector (along with $(age_{iv})^2$ and a constant). Given X_{iv}, individual i would produce $X_{iv}\tilde{\beta} + \tilde{\epsilon}_{iv}$ if selected as a bureaucrat, where $\tilde{\beta}$ is a parameter vector and $\tilde{\epsilon}_{iv}$ captures unobservable characteristics of i that would influence bureaucrat quality.

Within village v, the Chief is tasked with appointing several bureaucrats from the village. Naturally, the Chief's selection will depend upon the distribution of candidates within the village. If the Chief acts to maximize bureaucrat quality, then there will exist some village-specific threshold $\tilde{\delta}_v$ such that candidate i will be chosen if and only if $X_{iv}\tilde{\beta} + \tilde{\epsilon}_{iv} \geq \tilde{\delta}_v$. Given this framework, the probability that individual

i is selected in village v can be estimated using the linear probability model[81]

$$y_{iv} = X_{iv}\beta + \delta_v + \epsilon_{iv}, \tag{3.2}$$

where $y_{iv} \in \{0, 1\}$ equals 1 if individual i served as a bureaucrat in village v (and 0 otherwise), δ_v is a village fixed-effect and ϵ_{iv} is an idiosyncratic error term.[82] Note that the fixed-effect δ_v will capture village-specific characteristics that influence the threshold of whether or not individual i is selected in village v.

While this model relates X_{iv} to an individual being selected as a bureaucrat, it does not test whether Chiefs hire associates or exhibit other systematic biases. To test for such considerations, the specification[83]

$$y_{iv} = X_{iv}\beta + \theta_R R_{iv} + \theta_A A_{iv} + \sum_j \eta_j |x_{jiv} - x^C_{jv}| + \delta_v + \epsilon_{iv} \tag{3.3}$$

is estimated, where x^C_{jv} denotes characteristic j of the Chief in village v, and (R_{iv}, A_{iv}) are indicators that signify if i is related to or is an associate of the Chief, respectively.

[81] Parameter estimates from linear probability models are presented throughout the remainder of this section. This approach provides at least two advantages. First, the linear probability model can explicitly account for village fixed-effects, whereas this is not possible with a probit model (due to the incidental parameters problem). Second, one of the key relationships of interest is an interaction effect. Interaction effects are particularly difficult to interpret in a non-linear estimation setting (see, for instance Ai and Norton (2003)). Nevertheless, when estimated using the Probit model, the qualitative insights are the same (though in some cases, statistical significance is stronger or weaker).

[82] Cawley et al. (2001) provide evidence that in the U.S., schooling and measured cognitive ability are so correlated that one cannot separately identify the marginal effects of these characteristics on outcomes of interest, without imposing strong modeling assumptions. In particular, the authors show that it is inappropriate to model log wage as a linear function of schooling and cognitive ability. In the data utilized in this paper, however, the correlation between $school_{iv}$ and $intel_{iv}$ is 0.47. Moreover, there are some individuals with no schooling that scored in the 90th percentile of the $intel_{iv}$ score distribution, and individuals with 12 years of schooling that scored in the bottom 10th percentile of the $intel_{iv}$ score distribution.

[83] Note that in all regressions reported below, observations are weighted to account for the probability of being included in the sample.

Element j of the parameter vector β denotes the marginal effect of a change in characteristic j on an individual's probability of being appointed. A parameter estimate of $\beta_j > 0$ would indicate that, in general, Chiefs are more likely to appoint an individual with a larger amount of j.

The parameters θ_R and θ_A denote the additional probability that relatives and associates of the Chief are appointed as bureaucrats, after controlling for the characteristics X_{iv}. A parameter estimate of, say, $\theta_R > 0$ would imply a preference for appointing relatives: given two equally qualified candidates i and i' (i.e., $X_{iv} = X_{i'v}$), the Chief is significantly more likely to select i relative to i' if $R_{iv} = 1$ and $R_{i'v} = 0$.

In addition, the specification also tests to see if Chiefs exhibit a pattern of appointing individuals that have similar characteristics as him or herself. The variables $|x_{jiv} - x^C_{jv}|$ denote the absolute difference between characteristic j of candidate i and characteristic j of the Chief.[84] A parameter value $\eta_j < 0$ indicates that Chiefs are less likely to hire an individual that differs from the Chief according to characteristic j (i.e., appointments exhibit assortative matching). For instance, consider two 40-year-old individuals that are identical; one individual resides in village v, where the Chief is 40 years old, whereas the other individual resides in village v', where the Chief is 65 years old. The individual in village v' is therefore $25 \cdot \eta_{age}$ more likely to be appointed as a bureaucrat.

Coefficient estimates where $(\theta_R, \theta_A) > 0$ and $\eta < 0$ would imply that the Chief exhibits some bias towards hiring people that are like him or herself. As discussed in Section 3.3.1, however, this could be because the Chief is exhibiting favoritism (e.g.,

[84]The absolute difference $|association_{iv} - association^C_v|$ is excluded from the set of absolute deviations because the variable A_{iv} includes much of this information. In addition, the absolute difference $|(age_{iv})^2 - (age^C_v)^2|$ is omitted as well, as the difference $|age_{iv} - age^C_v|$ will contain the pertinent information.

giving decent paying jobs to close friends, or hiring associates to wield more control over the program), or because the Chief simply has better information regarding candidates that exhibit similar characteristics as the Chief.

Proposition 4, however, suggests that if information asymmetries are the main driving force behind an assortative bias (i.e., $\eta < 0$), such a bias should exist primarily for candidates that are outside of the Chief's network (i.e., for individuals where $R_{iv} = A_{iv} = 0$). The Chief should have relatively better information about individuals within his or her network, regardless of characteristics X_{iv}, and thus the *magnitude* of η should smaller or zero across individuals within the network. If, however, cronyism or nepotism is the primary source of the bias, then it may be the case that the Chief exhibits a particular preference for hiring individuals that are in the network and similar according to, say, age and ethnicity. In such an instance, η may even be more negative across individuals within the Chief's network.

To investigate whether the information asymmetry or cronyism mechanisms appear to dominate, the variable $T_{iv} = \max\{R_{iv}, A_{iv}\}$ is constructed in order to estimate

$$\begin{aligned} y_{iv} &= X_{iv}\beta + \theta_R R_{iv} + \theta_A A_{iv} + \theta_T T_{iv} \\ &\quad + \sum_j \Big(\eta_j + \zeta_j T_{iv}\Big)|x_{jiv} - x_{jv}^C| + \delta_v + \epsilon_{iv}, \end{aligned} \tag{3.4}$$

where T_{iv} is an indicator signifying that individual i is within the Chief's relative or associate network. The information-based theory for assortative matching would predict that $\zeta_j \geq 0$, i.e. within the Chief's network, there is less of a tendency towards appointing individuals with similar characteristics. A value of $\zeta_j < 0$, however, may be relatively more plausible if nepotism or cronyism is the dominant mechanism.

3.3.3 Estimates

First, estimation results are presented for models (3.2) and (3.3). Recall, however, that $intel_{iv}$ was only collected for one member of each randomly selected sample household. Consequently, information on intelligence is only available for a selected sample of individuals: the bureaucrats, the Chief and this restricted sample of villagers. Estimates are therefore reported for two versions of the bureaucrat selection model: Table 3.3 presents estimates when $intel_{iv}$ is omitted from the specification, whereas Table 3.4 includes $intel_{iv}$ as a covariate.

Column 1 of these tables reports parameter estimates for model (3.2), in which only X_{iv} is included as controls; Column 2 incorporates the relationship variables (R_{iv}, A_{iv}), whereas Column 3 reports the full estimation results of model (3.3). Note that coefficients are relatively small in magnitude, because the committee consisted of relatively few bureaucrats: while around seven individuals were appointed to the committee, samples villages each have several hundred (or several thousand) adult residents.

Across both tables, a similar pattern emerges. In general, individuals with greater human capital and cognitive ability, as measured by years of schooling ($school_{iv}$), intelligence ($intel_{iv}$) and social activity ($association_{iv}$), are statistically more likely to be appointed by the Chief.[85] Below, it is shown that individuals with greater intelligence and schooling perform better at the technical aspects of the targeting program. In addition, socially active individuals perform better at identifying households that villagers perceive are poor.

In both tables, the indicator variable for being related to the Chief (R_{iv}) is sta-

[85]Statistical significance holds for all three of these variables across all specifications, except $association_{iv}$ in Column 2 of Table 3.3.

Table 3.3: *Chief's selection criteria, omitting intelligence*

	[1]	[2]	[3]
$school_{iv}$	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)
$association_{iv}$	0.011***	0.004	0.004*
	(0.002)	(0.002)	(0.002)
age_{iv}	0.002***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)
$(age_{iv})^2$	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
$business_{iv}$	-0.007***	-0.007***	-0.005***
	(0.001)	(0.001)	(0.001)
$male_{iv}$	0.004***	0.004***	0.005*
	(0.001)	(0.001)	(0.002)
$minority_{iv}$	0.009**	0.008**	0.005
	(0.004)	(0.004)	(0.004)
R_{iv}		-0.000	-0.000
		(0.001)	(0.001)
A_{iv}		0.023***	0.023***
		(0.005)	(0.005)
$\lvert school_{iv} - school_v^c \rvert$			-0.001***
			(0.000)
$\lvert age_{iv} - age_v^c \rvert$			-0.000***
			(0.000)
$\lvert business_{iv} - business_v^c \rvert$			-0.002***
			(0.001)
$\lvert male_{iv} - male_v^c \rvert$			0.001
			(0.002)
$\lvert minority_{iv} - minority_v^c \rvert$			0.009**
			(0.004)
Constant	-0.047***	-0.047***	-0.029***
	(0.003)	(0.003)	(0.007)
Village FEs	Yes	Yes	Yes
R^2	0.018	0.020	0.021
N	10,468	10,468	10,468

Notes. Each column reports parameter estimates from the linear probability model that regresses an indicator for being selected as a bureaucrat on the variables listed in the rows. Each model includes village fixed-effects, and observations are sample weighted. Standard errors are clustered by village and reported in parenthesis. The star levels ***, ** and * correspond to statistical significance at 1%, 5% and 10%, respectively.

Table 3.4: *Chief's selection criteria, including intelligence*

	[1]	[2]	[3]
$school_{iv}$	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
$association_{iv}$	0.035***	0.016**	0.016**
	(0.007)	(0.008)	(0.008)
age_{iv}	0.005***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)
$(age_{iv})^2$	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
$business_{iv}$	-0.022***	-0.023***	-0.018***
	(0.003)	(0.003)	(0.006)
$male_{iv}$	0.018***	0.016***	0.018*
	(0.003)	(0.004)	(0.010)
$minority_{iv}$	0.021	0.022	0.012
	(0.014)	(0.014)	(0.014)
$intel_{iv}$	0.082***	0.083***	0.083***
	(0.010)	(0.010)	(0.011)
R_{iv}		-0.001	-0.002
		(0.004)	(0.005)
A_{iv}		0.063***	0.062***
		(0.016)	(0.016)
$\lvert school_{iv} - school_v^c \rvert$			-0.002**
			(0.001)
$\lvert age_{iv} - age_v^c \rvert$			-0.001
			(0.000)
$\lvert business_{iv} - business_v^c \rvert$			-0.007
			(0.006)
$\lvert male_{iv} - male_v^c \rvert$			0.002
			(0.009)
$\lvert minority_{iv} - minority_v^c \rvert$			0.032**
			(0.014)
$\lvert intel_{iv} - intel_v^c \rvert$			-0.003
			(0.012)
Constant	-0.168***	-0.168***	-0.116***
	(0.016)	(0.016)	(0.037)
Village FEs	Yes	Yes	Yes
R^2	0.058	0.062	0.063
N	4,438	4,438	4,438

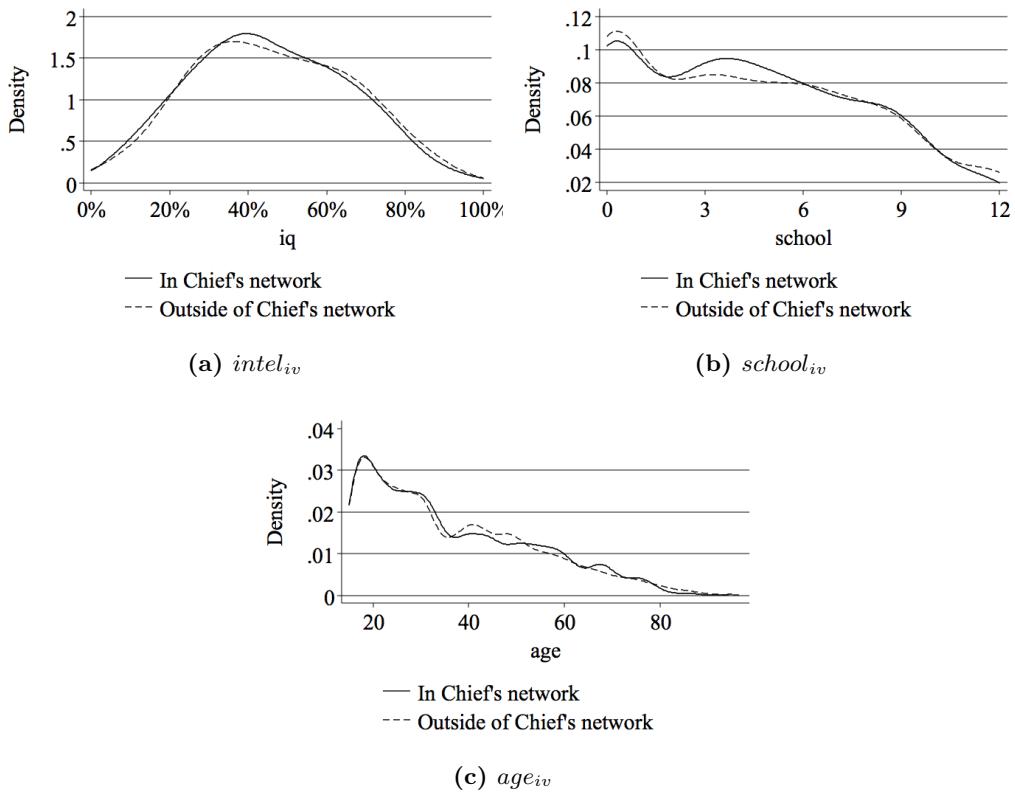
Note. See the notes for Table 3.3.

tistically equal to zero, whereas A_{iv} is positive and statistically significant; estimates suggest that associates of the Chief are between 2 and 6 percent more likely to be selected as a bureaucrat, relative to non-associates.

Finally, parameter estimates in the third columns of the tables provide some evidence that Chiefs do, indeed, exhibit a preference for appointing bureaucrats that exhibit similar characteristics as themselves. In Table 3.3, the η_j parameters are negative and significant for schooling, age and the indicator for whether or not individual i's main occupation is business-related. Interestingly, η_{male} is not statistically different from zero, whereas $\eta_{minority}$ is strictly positive at the 5 percent level. This may reflect the fact that the targeting program stipulated that bureaucrat selection be representative of minorities in the village, and that a certain fraction of females should be selected as committee members. In other words, Chiefs (who are predominantly male and non-minority) consciously select individuals that are different from himself along these dimensions. In Table 3.4, the only absolute difference variable that is statistically negative is for schooling, though the coefficients for age and the business indicator are both negative in sign. In particular, note that η_{iq} is not statistically different from zero, which is in accordance with the idea that Chiefs only receive a noisy signal about villagers' intelligence.

To summarize, the estimates in Tables 3.3 and 3.4 suggest that while Chiefs tend to select individuals with higher degrees of human capital, relative to individuals with lower degrees of human capital. While nepotism does not appear to be a systematic occurrence, Chiefs tend to appoint their associates, as evidenced by the positive θ_A coefficient. There is some evidence that Chiefs tend to select individuals with similar characteristics.

Next, the paper utilizes the prediction from Proposition 4 to test if the Chiefs' ten-



(a) $intel_{iv}$ **(b)** $school_{iv}$

(c) age_{iv}

Figure 3.3: *Distribution of individual characteristics, by* T_{iv}

dency to appoint individuals "like themselves" is predominantly driven by imperfect information. Interestingly, the distribution of intelligence, schooling and age for individuals within ($T_{iv} = 1$) and outside ($T_{iv} = 0$) of the Chief's network are quite similar. Figure 3.3 compares the respective distributions for the variables $intel_{iv}$, $school_{iv}$ and age_{iv}. The solid (dashed) lines correspond to the distribution of villagers for which $T_{iv} = 1$ ($T_{iv} = 0$).

Table 3.5 reports parameter estimates for (θ_T, η, ζ), as defined in specification (3.4). Proposition 4 indicated that if an information mechanism is primarily driving

Table 3.5: *Bias against households that are in and out of the leaders' network*

	[1]	[2]
T_{iv}	0.031***	0.073**
	(0.010)	(0.034)
$\lvert school_{iv} - school_v^c \rvert$	-0.000**	-0.001
	(0.000)	(0.001)
$\lvert school_{iv} - school_v^c \rvert \cdot T_{iv}$	-0.002***	-0.005***
	(0.001)	(0.002)
$\lvert age_{iv} - age_v^c \rvert$	-0.000	-0.000
	(0.000)	(0.000)
$\lvert age_{iv} - age_v^c \rvert \cdot T_{iv}$	-0.001***	-0.001***
	(0.000)	(0.000)
$\lvert business_{iv} - business_v^c \rvert$	-0.002	-0.005
	(0.001)	(0.007)
$\lvert business_{iv} - business_v^c \rvert \cdot T_{iv}$	-0.006**	-0.007
	(0.002)	(0.010)
$\lvert male_{iv} - male_v^c \rvert$	0.003	0.007
	(0.003)	(0.010)
$\lvert male_{iv} - male_v^c \rvert \cdot T_{iv}$	-0.009***	-0.020*
	(0.002)	(0.011)
$\lvert minority_{iv} - minority_v^c \rvert$	0.012***	0.040***
	(0.004)	(0.014)
$\lvert minority_{iv} - minority_v^c \rvert \cdot T_{iv}$	-0.014*	-0.039**
	(0.008)	(0.019)
$\lvert intel_{iv} - intel_v^c \rvert$		0.005
		(0.012)
$\lvert intel_{iv} - intel_v^c \rvert \cdot T_{iv}$		-0.022
		(0.026)
(X_{iv}, R_{iv}, A_{iv})	Yes	Yes
Village FEs	Yes	Yes
R^2	0.023	0.066
N	10,468	4,438

Notes. See the notes for Table 3.3. The models estimated in this table differ from Table 3.3 by incorporating the covariates listed in the rows. Parameters for with (X_{iv}, R_{iv}, A_{iv}) are unreported.

the observed pattern, then one should expect $\zeta \geq 0$, i.e. the Chief should exhibit *less bias towards selecting similar individuals* from within his or her network.

The parameters in Table 3.5, however, indicate that the opposite pattern tends to hold. In particular, the Chief exhibits greater tendency towards selecting individuals with similar education levels, a similar age, a similar sex and a similar ethnicity when appointing from his or her own network.[86] In fact, the only characteristic for which this does not occur at all is intelligence. This pattern is in stark contrast to the prediction provided in Part 2 of Proposition 4, but is not inconsistent with the idea of favoring close friends.

3.3.4 Additional analyses

To conclude this section, two additional dimensions of the program are investigated: first, the paper investigates whether Chiefs modify their selection of bureaucrats in response to an exogenous shock to the marginal benefit of the targeting program. During the year prior to the implementation of the targeting program, a drought affected a portion of the country. If Chiefs were partially keen on maximizing public welfare, then one would expect that the Chief would tend to select bureaucrats with higher human capital in response to such a shock, when households are particularly vulnerable. If cronyism is a dominant driver of the Chief's motivation, however, then selection may worsen if the relative power of common villagers declines in the response to such a shock.

To execute this test, the indicator variable $disaster_v$ is constructed. This variable equals 1 if the village suffered a drought in the previous year, and 0 otherwise.

[86] In addition, in Column 1, the negative $\zeta_{business}$ parameter suggests that Chiefs are more biased towards selecting individuals with the same type of occupation when selecting from their own network.

Table 3.6: *Selection of bureaucrats, by village characteristic*

	$I_v = disaster_v$		$I_v = vote_v$	
	[1]	[2]	[3]	[4]
$school_{iv}$	0.003***	0.007***	0.003***	0.005***
	(0.000)	(0.001)	(0.000)	(0.001)
$school_{iv} \cdot I_v$	-0.001***	-0.002**	0.000	0.001
	(0.000)	(0.001)	(0.000)	(0.001)
$association_{iv}$	0.001	0.012	0.004	0.019**
	(0.004)	(0.012)	(0.003)	(0.008)
$association_{iv} \cdot I_v$	0.004	0.006	-0.000	-0.005
	(0.005)	(0.015)	(0.005)	(0.016)
$intel_{iv}$		0.079***		0.085***
		(0.015)		(0.013)
$intel_{iv} \cdot I_v$		0.007		-0.006
		(0.020)		(0.020)
R_{iv}	-0.001	-0.003	0.002	0.005
	(0.002)	(0.008)	(0.002)	(0.005)
$R_{iv} \cdot I_v$	0.002	0.005	-0.004	-0.013
	(0.003)	(0.009)	(0.003)	(0.009)
A_{iv}	0.037***	0.103***	0.027***	0.066***
	(0.009)	(0.027)	(0.009)	(0.024)
$A_{iv} \cdot I_v$	-0.022**	-0.065**	-0.006	-0.004
	(0.011)	(0.032)	(0.011)	(0.033)
Village FEs	Yes	Yes	Yes	Yes
R^2	0.021	0.064	0.020	0.062
N	10,468	4,438	10,468	4,438

Notes. See the notes for Table 3.3. The model that generated the parameter estimates reported in this table differs from Table 3.3 by incorporating the I_v interaction terms listed in the rows. In Columns 1 and 2 (Columns 3 and 4), the indicator I_v equals 1 if the village suffered a natural disaster damaging crop production in the previous year (if half or more households reported that there was a public vote to select the bureaucrats), and 0 otherwise. The remaining variables included in X_{iv} are unreported.

Model (3.3) is then modified by interacting $disaster_v$ with the main measures of human capital and inteligence, $(school_{iv}, association_{iv}, intel_{iv})$, and the measures for connectedness to the Chief, (R_{iv}, A_{iv}).

Parameter estimates for this modified model are displayed in the first two columns of Table 3.6. These estimates provide a mixed picture regarding how Chiefs adjusts their appointments: in villages that experienced a shock, Chiefs tended to select individuals with lower levels of educational attainment, while simultaneously appointing fewer of their associates. These dual effects suggest that Chiefs may reduce favoritism for network members, but may not improve overall selection by much (if at all) in response to a negative shock.

Second, the paper investigates whether the final selection of appointees varied by whether or not households in the village reported that a public vote occurred to select the bureaucrats. The variable $vote_v$ is constructed, which equals 1 if a majority of household respondents reported that a public vote occurred (among those respondents that reported either yes or no to the question).[87] The variable $vote_v$ is then interacted with $(school_{iv}, association_{iv}, intel_{iv}, R_{iv}, A_{iv})$.

If elections are an effective mechanism for improving the selection of these public servants, then one should expect that the parameters multiplying the interaction terms should be positive: when given the chance to elect the bureaucrats, better quality bureaucrats are selected. To be sure, there may be omitted village-level variables that are correlated with $vote_v$ and the overall quality of bureaucrat selection in the village. In particular, if the Chief organized a public vote, then it is most likely the case that the Chief also follows good governance procedures, and thus selection of

[87] In particular, all households that reportedly "did not know" whether a vote occurred are omitted from consideration.

bureaucrats may be better in such villages. Consequently, if anything, there should be a positive bias for the parameter multiplying the interaction term between $vote_v$ and $(school_{iv}, association_{iv}, intel_{iv})$.

Columns 3-4 in Table 3.6 report the parameter estimates for this modified model. All of the parameters multiplying the interaction terms are insignificant, which provides evidence against the fact that a public vote influences the final bureaucrat selection. This provides additional evidence that the Chief maintains significant influence over appointing bureaucrats.

3.4 Bureaucrat performance

Given the evidence of Chiefs' biases presented above, one question is the extent to which bureaucrats with higher human capital actually perform better, and the extent to which suboptimal selection reduces program performance. This section proceeds in two steps. First, the paper examines the bureaucrat characteristics that are associated with strong performance in administering the targeting program's procedures. Here, both bureaucrats' technical capacity to implement the PMT, as well as the overall committee's ability to produce an accuracy list of beneficiaries, are examined. Using these findings, the paper estimates how overall targeting accuracy could have been improved, had the Chief appointed the "optimal" committee.

Successful implementation of the targeting program's PMT requires a minimum amount of human capital. From a technical point of view, the ability to think critically and perform basic algebraic tasks is necessary to properly implement the PMT. Consequently, a bureaucrat's innate cognitive ability and years of completed schooling should correspond to a higher technical ability. To assess the relationship between

the characteristics X_{iv} and a bureaucrat's technical skill, the model

$$PMT\_exam_{iv} = X_{iv}\beta + \theta_R R_{iv} + \theta_A A_{iv} + \epsilon_{iv} \quad (3.5)$$

is estimated (recall that PMT_exam_{iv} is the percent of questions that i answered correctly on the PMT exam). Of course, PMT_exam_{iv} is only observed for the bureaucrats that were actually appointed by the Chief. Consequently, estimating (3.5) via OLS may lead to sample selection bias. Therefore, (3.5) is estimated using both OLS estimator, and by employing the Heckman correction model to adjust for sample selection.[88] For the selection equation, model (3.4) is utilized, which incorporated the $\{|x_{jiv} - x_{jv}^C|\}_j$ terms measuring the differences between i's and the Chief's characteristics. These difference terms are utilized as the exclusion restrictions; as was shown above, these terms are related to the probability that i was appointed as a bureaucrat, but these terms should not be related to i's ability to properly implement the PMT.

Table 3.7 reports estimation results for these models, where the OLS (Heckman) results are displayed in the first (last) two columns. Across all specifications, the coefficients for the schooling and intelligence are positive and statistically significant at the 1 percent level. Parameter estimates are quite similar across the two estimation methods. In fact, the coefficient for the Mill's λ in the Heckman specification is very close to 0, and highly insignificant. This suggests that OLS estimates of (3.5) are not biased by sample selection. A rationale for this may be that sample selection bias is inherently idiosyncratic for each Chief. By pooling data across a number of villages, the observed sample of bureaucrats may actually capture much of the variability of X_{iv} in the population, allowing for the overall population estimate of β to be unbiased

[88] OLS estimates are qualitatively the same if village fixed-effects are incorporated into the specification.

Table 3.7: *Technical ability of bureaucrats*

	OLS		Heckman (MLE)	
	[1]	[2]	[3]	[4]
$school_{iv}$	0.013***	0.011***	0.013***	0.011***
	(0.002)	(0.002)	(0.004)	(0.003)
$association_{iv}$	0.011	0.009	0.011	0.009
	(0.009)	(0.009)	(0.010)	(0.010)
age_{iv}	-0.001	0.002	-0.001	0.002
	(0.002)	(0.002)	(0.004)	(0.003)
$(age_{iv})^2$	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$business_{iv}$	-0.005	-0.012	-0.005	-0.012
	(0.018)	(0.018)	(0.020)	(0.018)
$male_{iv}$	0.006	-0.002	0.005	-0.002
	(0.013)	(0.012)	(0.014)	(0.014)
$minority_{iv}$	-0.001	-0.007	-0.002	-0.007
	(0.025)	(0.021)	(0.026)	(0.022)
R_{iv}	0.005	0.008	0.005	0.008
	(0.012)	(0.011)	(0.012)	(0.011)
A_{iv}	0.002	0.006	0.002	0.006
	(0.016)	(0.016)	(0.021)	(0.020)
$intel_{iv}$		0.215***		0.216***
		(0.029)		(0.044)
Mill's λ			-0.001	0.001
			(0.038)	(0.030)
N	1,543	1,543	10,452	4,422

Notes. Each column reports parameter estimates from the model that regresses the bureaucrat's PMT exam score on the variables listed in the rows. Columns 1 and 2 report OLS parameter estimates, whereas Columns 3 and 4 report estimates from the (maximum likelihood) Heckman correction model; estimates from the selection equation are unreported. Observations are sample weighted. Standard errors are clustered by village and reported in parenthesis. The star level *** corresponds to statistical significance at 1 percent.

in the bureaucrat sample.

These estimates confirm that, as expected, greater schooling and intelligence are associated with better performance on the PMT exam. It is a question, however, of whether individuals with high human capital and intelligence also perform best in the field, when local knowledge can be utilized to improve accuracy. In particular, some types of bureaucrats may be particularly prone to, say, inappropriately include non-poor households.

Unfortunately, the data do not allow for a bureaucrat to be accurately linked to the individual household that the bureaucrat interviewed.[89] Consequently, the analysis of program accuracy is conducted using aggregated, village-level data. This approach allows us to compare targeting accuracy *across villages*, to investigate how the average characteristics of the committee influences the accuracy of beneficiary selection. The specification

$$\overline{correct}_{lv} = \overline{X}_v\beta_l + \theta_{R,l}\overline{R}_v + \theta_{A,l}\overline{A}_v + Z_v\gamma_l + \delta_s + \epsilon_{lv} \tag{3.6}$$

is estimated, where $\overline{correct}_{lv}$ is described in detail below, $(\overline{X}_v, \overline{R}_v, \overline{A}_v)$ denotes the average of the bureaucrats' characteristics in village v, δ_s is a state fixed-effect and ϵ_{lv} is an error term. The vector Z_v contains village characteristics that may repre-

[89] While collecting the data utilized in this paper, the professional enumerator team conducted a meeting with the bureaucrats in each village. During this meeting, the committee was asked to identify the bureaucrat that interviewed each of the ten sample households in the village. In addition, the main respondent from each sample household was asked to identify the bureaucrat that interviewed the household. For only 29 percent of households, both the committee and respondent household member identified a bureaucrat that conducted an interview. Among this 29 percent of households, the committee and respondent household member identified the same bureaucrat interviewer for only 58 percent of cases. Several possible rationales for these discrepancies are: (i) the household respondent that the professional enumerator interviewed may have been different from the household member that the bureaucrat interviewed; (ii) some bureaucrats may have shirked their interview responsibilities; and (iii) committee members may have not disclosed accurate information during the meeting (e.g., so bureaucrats could not be linked to households for evaluation purposes).

sent potentially important omitted variables. These controls include several of the covariates used in other papers that investigate the effectiveness of decentralized programs in developing countries, such as Olken (2006), Galasso and Ravallion (2005) and Bardhan and Mookherjee (2006), and include the variables listed in Table 3.2.[90]

The dependent variable $\overline{correct}_{lv}$ equals the percent of sample households of livelihood l in village v that were correctly categorized by the targeting program. Two livelihood categories l are considered: households that are poor, and households that are non-poor (according to a livelihood benchmark). For each poor (non-poor) household i, the variable $correct_{ilv}$ is constructed such that this variable equals 1 if i was included on (excluded from) the list of beneficiaries, and 0 otherwise. The variable $\overline{correct}_{lv}$ is then just the weighted sum over the sample households in village v of livelihood l. Two separate ways of classifying households as poor and non-poor are considered: villagers' perceptions of which households are poor and the targeting program's objective PMT decision rule.

Separately considering the share of correctly identified poor and non-poor households allows for the examination of whether some types of committees tend to exclude poor households from the beneficiary list, whereas other types of committees tend to include non-poor households. In particular, it may be the case that committees that are particularly connected to the Chief may be systematically prone to incorrectly include non-poor households.

Table 3.8 displays parameter estimates for specification (3.6). Columns 1 and 2 of the table report parameter estimates when the targeting program's benchmark poverty measure (villagers' livelihood perceptions) is utilized to partition households

[90]The natural log, as opposed to the level, of the number of households in the village and the distance to the state administrative center, are included in Z_v.

Table 3.8: *Committee characteristics associated with accurate selection*

	Perception benchmark		**PMT benchmark**	
	[1]	[2]	[3]	[4]
	Poor subsample	Non-poor subsample	Poor subsample	Non-poor subsample
$\overline{school}_v$	0.016	0.011*	0.014	0.001
	(0.011)	(0.007)	(0.015)	(0.007)
$\overline{association}_v$	0.102**	0.051*	0.036	0.013
	(0.041)	(0.026)	(0.058)	(0.025)
$\overline{age}_v$	-0.004	-0.009	0.006	0.000
	(0.015)	(0.009)	(0.022)	(0.009)
$(\overline{age}_v)^2$	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$\overline{business}_v$	-0.023	0.047	0.112	-0.022
	(0.117)	(0.060)	(0.148)	(0.063)
$\overline{male}_v$	-0.005	0.036	-0.018	0.139***
	(0.065)	(0.043)	(0.083)	(0.047)
$\overline{minority}_v$	-0.287***	-0.020	-0.330***	0.001
	(0.059)	(0.069)	(0.103)	(0.059)
$\overline{iq}_v$	0.215*	-0.163**	-0.068	-0.192**
	(0.130)	(0.078)	(0.152)	(0.088)
$\overline{R}_v$	-0.023	-0.088*	-0.173*	-0.108**
	(0.066)	(0.046)	(0.102)	(0.046)
$\overline{A}_v$	-0.190**	-0.033	-0.093	0.001
	(0.074)	(0.054)	(0.087)	(0.043)
Village controls	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes
R^2	0.219	0.214	0.144	0.287
N	291	293	265	294

Notes. Each column reports parameter estimates from the model that regresses the percentage of sample households that were correctly categorized by the targeting program on the village's average bureaucrat characteristics (as defined in Table 3.1), village- and Council jurisdiction-level controls (as defined in Table 3.2) and state fixed-effects. In Columns 1 and 3, the dependent variable is the percentage of poor sample households that were *correctly included on the beneficiary list*. In Columns 2 and 4, the dependent variable is the percentage of non-poor sample households that were *correctly omitted from the beneficiary list*. Observations are sample weighted. Standard errors are robust to heteroskedasticity and reported in parenthesis. The star levels ***, ** and * correspond to statistical significance at 1, 5 and 10 percent, respectively.

into poor and non-poor subsamples. The last two columns, on the other hand, use the targeting program's official, objective method for classifying households (the PMT) to partition households into poor and non-poor subsamples. The dependent variable in Columns 1 and 3 is the share of poor households in the village that were *included on the list of beneficiaries*, whereas the dependent variable in Columns 2 and 4 is the share of non-poor households that were *excluded from the list of beneficiaries*.[91] In other words, positive parameter estimates will correspond to characteristics that are associated with better bureaucrat performance.

Committees that include more socially active bureaucrats ($\overline{association}_v$) perform better at including poor (excluding non-poor) households when villagers' perceptions are used to partition households into poor and non-poor subsamples. Committees where bureaucrats belonged to, on average, one additional social organization correctly identify 10 percent more poor households as beneficiaries, while simultaneously excluding 5 percent more non-poor households as beneficiaries. A similar pattern does not hold, however, when the targeting program's PMT is used to identify households as poor and non-poor. As discussed above, such a pattern could occur because villagers that are active in the community may have better information regarding the economic condition of other households in their village. Socially active individuals, however, do not appear better equipped to adhere to the PMT, as was shown in Table 3.7.

After controlling for characteristics $\overline{X}_v$, committees that include more relatives or associates of the Chief perform worse at properly identifying households. Seven of

[91] The sample size varies across specifications because in some villages, none of the households were classified as poor (or non-poor) by the benchmark. For instance, in three villages, none of the sample households were perceived as poor by villagers; consequently, the sample size in Column 1 is 291, as opposed to 294.

the eight $(\overline{R}_v, \overline{A}_v)$ coefficients in Table 3.8 are negative in sign, and four of these coefficients are statistically significant (at the 5 or 10 percent level). For instance, about 9 percent more non-poor households (as perceived by other villagers) are included as beneficiaries if a committee is completely comprised of relatives of the Chief, when compared to a committee that contains no relatives of the Chief. For relatives of the Chief, an even stronger pattern emerges when the targeting program's PMT is utilized to partition the households.

While bureaucrats with more schooling perform better at excluding the non-poor from the program, more intelligent bureaucrats are actually *more likely to included non-poor households as beneficiaries*: a 10 percentage point increase in $\overline{intel}_v$ is associated with an additional 1.6 percent of non-poor households being identified as eligible. Interestingly, the same pattern also holds when the targeting program's PMT is utilized to classify households as non-poor; in fact, the magnitude of the coefficient is quite similar. Committees comprised of individuals with higher intelligence, however, also perform better at including households that villagers perceive as poor on the list of beneficiaries.

To summarize, parameter estimates in Tables 3.7 and 3.8 thus suggest that bureaucrats with higher intelligence perform better at properly administering the PMT and including poor households as beneficiaries, but simultaneously tend to include more non-poor households as beneficiaries as well. One possible rationale for this pattern is that intelligent, clever bureaucrats may be better equipped to exploit the targeting program for private gain by strategically including non-poor households on the list of beneficiaries. In addition, committees that include relatives and associates of the Chief tend to perform worse at properly executing the targeting program's procedures. Bureaucrats that are more socially active, however, tend to perform better

at properly identifying households that villagers perceive as poor, and not identifying households that villagers perceive as non-poor.

To conclude this section, parameter estimates from model (3.6) are used to estimate the degree to which suboptimal selection of bureaucrats, from the point of view of the accurate selection of poor households, reduces overall targeting accuracy. In particular, parameter estimates of (3.6), omitting the variable $\overline{intel}_v$, are utilized to predict the percentage of poor and non-poor households that individuals in the sample would have correctly identified, had they been appointed as a bureaucrat.

This procedure utilizes villagers' livelihood perceptions as the relevant benchmark when estimating (3.6), as this measure was the targeting program's ideal objective. Equipped with these hypothetical targeting accuracy estimates, individuals are then ranked within each village from the individual that would have most accurately identified poor households in the village, to the individual that would have performed the worst at identifying poor households in the village. The top-performing b_v individuals are then selected as the hypothetical bureaucrat committee, where b_v denotes the number of committee members in village v. The targeting accuracies of the b_v best performers are then averaged, leading to the estimates $(\overline{\hat{correct}}_{poor,v}, \overline{\hat{correct}}_{non-poor,v})$ for each village.

This procedure implicitly assumes several things. First, this procedure assumes that the committee's "accuracy production function" is linear and additively separable, i.e. $f(X_{1v}, ..., X_{b_v v}) = (\sum_{i=1}^{b_v} X_{iv}/b_v)\beta_l$. This assumption is consistent with the functional form used to estimate (3.6). Second, this procedure assumes that parameter estimates for (3.6) are not biased due to sample selection. In other words, the assumption is that these parameters represent the marginal effects of committee members' characteristics across the entire population of adults in the village. Given

Table 3.9: *Percentage of households correctly identified: Averaged across villages*

	Implementation [1]	Optimal [2]	Difference ([2]-[1]) [3]
Poor households	0.75	0.86	0.11
Non-poor households	0.88	0.94	0.06

Notes. Column 1 reports the average percentage of households correctly categorized by the targeting program, according to the village perceptions benchmark. Column 2 reports the predicted average percentage of households correctly categorized by each village's optimal committee, where the optimal committee was selected to maximize the predicted percentage of poor households identified as beneficiaries (see the text).

that sample selection was rejected for the Heckman model estimation results presented in Table 3.8, this assumption may not be unwarranted. Third, this procedure computes the unconstrained optimal committee; in particular, it does not explicitly take into account the guidelines suggested by the targeting program, or constraints on who the Chief knows in the village. Interestingly, however, the average number of women identified as optimal committee members is 46 percent, implying that the one "hard" constraint imposed by the program (25 percent or more of committee members should be women) is not binding. And fourth, this procedure represents a lower bound on optimal committee performance, as the random sample of villagers is only a small subset of the total distribution of candidate bureaucrats in the village.

Table 3.9 reports the actual and estimated optimal number of households that were correctly identified by the targeting program. The first column reports the share of households that were correctly identified by the targeting program, by the poor household and non-poor household subsamples, in the average village. Column 2, on the other hand, reports the estimated optimal share that were correctly identified, using the procedure outlined in the preceding paragraphs. In the aver-

age village, bureaucrats identified about 75 percent of poor households as eligible for benefits, whereas the optimal bureaucrat committee would have correctly identified 86 percent of households. Suboptimal bureaucrat selection thus reduces the number of beneficiary poor households by about 13 percent (i.e., 11 percentage points). Simultaneously, about 12 percent of non-poor households were improperly identified as beneficiaries in the average village, whereas the optimal committee would have included half as many non-poor households (only 6 percent overall).

3.5 Conclusion

This paper has studied the hiring decisions of local government officials, and evidence suggests that village Chiefs exhibit a tendency to select individuals that are "like them." While in general, Chiefs tend to appoint individuals with high levels of human capital and cognitive ability, they nevertheless also tend to appoint their social associates, as well as individuals that exhibit similar observable characteristics. The paper considers whether nepotism, cronyism or information asymmetries drive this pattern, and some preliminary evidence is consistent with the cronyism mechanism. This hiring pattern leads to a suboptimal bureaucrat selection, and an overall reduction in targeting accuracy.

At this stage, the paper can be improved along several dimensions. First, the paper must address the issue of sample selection with respect to cognitive ability: the cognitive test was only administered to one (non-randomly selected) member per household, indicating that this sample of candidate bureaucrats is selected. Second, the paper must delve more deeply into understanding the various possible mechanisms for why the appointment patterns occur, and improving the test for determining which

mechanism is primarily driving the pattern. Finally, and perhaps most importantly, it will be important to address the potential endogeneity of the village Chief's characteristics (i.e., across villages).

Appendix A

Data description for Chapter 1

The first five subsections in this appendix describe each of our five raw data sources. The sixth subsection then describes how these data are merged together. The seventh subsection describes the International City/County Management Association (ICMA) political activity data.

County and City Data Book (CCDB)

The CCDB includes demographic information from the decennial Census of Population and Housing for all US cities with population 25,000 or more. The Interuniversity Consortium for Political and Social Research (ICPSR) has combined all rounds of the CCDB from 1944 to 1977 into one file (ICPSR code number 7735), and also has posted the 1983 CCDB (ICPSR code number 8265). The University of Virginia (UVA) has publicly posted CCDB information for 1994 and 2000.

We extract information for the covariates that are included in our regressions for 1970, 1980, 1990 and 2000. In addition, we also extract information for median household income and population for 1960. For all years, median household income

is adjusted for inflation to 2005 dollars.

Unfortunately, information for several variables is missing for some years in the panel. We manually update the CCDB for these problems in the following way:

- The 1994 CCDB omits the unemployment rate for 30 cities. We acquire the missing information from the Census Bureau's American FactFinder website.

- The 2000 CCDB omits several key variables for cities with population between 25,000 and 100,000.[92] We acquire the missing information from the Census Bureau's American FactFinder website.

- Across the panel of CCDBs, information for serious crimes per 100,000 inhabitants is missing for a number of cities (24 cities in 1970, 6 cities in 1980, 101 cities in 1990 and 97 cities in 2000). The CCDB compiles its crime data from the FBI's Uniform Crime Reporting Program. We update the missing information by manually searching the FBI's Uniform Crime Reporting Program documents back to the late 1960s, which are available from the ICPSR. For the 2000 CCDB, data were also gathered from the Illinois State Police's Uniform Crime Reporting website, which lists detailed crime statistics for cities in the state of Illinois. A document listing the source for each value is available from the authors upon request.

Census of Governments (CoG)

Every five years, the US Census Bureau conducts the CoG, which collects detailed financial and employment information from all government bodies within the United

[92]The variables are the percentage of high school graduates residing in the city; the proportion of college graduates residing in the city; the percentage of city residents living below the poverty line; and the city's median household income.

States. We gather information directly from the Census Bureau for 1972, 1982, 1992 and 2002. We extract information on operational and capital expenditures, the aggregate municipal wage bill paid to full-time employees and the wage bill disaggregated for the five services we study. Spending information is adjusted for inflation to 2005 dollars.

NBER Public Sector Collective Bargaining Law Data Set

This data set provides information regarding state-level labor laws for public sector workers. Information is reported for five types of public sector employees: state employees, municipal police, municipal fire fighters, non-college teachers and other local employees. We extract information on workers' right to bargain collectively for (i) municipal police, (ii) municipal fire fighters and (iii) other local employees.

The collective bargaining rights variable is coded into seven categories:

0 = No provision regarding collective bargaining

1 = Collective bargaining is prohibited

2 = Municipality is authorized but not required to bargain with union

3 = Union has right to present proposals

4 = Union has right to meet and confer with municipality

5 = Municipality implicitly obligated to bargain with union

6 = Municipality explicitly obligated to bargain with union

For each state, year and employee category, we define the employee category as having legal bargaining power if the collective bargaining rights variable equals 5 or 6.

The NBER data set includes information from 1955 through 1985. We extract information for 1972 and 1982. We manually update the data set for 1992 and 2002; information on these categorizations and the source of information is available from the authors upon request.[93]

County-Level Political Variables

We compute our measure of resident turnover using the ICPSR county-level data derived from the decennial Census of Population and Housing. These data include variables that allow for computing the fraction of county residents that remain living in a county through time.[94] We downloaded the relevant county data for 1970, 1980, 1990 and 2000 (ICPSR code number 9694, 9693, 2889 and 13402, respectively).

To compute county population 5 years prior, we interpolate between Census years assuming a constant exponential population growth rate between $t-10$ and t. To interpolate a value for 1965, we acquire information by downloading the NBER's Decennial County Population Data 1900-1990,[95] which reports county population in 1960. As discussed in the text, the variable $stayers_{it}$ is constructed by computing the number of county i's year t residents that lived in county i in $t-5$, and then dividing this number by the total number of residents that lived in county i in year $t-5$.

County-level Presidential voting data is downloaded from the ICPSR ("General Election Data for the United States," ICPSR code number 13). This data contains voting data for all counties through 1990. County-level voting data for the 2000

[93] There were relatively few changes in state laws between 1982 and 2002.

[94] One variable reports county population in each census year. A second set of variables identify the number of county residents that lived in the same county 5 years prior to the census year. See the paper's main text for a description of how we compute our variable of interest.

[95] The ICPSR data series begin in 1970. The NBER data set is available for download from `http://www.nber.org/data/census-decennial-population.html`.

Presidential election is downloaded from an online database create by American University's Center for Congressional and Presidential Studies.[96] We define the variable dem_share_{it} by dividing the number of votes for the Democratic Presidential nominee by the sum of the number of votes for the Democratic and Republican nominees. We match the 1968 (1980) [1988] {2000} voting data to the 1970 (1980) [1990] {2000} Decennial Census data.

National Oceanic and Atmospheric Administration (NOAA) US Cities Location List

The NOAA website contains a document that lists the latitudinal and longitudinal location of all US cities. This information is used to compute the distances between all cities in our sample. We employ the haversine formula to compute the spherical distance.

Creating the Panel

To create the panel for our main analysis, we merge together the CCDB data (for the Census years 1960, 1970, 1980, 1990 and 2000) with the CoG and NBER data (1972, 1982, 1992 and 2002) and the county-level information. Our sample is selected according to the following criteria:

1. First, we retain all cities which had a population of at least 25,000 for all census years between 1960 and 2000. This results in 619 cities.

2. We drop the six cities that did not report median household income information in 1960.[97]

[96] See `http://www.american.edu/spa/ccps/Data-Sets.cfm`.

[97] The dropped cities are: Bell Gardens, CA; Temple City, CA; Carson, CA; West Haven, CT;

3. We drop the two cities in Alaska and Hawaii (Anchorage, AK; Honolulu, HI).

4. We drop the 15 cities for which we could not locate crime information corresponding to one or more census years from 1970 to 2000.[98]

5. We drop one city that did not report wage bill information in 1990 (East Point, GA).

This results in a final sample of 595 cities.

International City/County Management Association (ICMA)

In 1988 and 1999, the ICMA conducted surveys of municipal governments regarding public sector labor relations.[99] One survey question stated, "Please indicate whether employee unions/associations have engaged in the following political activities since July 1, 1988. (Check all applicable.)" The questionnaire displayed a five-by-seven table with blank spaces for the respondent to fill in.

The five rows listed five separate employee groups: police protection (sworn), fire protection (sworn), sanitation, public works (nonsanitation) and all other employees (except education personnel).

The seven columns listed seven types of political activities: candidate endorsements, candidate financial contributions, time or in-kind campaign contributions, mismanagement disclosure threats, state level lobbying, publicity campaigns and taking issues to referendum.

Overland Park, KS; Dearborn Heights, MI.

[98]The dropped cities are: Clinton, IA; Lawrence, KS; Hutchinson, KS; Salina, KS; Rockville, MD; Holyoke, MA; Chelsea, MA; Biloxi, MS; Santa Fe, NM; Valley Stream, NY; East Cleveland, OH; Parma, OH; Findlay, OH; Elyria, OH; Warren, OH.

[99]This data was purchased directly from the ICMA.

We define the indicator variable $activity_{it}$ to equal 1 if any of the five employee groups engaged in either (i) candidate endorsements, (ii) candidate financial contributions or (iii) time or in-kind campaign contributions, as these activities correspond to offering political support to candidates for local office. In general, $activity_{it} = 1$ only if either police or fire protection workers engaged in these activities.

We then merge the 1988 (1999) $activity_{it}$ variable to the 1990 (2000) Census data defined above. Sample size is as follows:

- In 1988, the ICMA mailed surveys to 2,715 municipalities, of which 1,373 (50.6%) responded. Of the respondents, 238 cities match our sample of 595 municipalities.
- In 1999, the ICMA mailed surveys to 2,881 municipalities, of which 1,401 (48.6%) responded. Of the respondents, 257 cities match our sample of 595 municipalities.
- Out of these municipalities, 136 responded to both the 1988 and 1999 surveys.

We retain these 136 municipal observations for our fixed effect estimation.

Appendix B

Additional tables for Chapter 1

Table B.1: *Aggregate wage bill, in the cross-section*

	[1] 1970	[2] 1980	[3] 1990	[4] 2000
$\log(z/z_{-10})$	-0.065	-0.149	0.219	-0.266
	(0.126)	(0.157)	(0.184)	(0.153)
$\log(y/y_{-10})$	0.558	0.779**	-0.043	-0.309
	(0.320)	(0.268)	(0.258)	(0.283)
$\log(z)$	1.169***	1.158***	1.139***	1.145***
	(0.022)	(0.021)	(0.022)	(0.022)
$\log(y)$	0.525	0.118	0.945***	0.724**
	(0.313)	(0.297)	(0.223)	(0.271)
unem	0.758	0.958	0.789	1.100
	(1.697)	(0.955)	(1.075)	(1.171)
poverty	1.829*	-0.259	2.730***	0.072
	(0.819)	(0.810)	(0.777)	(0.698)
college	0.126	0.443	-0.343	-0.255
	(0.494)	(0.476)	(0.373)	(0.377)
schage	-2.168*	0.503	-4.011***	0.270
	(0.874)	(1.105)	(1.083)	(0.909)
old	1.779**	1.298*	1.684**	1.672**
	(0.641)	(0.593)	(0.566)	(0.636)
black	-1.246	0.271	0.258	0.079
	(0.895)	(0.403)	(0.287)	(0.256)
white	-1.730	0.011	0.077	-0.179
	(0.891)	(0.395)	(0.291)	(0.235)
log(ineq)	0.137	0.913**	0.471	1.294***
	(0.291)	(0.343)	(0.263)	(0.221)
log(crime)	0.024	0.150**	0.138***	0.199***
	(0.048)	(0.050)	(0.041)	(0.049)
log(land area pc)	0.127***	0.079*	0.070*	0.083*
	(0.031)	(0.034)	(0.035)	(0.036)
boundary	-0.043	-0.086	-0.051	0.352*
	(0.057)	(0.095)	(0.123)	(0.179)
Number of Cities	595	595	595	595

Notes. The dependent variable in each regression is the natural log of the wage bill paid to full time municipal employees. Each column represents an estimation using a cross-section of the data, where the year is denoted by column. All regressions include state fixed effects, which are unreported. Heteroskedastic-robust standard errors are reported in parentheses. Star levels ***, ** and * correspond to statistical significance at the 1, 5 and 10 percent levels, respectively.

Table B.2: *Fixed-effect estimations incorporating political activity*

	[1]	[2]
decline	0.029	-0.008
	(0.058)	(0.059)
decline * activity		0.088
		(0.057)
activity		-0.003
		(0.061)
B	0.098	0.091
	(0.091)	(0.124)
vote share dem	0.238	0.224
	(0.415)	(0.418)
stayers	-0.546	-0.284
	(0.854)	(0.925)
log(z)	0.642**	0.594**
	(0.208)	(0.197)
log(y)	0.764	0.715
	(0.487)	(0.490)
unem	0.197	0.069
	(1.638)	(1.668)
poverty	-0.181	-0.159
	(0.662)	(0.646)
college	-0.278	-0.224
	(0.943)	(0.977)
schage	-1.907	-2.144
	(2.321)	(2.354)
old	0.332	-0.084
	(1.626)	(1.635)
black	0.755	0.613
	(0.843)	(0.863)
white	0.904	0.916
	(0.667)	(0.664)
log($ineq$)	0.072	0.017
	(0.653)	(0.656)
log($crime$)	-0.030	-0.021
	(0.051)	(0.050)
land area pc	-0.371	-0.376
	(0.278)	(0.279)
boundary	-0.013	-0.014
	(0.121)	(0.116)
Number of Cities	136	136

Notes. This table reports fixed effect coefficient estimates after merging the data with the ICMA data set. The data are a balanced panel of 136 cities across two time periods (1990 and 2000). The dependent variable in each regression is the natural log of the wage bill paid to full time municipal employees. The variable $decline_{it}$ equals 1 if city i experienced population decline between $t-10$ and t, and 0 otherwise. The variable $activity_{it}$ equals 1 if city i's public workers engaged in political activity at time t (as defined in the text) and 0 otherwise. Both regressions include city and year fixed effects, which are unreported. Heteroskedastic-robust standard errors are reported in parentheses. Star levels ***, ** and * correspond to statistical significance at the 1, 5 and 10 percent levels, respectively.

Table B.3: *Full estimation results for total wage bill: By service*

	Fire		Police		Roads		Parks		Admin	
Variable	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
$\log(z/z_{-10})$	-0.158	-3.313	-0.204	-3.333	0.027	0.244	-0.008	-0.069	-0.037	-0.512
$\log(y/y_{-10})$	-0.327	-3.970	-0.246	-2.359	-0.282	-1.439	-0.139	-0.673	0.022	0.175
B	0.058	3.005	0.084	3.463	0.136	2.826	0.143	2.745	0.132	4.357
vote share dem	-0.149	-1.210	-0.166	-1.077	0.308	1.049	-0.593	-1.988	0.337	1.777
stayers	0.034	0.144	0.039	0.129	0.192	0.329	0.365	0.611	0.252	0.676
rho	0.149	25.187	0.111	21.515	-0.011	-0.234	0.065	2.497	0.055	2.432
$\log(z)$	0.799	18.852	0.763	14.628	0.867	8.678	0.827	8.063	0.850	13.269
$\log(y)$	0.655	5.352	0.601	3.878	0.438	1.488	0.636	2.064	0.450	2.389
unem	-0.852	-2.315	-0.172	-0.367	-1.197	-1.349	-1.398	-1.474	-1.682	-2.955
poverty	-0.565	-2.114	-0.502	-1.494	-0.021	-0.033	-0.060	-0.090	-0.470	-1.139
college	-0.127	-0.734	-0.037	-0.170	0.388	0.930	-0.051	-0.116	0.090	0.336
schage	0.221	0.522	-0.192	-0.346	2.836	2.802	-2.117	-2.032	0.516	0.799
old	-0.238	-0.757	0.370	0.923	2.344	3.089	1.654	2.103	0.920	1.895
black	0.412	2.559	0.276	1.295	-0.154	-0.382	-0.199	-0.493	0.189	0.765
white	-0.306	-2.202	-0.250	-1.385	0.083	0.248	-0.191	-0.561	-0.306	-1.445
$\log(ineq)$	0.098	0.884	0.127	0.902	0.024	0.092	-0.084	-0.309	0.489	2.861
$\log(crime)$	0.053	3.413	0.069	3.551	0.032	0.855	0.029	0.768	0.009	0.407
land area pc	0.007	0.239	-0.045	-1.152	0.126	1.662	0.133	1.688	0.230	4.697
boundary	0.019	0.730	0.038	1.165	-0.074	-1.170	-0.078	-1.187	-0.065	-1.605

Note. This table is continued on the following page.

Table B.3: *Continued*

	Fire		Police		Roads		Parks		Admin	
Variable	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
W-log(z/z_{-10})	-0.031	-0.383	-0.020	-0.199	-0.279	-1.412	0.065	0.315	-0.140	-1.096
W-log(y/y_{-10})	0.185	1.672	0.216	1.563	0.662	2.472	-0.004	-0.017	0.157	0.925
W-vote share dem	0.153	0.933	0.208	1.011	0.364	0.927	1.150	2.821	-0.383	-1.514
W-stayers	0.304	0.864	0.754	1.728	1.294	1.530	-0.132	-0.147	-0.206	-0.380
W-log(z)	-0.037	-0.615	-0.013	-0.180	0.216	1.344	0.328	2.065	0.179	1.833
W-log(y)	0.011	0.067	-0.337	-1.521	0.217	0.516	-0.564	-1.275	0.132	0.482
W-unem	0.037	0.076	0.890	1.433	0.788	0.661	-0.160	-0.128	0.532	0.695
W-poverty	-0.565	-1.344	-0.984	-1.920	-0.040	-0.040	-1.841	-1.764	-0.573	-0.901
W-college	-0.932	-3.579	-0.549	-1.740	-0.273	-0.454	0.560	0.867	-0.154	-0.393
W-schage	-0.214	-0.287	0.140	0.149	0.572	0.317	0.755	0.413	0.452	0.392
W-old	-0.651	-1.149	0.072	0.100	1.215	0.886	0.857	0.623	-0.296	-0.338
W-black	0.100	0.348	0.234	0.642	0.735	1.063	0.683	0.948	0.237	0.537
W-white	-0.134	-0.639	-0.064	-0.240	0.440	0.887	-0.275	-0.527	-0.440	-1.384
W-log($ineq$)	0.490	2.505	0.435	1.774	0.211	0.453	0.794	1.693	0.025	0.083
W-log($crime$)	0.084	2.675	0.048	1.243	0.057	0.758	0.142	1.757	0.089	1.833
W-land area pc	0.012	0.226	-0.054	-0.780	0.125	0.945	0.284	2.036	0.071	0.838
W-boundary	0.020	0.525	-0.030	-0.613	-0.123	-1.303	-0.201	-2.038	0.014	0.240
Number of Cities	589		563		576		547		595	

Notes. This table reports the full estimation results from Table 8 in the text. The dependent variable in each regression is the natural log of the wage bill paid to full time municipal employees that produce the public good specified in the column name. Significance levels are not starred.

Appendix C

Omitted proofs for Chapter 1

Proof of Lemma 1

We first establish a lemma that will be referenced in the proof of Lemma 1.

Lemma 2. *Let T be an operator over the space of continuous, bounded functions on $G \times S$ such that*

$$T(\hat{U})(g_{-1}, s) = \max_{g \in G(s)} \left\{ u(g, g_{-1}) - \phi\Big(\lambda\tilde{v}(s) - \lambda v(g, s)\Big) + \beta \int \hat{U}(g, \gamma s + \epsilon) d\Gamma(\epsilon) \right\}.$$

T has a unique fixed point U. $\forall s \in S$, U is differentiable, strictly concave and strictly increasing in g_{-1}. Moreover, $\exists!$ function $g^*(g_{-1}, s)$ such that

$$\begin{aligned} g^*(g_{-1}, s) = \arg\max_{g \in G(s)} \Bigg\{ & u(g, g_{-1}) - \phi\Big(\lambda\tilde{v}(s) - \lambda v(g, s)\Big) \\ & + \beta \int U(g, \gamma s + \epsilon) d\Gamma(\epsilon) \Bigg\} \end{aligned} \tag{C.1}$$

where, $\forall s \in S$, $g^*(\cdot, s)$ is continuous.

Proof of Lemma 2. The proof follows from standard dynamic programing arguments,

as discussed in Stokey, Lucas, and Prescott (1989). Note that $G \subset \Re_+$ and $S \subset \Re_+^3$ are compact and convex, while $\Gamma(\epsilon)$ satisfies the Feller property and $G(s)$ is nonempty, compact-valued and continuous. Moreover, $u(\cdot)$ is continuously differentiable, bounded, strictly concave over G^2 and strictly increasing in g_{-1}.

Note, however, that $\lim_{g \to x} v(g,s) = -\infty$ for $x \in \{0, yz\}$, which poses a technical problem. As is standard, we circumvent this problem by redefining the constraint correspondence to be $G^\epsilon(s) = [\epsilon, yz - \epsilon]$ for some $\epsilon > 0$ small.[100] Then $v(\cdot, s)$ is continuously differentiable, bounded and strictly concave over $G^\epsilon(s)$, which implies that $-\phi(\lambda\tilde{v}(s) - \lambda v(g,s))$ satisfies the same properties.[101] Theorems 9.6-9.8 and 9.10 of Stokey, Lucas, and Prescott (1989) imply the lemma. □

Proof of Lemma 1. Let $g^*(g_{-1}, s)$ be as defined in (C.1),

$$C^*(g_{-1}, s) \quad = \quad \lambda\tilde{v}(s) - \lambda v(g^*(g_{-1}, s), s) \quad > \quad 0$$

and $g^{**}(g_{-1}, s, \hat{g}, \hat{C}) = \hat{g}$. We show that $(g^*(\cdot), C^*(\cdot), g^{**}(\cdot))$ constitutes an MPE. Note that $(g^*(\cdot), C^*(\cdot))$ is continuous in g_{-1} by Lemma 2 and $v(\cdot)$ continuous. Given

[100] Here, we define a sufficient ϵ. Note that $g^*(g_{-1}, s) > \tilde{g}(s)$ by $u_g, U_g, \phi' > 0$. Consequently, $\epsilon < \min_{s \in S} \tilde{g}(s)$ does not constrain $g^*(\cdot)$ from the left. Moreover, $\lim_{g \to xy} \lambda v_g(g,s) \cdot \phi'(\lambda\tilde{v}(s) - \lambda v(g,s)) = -\infty$, whereas

$$\sup_{(s, g_{-1}) \in S \times G} \left\{ u_g(\tilde{g}(s), g_{-1}) + \beta \int U_g(\tilde{g}(s), \gamma s + \epsilon) d\Gamma(\epsilon) \right\} \quad < \quad H$$

for some $H \in \Re_{++}$. Then we can implicitly define $\hat{g}(s)$ as $\lambda v_g(\hat{g}(s), s) \cdot \phi'[\lambda\tilde{v}(s) - \lambda v(\hat{g}(s), s)] = -H$. Note that $\hat{g}(s)$ is continuous and satisfies $\hat{g}(s) < yz$. For any $g' \in (\hat{g}(s), yz)$, note that

$$u_g(g', g_{-1}) + \lambda v_g(g', s) \cdot \phi'\Big(\lambda\tilde{v}(s) - \lambda v(g', s)\Big) + \beta \int U_g(g', \gamma s + \epsilon) d\Gamma(\epsilon) \quad < \quad 0,$$

and thus restricting $g < \hat{g}(s)$ does not constraint $g^*(\cdot)$ from the right. Therefore, selecting $\hat{\epsilon} = \min\{\min_{s \in S} \tilde{g}(s), \min_{s \in S} yz - \hat{g}(s)\}$ is without loss of generality.

[101] To see strict concavity, note that the second order condition is $\lambda v_{gg}\phi' - [\lambda v_g]^2 \phi'' < 0$, where the inequality follows from $v_{gg} < 0$, $\phi' > 0$ and $\phi'' \geq 0$.

these strategies, we can recursively define the functions

$$\begin{aligned} P(g_{-1},s) &= \lambda v\Big(g^*(g_{-1},s),s\Big) \;+\; C^*(g_{-1},s) \;+\; \beta\int P\Big(g^*(g_{-1},s),\gamma s+\epsilon\Big)\Gamma(\epsilon) \\ U(g_{-1},s) &= u\Big(g^*(g_{-1},s),g_{-1}\Big) \;-\; \phi\Big(C^*(g_{-1},s)\Big) \\ &\quad + \beta\int U\Big(g^*(g_{-1},s),\gamma s+\epsilon\Big)\Gamma(\epsilon), \end{aligned}$$

which are continuous in g_{-1}. Since $P(\cdot,s)$ is continuous, then

$$\tilde{P}(s) \;=\; \max_{g\in G(s)}\left\{\lambda v(g,s)+\beta\int P(g,\gamma s+\epsilon)\Gamma(\epsilon)\right\}$$

is well-defined. A union offer (g',C') must satisfy

$$\lambda v\Big(g',s\Big)+C'+\beta\int P(g',s),\gamma s+\epsilon)d\Gamma(\epsilon) \;\geq\; \tilde{P}(s) \tag{C.2}$$

to incentivize the party to select (g',C'). The union's problem is thus

$$\max_{(\hat{g},\hat{C})\in G(s)\times\Re_+}\left\{u(\hat{g},g_{-1})-\phi(\hat{C})+\beta\int U(\hat{g},\gamma s+\epsilon)d\Gamma(\epsilon)\right\}$$

subject to (C.2). The constraint (C.2) is clearly solved with equality, and thus for every $(g_{-1},s)\in G\times S$, the party receives $\tilde{P}(s)$ value. Thus, (C.2) can be reduced to

$$\lambda v(\hat{g},s)+\hat{C}+\beta\int\tilde{P}(\gamma s+\epsilon)d\Gamma(\epsilon) \;=\; \lambda\tilde{v}(s)+\beta\int\tilde{P}(\gamma s+\epsilon)d\Gamma(\epsilon),$$

and thus $\hat{C}=\lambda\tilde{v}(s)-\lambda v(\hat{g},s)$. Consequently, the union's problem is

$$\max_{\hat{g}\in G(s)}\left\{u(\hat{g},g_{-1})-\phi\Big(\lambda\tilde{v}(s)-\lambda v(\hat{g},s)\Big)+\beta\int U(\hat{g},\gamma s+\epsilon)d\Gamma(\epsilon)\right\}, \tag{C.3}$$

which is the problem studied in Lemma 2. □

Proof of Proposition 1

The Euler Equation from (C.3) is

$$\begin{aligned} u_g(g, g_{-1}) + \lambda v_g(g, s)\phi'\Big(\lambda\tilde{v}(s) - \lambda v(g, s)\Big) & \\ + \beta \int U_g(g, \gamma s + \epsilon) d\Gamma(\epsilon) &= 0. \end{aligned} \tag{C.4}$$

Note that $g^*(g_{-1}, s) > \tilde{g}(s)$ by $u_g, U_g, \phi' > 0$. The Implicit Function Theorem implies[102]

$$\frac{dg^*}{dg_{-1}} = -\frac{u_{gg_{-1}}}{u_{gg} + \lambda v_{gg}\phi' - [\lambda v_g]^2\phi'' + \beta \int U_{gg} d\Gamma(\epsilon)}$$

which is strictly positive by $u_{gg_{-1}} > 0$ and the denominator being strictly negative (by (C.3) being strictly concave). Again, the Implicit Function Theorem implies that for $x \in \{y, z\}$,

$$\frac{dg^*}{dx} = -\frac{\lambda v_{gx}\phi' + \lambda^2 v_g[\tilde{v}_x - v_x]\phi'' + \beta\gamma_x \int U_{gx} d\Gamma(\epsilon)}{u_{gg} + \lambda v_{gg}\phi' - [\lambda v_g]^2\phi'' + \beta \int U_{gg} d\Gamma(\epsilon)}. \tag{C.5}$$

Consider the case for y. Recall that (i) $v_{gy} > 0$, which implies that $\tilde{v}_y(s) < v_y(g^*, s)$; (ii) $v_g(g^*, s) < 0$ by $g^*(g_{-1}, s) > \tilde{g}(s)$; and (iii) the denominator is strictly negative. Consequently, $U_{gy}(g^*, s') \geq 0$ is sufficient for $dg^*/dy > 0$. Let s^j be the history of the exogenous state vector from the current state, s_0, through the j-period ahead state,

[102] As is standard, we assume that $U(\cdot)$ is twice-differentiable.

s_j. Let

$$g^{*,j}(g_{-1}, s^j) \quad = \quad g^*(\underbrace{g^*(...g^*(g_{-1}, s_0), ...)}_{j-1 \text{ nested policies}}, s_j)$$

be the j-step ahead equilibrium policy, given g_{-1} and history s^j. Since $g^*_{g_{-1}}(g_{-1}, s) > 0$, then $\forall j$, $g^{*,j}_{g_{-1}}(g_{-1}, s^j) > 0$. Consequently,

$$\begin{aligned} U_{yg}(g_{-1}, s) \quad = \quad & \sum_{j=0}^{\infty} (\beta\gamma_y)^j E_{s^j|s_0} \Bigg\{ \Big[\lambda v_{gy}(g^{*,j}, s_j)\phi'(\cdot) \\ & + \quad \lambda^2 v_g(g^{*,j}, s_j)\big[\tilde{v}_y(s) - v_y(g^{*,j}, s_j)\big]\phi''(\cdot)\Big] \cdot \prod_{i=0}^{j} g^{*,i}_{g_{-1}}(g_{-1}, s^j) \Bigg\} \end{aligned}$$

is strictly positive. To conclude, note that $dg^*/dy_{-1} = [dg^*/dg_{-1}] \cdot [dg_{-1}/dy_{-1}] > 0$. The argument for dg^*/dz_{-1} is analogous.

Proof of Proposition 2

Proposition 1 established that $dg^*/dx_{-1} = [dg^*/dg_{-1}] \cdot [dg_{-1}/dx_{-1}] > 0$ for $x_{-1} \in \{y_{-1}, z_{-1}\}$. Consequently, $y \cdot [dg/dy] > z \cdot [dg/dz]$ is sufficient for the result. From (C.5), it follows that if $\gamma_z = 0$, then it is sufficient to show that

$$y \cdot \Big[\lambda v_{gy}\phi' + \lambda^2 v_g[\tilde{v}_y - v_y]\phi''\Big] \quad > \quad z \cdot \Big[\lambda v_{gz}\phi' + \lambda^2 v_g[\tilde{v}_z - v_z]\phi''\Big].$$

First, we show that $yv_{gy}(g,s) > zv_{gz}(g,s)$ for $g > \tilde{g}(s)$:[103]

$$\begin{aligned} y \cdot \frac{\partial^2 v}{\partial g \partial y}(g,s) &= -(\sigma-1)(1-\eta)yz^{2-\sigma}[zy-g]^{\sigma-2} \\ &> -(\sigma-1)(1-\eta)yz^{2-\sigma}[zy-g]^{\sigma-2} \quad - \quad (\sigma-1)\frac{\partial v}{\partial g}(g,s) \\ &= z \cdot \frac{\partial^2 v}{\partial g \partial z}(g,s), \end{aligned}$$

where the inequality follows from $v_g(g,s) < 0$ when $g > \tilde{g}(s)$. Next, we show that $y[\tilde{v}_y(s) - v_y(g,s)] \leq z[\tilde{v}_z(s) - v_z(g,s)]$:[104]

$$\begin{aligned} y[\tilde{v}_y(s) - v_y(g,s)] &= (1-\eta)yz^{2-\sigma}\Big[[zy-\tilde{g}(s)]^{\sigma-1} - [zy-g]^{\sigma-1}\Big] \\ &\leq (1-\eta)yz^{2-\sigma}\Big[[zy-\tilde{g}(s)]^{\sigma-1} - [zy-g]^{\sigma-1}\Big] \\ &\quad - \quad (\sigma-1)\Big[\tilde{v}(s) - v(g,s)\Big] \\ &= z[\tilde{v}_z(s) - v_z(g,s)], \end{aligned}$$

where the inequality follows from $\tilde{v}(s) \geq v(g,s)$.

Proof of Proposition 3

From (C.4), the Implicit Function Theorem implies that

$$\frac{dg^*}{dw} = -\frac{\lambda v_{gw}\phi' + \lambda^2 v_g[\tilde{v}_w - v_w]\phi'' + \beta\gamma_w \int U_{gw} d\Gamma(\epsilon)}{u_{gg} + \lambda v_{gg}\phi' - [\lambda v_g]^2\phi'' + \beta \int U_{gg} d\Gamma(\epsilon)}$$

[103] Recall that $g^*(\cdot) > \tilde{g}(s)$ from Proposition 1.
[104] Recall that $v_g < 0$ when $g > \tilde{g}(s)$.

which is positive by $v_{gw}(g,s) > 0$ and analogous arguments as in the proof of Proposition 1. If $\gamma_w = 0$ and $\phi''(\cdot) = 0$, then

$$\begin{aligned}\frac{\partial g^*}{\partial w}\cdot\frac{w}{g^*} &= -\frac{\lambda v_{gw}(g^*,w)\cdot\phi'(\cdot)}{u_{gg}(g^*,g_{-1})+\lambda v_{gg}(g^*,w)\cdot\phi'(\cdot)+\beta\int U_{gg}(g^*,\gamma s+\epsilon)d\Gamma(\epsilon)}\cdot\frac{w}{g^*}\\ &< -\frac{v_{gw}(g^*,s)}{v_{gg}(g^*,s)}\cdot\frac{w}{g^*},\end{aligned}$$

where the inequality follows from $-u_{gg}(g^*,g_{-1})-\beta\int U_{gg}(g^*,\gamma s+\epsilon)d\Gamma(\epsilon) > 0$. Define $f(g,s) = -[v_{gw}(g,s)/v_{gg}(g,s)]\cdot[w/g]$. We show that $f_g(g,s) < 0$. Note that

$$\begin{aligned}\frac{\partial f}{\partial g} &= \frac{-v_{ggw}v_{gg}g + v_{gw}[v_{gg}+v_{ggg}g]}{[gv_{gg}]^2}\cdot w\\ &= \frac{w}{[gv_{gg}]^2}\cdot\Bigg\{\frac{z^{2-2\sigma}\eta^2 g^{2\sigma-3}}{w^{2\sigma+1}}\cdot\underbrace{\Big[\sigma(\sigma-1)^2-\sigma(\sigma-1)-\sigma(\sigma-1)(\sigma-2)\Big]}_{A}\\ &\quad+\frac{z^{2-2\sigma}\eta(1-\eta)g^{\sigma-1}[y-\frac{g}{z}]^{\sigma-3}}{w^{\sigma+1}}\\ &\quad\cdot\underbrace{\Big[[yz-g]\Big[\sigma(\sigma-1)^2-\sigma(\sigma-1)\Big]+g\Big[\sigma(\sigma-1)(\sigma-2)\Big]\Big]}_{B}\Bigg\}.\end{aligned}$$

Note that $A = \sigma(\sigma-1)[\sigma-1-1-\sigma+2] = 0$. Moreover, $B = yz[\sigma(\sigma-1)(\sigma-2)] < 0$, where the inequality follows from $\sigma < 0$. Consequently, $f_g(g,s) < 0$. Since $g^*(g_{-1},s) > \tilde{g}(s)$ (by Proposition 1), then

$$-\frac{v_{gw}(g^*,s)}{v_{gg}(g^*,s)}\cdot\frac{w}{g^*} < -\frac{v_{gw}(\tilde{g},s)}{v_{gg}(\tilde{g},s)}\cdot\frac{w}{\tilde{g}} = \frac{\partial\tilde{g}}{\partial w}\cdot\frac{w}{\tilde{g}}.$$

Bibliography

Ai, C. and E. C. Norton (2003): "Interaction terms in logit and probit models," *Economics Letters*, 80, 123–129.

Aigner, D. J. and G. G. Cain (1977): "Statistical theories of discrimination in labor markets," *Industrial and Labor Relations Review*, 30, 175–187.

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias (2010): "Targeting the poor: Evidence from a field experiment in Indonesia," Tech. rep., National Bureau of Economic Research.

Alderman, H. (2002): "Do local officials know something we don't? Decentralization of targeted transfers in Albania," *Journal of Public Economics*, 83, 375–404.

Alesina, A., R. Baqir, and W. Easterly (1999): "Public goods and ethnic divisions," *Quarterly Journal of Economics*, 114, 1243–1284.

——— (2000): "Redistributive public employment," *Journal of Urban Economics*, 48, 219241.

Alesina, A. and G. Tabellini (2007): "Bureaucrats or politicians? Part I: A single policy task," *The American Economic Review*, 97, 169–179.

——— (2008): "Bureaucrats or politicians? Part II: Multiple policy tasks," *Journal of Public Economics*, 92, 426–447.

Anderson, K. (2010): *The political economy of agricultural price distortions*, Cambridge University Press.

Anonymous (2011): "(Government) workers of the world unite!" *The Economist*.

Anselin, L., J. L. Gallo, and H. Jayet (2006): "Spatial panel econometrics," *The Econometrics of Panel Data*, 625–660.

Araujo, M. C., F. H. Ferreira, P. Lanjouw, and B. Özler (2008): "Local inequality and project choice: Theory and evidence from Ecuador," *Journal of Public Economics*, 92, 1022–1046.

Banks, J. S. and R. K. Sundaram (1993): "Adverse selection and moral hazard in a repeated elections model," in *Political economy: Institutions, competition, and representation: Proceedings of the Seventh International Symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett and M. J. Hinich, Cambridge University Press, 295–311.

Baqir, R. (2002): "Districting and government overspending," *Journal of Political Economy*, 110, 1318–1354.

Bardhan, P. and D. Mookherjee (2005): "Decentralizing antipoverty program delivery in developing countries," *Journal of Public Economics*, 89, 675–704.

——— (2006): "Pro-poor targeting and accountability of local governments in West Bengal," *Journal of Development Economics*, 79, 303–327.

——— (2010): "Determinants of redistributive politics: An empirical analysis of land reforms in West Bengal, India," *American Economic Review*, 100, 1572–1600.

Bassetto, M. and L. McGranahan (2009): "On the relationship between mobility, population growth, and capital spending in the United States," Working paper.

Beer, C. and A. Riedl (2009): "Modeling spatial externalities: A panel data approach," *mimeo*.

Bergstrom, T. C., D. L. Rubinfeld, and P. Shapiro (1982): "Micro-based estimates of demand functions for local school expenditures," *Econometrica*, 1183–1205.

Besley, T. (2006): *Principled agents?: The political economy of good government*, Oxford University Press, USA.

Besley, T. and R. Burgess (2002): "The political economy of government responsiveness: Theory and evidence from India," *Quarterly Journal of Economics*, 117, 1415–1451.

Besley, T. and S. Coate (2003): "Elected versus appointed regulators: Theory and evidence," *Journal of the European Economic Association*, 1, 1176–1206.

Besley, T., R. Pande, and V. Rao (2005): "Political selection and the quality of government: Evidence from South India," *CEPR Discussion Paper No. 5201*.

——— (2011): "Just rewards? Local politics and public resource allocation in South India," *The World Bank Economic Review*.

Brouwers, S. A., F. J. Van de Vijver, and D. A. Van Hemert (2009): "Variation in Raven's Progressive Matrices scores across time and place," *Learning and Individual Differences*, 19, 330–338.

Camacho, A. and E. Conover (2011): "Manipulation of social program eligibility," *American Economic Journal: Economic Policy*, 3, 41–65.

Castañeda, T. (2005): "Targeting social spending to the poor with proxy-means testing: Colombia's SISBEN system," *Social Protection Discussion Papers*.

Cawley, J., J. Heckman, and E. Vytlacil (2001): "Three observations on wages and measured cognitive ability," *Labour Economics*, 8, 419–442.

Coate, S. and B. G. Knight (2009): "Government form and public spending: Theory and evidence from US municipalities," *NBER Working Paper*.

Coate, S. and S. Morris (1999): "Policy persistence," *American Economic Review*, 89, 1327–1336.

Corman, H. and H. N. Mocan (2000): "A time-series analysis of crime, deterrence, and drug abuse in New York City," *The American Economic Review*, 90, 584–604.

Day, K. M. and S. L. Winer (2006): "Policy-induced internal migration: An empirical investigation of the Canadian case," *International Tax and Public Finance*, 13, 535–564.

Dixit, A. and J. Londregan (1996): "The determinants of success of special interests in redistributive politics," *Journal of Politics*, 1132–1155.

Drazen, A. and M. Eslava (2010): "Electoral manipulation via voter-friendly spending: Theory and evidence," *Journal of Development Economics*, 92, 39–52.

Drèze, J. and A. Sen (1989): *Hunger and public action*, Oxford University Press.

Elhorst, J. P. (2003): "Specification and estimation of spatial panel data models," *International Regional Science Review*, 26, 244–268.

Enikolopov, R. (2010): "Politicians, bureaucrats and targeted redistribution: The role of career concerns," Working paper.

Ferraz, C. and F. Finan (2011): "Electoral accountability and corruption: Evidence from the audits of local governments," *American Economic Review*, 101, 1274–1311.

Ferreira, F. and J. Gyourko (2009): "Do political parties matter? Evidence from US cities," *Quarterly Journal of Economics*, 124, 399–422.

Freeman, R. B. (1986): "Unionism comes to the public sector," *Journal of Economic Literature*, 41–86.

Freeman, R. B., C. Ichniowski, and H. Lauer (1989): "Collective bargaining laws, threat effects, and the determination of police compensation," *Journal of Labor Economics*, 7, 191–209.

Galasso, E. and M. Ravallion (2005): "Decentralized targeting of an antipoverty program," *Journal of Public Economics*, 89, 705–727.

Goldberg, P. K. and G. Maggi (1999): "Protection for sale: An empirical investigation," *American Economic Review*, 89, 1135–1155.

Gray, S. (2009): "Mayor Dave Bing: Can he stop the slide in Detroit?" *Time*.

Greenwood, M. J. (1997): "Internal migration in developed countries," *Handbook of Population and Family Economics*, 1, 647–720.

Grossman, G. M. and E. Helpman (1994): "Protection for sale," *American Economic Review*, 84, 833–850.

Hamermesh, D. S. and G. A. Pfann (1996): "Adjustment costs in factor demand," *Journal of Economic Literature*, 34, 1264–1292.

Hoxby, C. M. (1996): "How teachers' unions affect education production," *The Quarterly Journal of Economics*, 671–718.

Iyer, L. and A. Mani (forthcoming): "Traveling agents: Political change and bureaucratic turnover in India," *The Review of Economics and Statistics*.

Jones, C. (2008): "Vallejo votes to declare Chapter 9 bankruptcy," *San Francisco Chronicle*.

Kneebone, R. D. and K. J. McKenzie (2001): "Electoral and partisan cycles in fiscal policy: An examination of Canadian provinces," *International Tax and Public Finance*, 8, 753–774.

Krueger, A. (1993): *Political economy of policy reform in developing countries*, MIT Press.

Ladd, H. F. (1994): "Fiscal impacts of local population growth: A conceptual and empirical analysis," *Regional Science and Urban Economics*, 24, 661–686.

Lee, L. and J. Yu (2010): "Estimation of spatial autoregressive panel data models with fixed effects," *Journal of Econometrics*, 154, 165–185.

LeSage, J. and R. K. Pace (2009): *Introduction to spatial econometrics*, Chapman & Hall/CRC.

Levitt, S. D. (2002): "Using electoral cycles in police hiring to estimate the effect of police on crime: Reply," *The American Economic Review*, 92, 1244–50.

Lindbeck, A. and J. W. Weibull (1987): "Balanced-budget redistribution as the outcome of political competition," *Public Choice*, 52, 273–297.

MacDonald, L. (2008): "The impact of government structure on local public expenditures," *Public Choice*, 136, 457–473.

Maskin, E. and J. Tirole (2004): "The politician and the judge: Accountability in government," *The American Economic Review*, 94, 1034–1054.

McFadden, D. (1978): "Modelling the choice of residential location," *Spatial Interaction Theory and Planning Models*, 25, 75–96.

Moe, T. M. (2006): "Political control and the power of the agent," *Journal of Law, Economics, and Organization*, 22, 1–29.

Najita, J. M. and J. L. Stern (2001): *Collective bargaining in the public sector: The experience of eight states*, ME Sharpe Inc.

Niehaus, P., A. Atanassova, M. Bertrand, and S. Mullainathan (2011): "Targeting with agents," Working paper.

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

——— (2006): "On the theory and practice of fiscal decentralization," *Institute for Federalism and Intergovernmental Relations Working Paper*.

Olken, B. A. (2006): "Corruption and the costs of redistribution: Micro evidence from Indonesia," *Journal of Public Economics*, 90, 853–870.

——— (2007): "Monitoring corruption: Evidence from a field experiment in Indonesia," *Journal of Political Economy*, 115, 200–249.

Phelps, E. S. (1972): "The statistical theory of racism and sexism," *The American Economic Review*, 62, 659–661.

Reinikka, R. and J. Svensson (2004): "Local capture: Evidence from a central government transfer program in Uganda," *Quarterly Journal of Economics*, 119, 679–705.

Rhode, P. W. and K. S. Strumpf (2003): "Assessing the importance of Tiebout sorting: Local heterogeneity from 1850 to 1990," *The American Economic Review*, 93, 1648–1677.

Schady, N. R. (2000): "The political economy of expenditures by the Peruvian Social Fund (FONCODES), 1991-95," *American Political Science Review*, 289–304.

Stokey, N. L., R. E. Lucas, and E. C. Prescott (1989): *Recursive methods in economic dynamics*, Harvard University Press.

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

Varghese, R. (2010): "Harrisburg to make bond payment with state help," *The Wall Street Journal*.

Wooldridge, J. M. (2002): *Econometric analysis of cross section and panel data*, The MIT press.