On the Optimal Number of Representatives

Emmanuelle Auriol* and Robert Gary-Bobo†

January 23, 2008

[1]Toulouse School of Economics

[2]Ecole d'Economie de Paris

INTRODUCTION

♣ Democracy is the direct power by, and for, the people.

♣ In spite of the extension of democratic values in recent period direct democracy is rare.

♣ In practice the decisions relative to public goods are made by a subsets of agents: *representative democracy.*

National ⇒ parliament

Local ⇒ councils

Clubs ⇒ committees

♣ Traditional literature on public goods focuses on direct democracy.

♣ Most of the work in formal political science uses voting models of direct representation.

♣ Goal: to understand the recourse to restricted representation, and its welfare implications.

The costs of direct democracy:

⇒ Size of the Government.

⇒ Competence of the citizens.

⇒ Structural instability.

THE INGREDIENTS

♣ A problem of public decision-making in decentralized economies: Information is *decentralized*

⇒ no common knowledge

⇒ adverse selection.

♣ There is no Super-agent: the Principal is defined as the agent or set of agents in charge of the public decisions.

⇒ The Principal rationally pursues his (her) private interest (opportunism).

♣ The representatives are n agents drawn at random out of the population of size N.

$\Rightarrow$ produce information for collective decision-making.

Remark : the representatives are drawn at random while most political systems rely on elections.

$\Rightarrow$ Random sampling is a radical way to ensure that representatives will resemble the represented people.

♣ The approach takes three types of constraint into consideration:

(i) constraint associated to the absence of common knowledge;

(ii) adverse selection constraint;

(iii) robustness to opportunistic manipulation constraint.

THE MODEL

♣ $N+1$ agents: $i = 0, \ldots, N$.

♣ Public decision $x \in X$ and taxes/subsidies t_i.

Assumption 1 Quasi-linear preferences, $v_i(x) - t_i$, with quadratic valuation function:

$$v_i(x) = \theta_i x - \frac{x^2}{2}$$

Assumption 2 θ_i are IID from distribution P on Θ, with $E_P(\theta) = \mu$ and $Var_P(\theta) = \sigma^2$

Assumption 3 Reporting information is costly: $F_i = F > 0$ if i is a representative; $F_i = 0$ otherwise.

Assumption 4 The n representatives are independent random drawings in the probability distribution P.

Definition 1 A *representation mechanism* is an array of functions (f, t), where f is a collective decision rule mapping representatives' reports about preferences $\widehat{\theta} = (\widehat{\theta}_1, ..., \widehat{\theta}_n)$ into X, i.e., $x = f(\widehat{\theta})$, and a list of tax functions denoted $t = (t_0, t_1, ...t_N)$, satisfying the budget constraint $\sum_{i=0}^{N} t_i = 0$.

GOAL: Find optimal mechanism (f, t, n)

FIRST BEST OPTIMUM

♣ There exists a benevolent, bayesian, utilitarian Planner who knows P (μ and σ^2)

♣ The Planner

a) chooses a sample size $n \leq N$;

b) randomly draws n agents relabeled from 1 to n;

d) checks type $\theta = (\theta_1, ..., \theta_n)$;

e) chooses public production: $x = f(\theta)$;

f) taxes all agents so that: $\sum_{i=0}^{N} t_i(\theta) = 0$.

♣ The first best objective is to choose x and n so as to maximize :

$$EW = E_P \left\{ \sum_{i=0}^{N} (v_i(x) - t_i) \mid (\theta_1, ..., \theta_n) \right\} - nF$$

♣ The first best objective is to choose x so as to maximize :

$$W(\theta, n) = \Big\{\sum_{i=1}^{n} \theta_i + (N+1-n)\mu\Big\}x - (N+1)\frac{x^2}{2} - nF$$

$\Rightarrow$ The "Statistical" Samuelson production rule:

$$f^*(\theta) = \frac{1}{N+1}\Big[\sum_{i=1}^{n} \theta_i + (N+1-n)\mu\Big]$$

♣ Substituting $x = f^*(\theta)$ in $W(\theta, n)$ and taking the expected value wrt P yields:

$$W(n) = \frac{n\sigma^2}{2(N+1)} + \frac{(N+1)\mu^2}{2} - nF$$

Proposition 1 The first best optimum is either a form of *Direct Democracy*, $n^* = N+1$ and $x^* = \frac{1}{N+1}\sum_{i=0}^{N}\theta_i$, or the exact opposite, an ideal *Reign of Tradition*, $n^* = 0$ and $x^* = \mu$.

SECOND BEST OPTIMUM

(a) There is no benevolent planner $\Rightarrow$ the agent(s) carrying out the mechanism has preferences, limited knowledge, and can manipulate decisions.

(b) θ_i are not observable $\Rightarrow$ revelation problem.

(c) P is not common knowledge $\Rightarrow$ decentralized knowledge problem.

Lemma 1 The first best optimum is not implementable under (a), (b), (c).

PROOF:

+ The Principal is an economic agent.

+ Since μ is not known, and not common knowledge, the Principal chooses μ according to her own preferences.

$\Rightarrow$ the ideal *Reign of Tradition* becomes dictatorship.

Definition 2 (Separation of Power) The executive cannot be a representative.

Definition 3 (Subsidiarity) If the parameters needed to implement (f, t) are not specified in the constitution, and are not provided for by the representatives according to constitutional rules, then, they are chosen by the executive.

Definition 4 (Non-Manipulability) A representation mechanism (f, t) is nonmanipulable if it is revealing in dominant strategies and if all its parameters are specified in the constitution.

Incentive Compatibility
Auriol and Gary-Bobo (2007)

Proposition 2. *Assume that the separation-of-powers, subsidiarity and anonymity principles hold. Assume that any utility function v is possible. Then, (f, t) is nonmanipulable if and only if the following conditions hold:*

(1) $f(\widehat{v}) \in \arg\max_{x \in X} \{\sum_{i=1}^{n} \widehat{v}_i(x) + k(x)\}$

and for all $i = 1, ..., n$;

(2) $t_i(\widehat{v}) = -\sum_{j \neq i} \widehat{v}_i(f(\widehat{v})) - k(f(\widehat{v})) + m(\widehat{v}_{-i})$,

where m doesn't depend on v_i; and finally,

(3) *k and m are fixed in the constitution.*

♣ Proof: "only if" part is difficult.

♣ When $k \equiv 0$ we obtain Clarkes-Groves mechanisms restricted to a random subset of agents

$\Rightarrow$ budget-balanced by construction.

Robust Representation Mechanisms

♣ Benevolent, Bayesian, and Utilitarian, Founding Fathers (FF) write the constitution;

♣ There is no omniscient, impartial and benevolent individual to carry out public decisions;

♣ FF don't know the probability distribution P, and they know that nobody does. They have a vague prior, B, on possible distributions P.

♣ Let B be the vague (non informative) FF prior on P. We assume that:

$$E_B(\mu_P) = \widehat{\mu},\ E_B(\sigma_P^2) = \widehat{\sigma}^2, Var_B(\mu_P) = \widehat{z}^2$$

♣ Prior variance of θ from FF point of view:

$$Var_{FF}(\theta) = Var_B[E(\theta|P)] + E_B[Var(\theta|P)]$$

$$\Leftrightarrow \qquad Var_{FF}(\theta) = \widehat{z}^2 + \widehat{\sigma}^2$$

Definition 5 (Robust Representation Mechanism). A mechanism (f, t) is robust if it is the limit of a sequence (f_k, t_k) of mechanisms, such that each (f_k, t_k) maximizes $E_{B_k}(E_P W)$ on the set of nonmanipulable mechanisms, where (B_k) is a sequence of priors with the property that that $\widehat{z}_k^2$ goes to $+\infty$, while $\widehat{\sigma}_k^2/\widehat{z}_k^2$ goes to zero.

Proposition 3. *Under Assumptions 1-4, the only robust mechanism* $f^*(\widehat{\theta})$ *maximizes* $\sum_{i=1}^{n} \widehat{v}_i(x)$, *with transfers* t *given by Proposition 2 above.*

$$\Rightarrow f^*(\widehat{\theta}_1, ..., \widehat{\theta}_n) = \frac{\sum_{i=1}^{n} \widehat{\theta}_i}{n}.$$

$\Rightarrow$ Sampling Clarke-Groves mechanisms.

PROOF: Let $v_i(x) = \theta_i x - x^2/2$.

♣ Non-manipulability forces to choose α and $\beta \geq 0$ s.t.
$f^* \in \arg\max_x \left\{ x \sum_{i=1}^n \widehat{\theta}_i - \frac{nx^2}{2} + \beta \left(\alpha x - \frac{x^2}{2} \right) \right\}$.

$\Rightarrow f^*(\widehat{\theta}) = \frac{\sum_{i=1}^n \widehat{\theta}_i + \alpha\beta}{n+\beta}$.

♣ $W_P(\alpha, \beta) = E_P \left\{ f^* \sum_{i=0}^N \theta_i - \frac{(N+1) f^{*2}}{2} \right\} - nF$.

♣ $E_B W_P = (n + \beta - \frac{N+1}{2}) \frac{n\widehat{\sigma}^2}{(n+\beta)^2} + \frac{b^2(N+1)}{2(n+\beta)^2} (2\alpha\widehat{\mu} - \alpha^2) + \frac{n(N+1)}{(n+\beta)^2} (\frac{n}{2} + \beta)(\widehat{\mu}^2 + \widehat{z}^2) - nF$.

Lemma 2. *For given B, the optimal values of α and β are*

$$\alpha^* = \widehat{\mu} \quad \text{and } \beta^* = \frac{(N+1-n)\widehat{\sigma}^2}{\widehat{\sigma}^2 + (N+1)\widehat{z}^2}.$$

$$\Rightarrow \lim_{\widehat{z}^2 \to +\infty} \beta^* = 0.$$

♣ The only robust mechanism entails $v_0(x) = \widehat{\mu}x - x^2/2$ and $\beta^* = 0$. Therefore the arbitrary function k must be set identically equal to 0.

Optimal Number of Representatives

Lemma 2 *Welfare loss due to sampling is:*

$$L(n) = \left[\frac{1}{n} - \frac{1}{N+1}\right] \frac{(N+1)\widehat{\sigma}^2}{2}.$$

PROOF:

$\rightarrow$ Substituting $x^* = (1/n)\sum_{i=1}^{n}\theta_i$ in the second best expected welfare yields:

$$\overline{W} = \tfrac{N+1}{2}(\widehat{\mu}^2 + \widehat{z}^2) + \tfrac{\widehat{\sigma}^2}{2} - \left[\tfrac{1}{n} - \tfrac{1}{N+1}\right]\tfrac{(N+1)\widehat{\sigma}^2}{2} - nF.$$

$\rightarrow$ First best expected welfare yields:

$$E_B E_P(W^*) = \tfrac{\widehat{\sigma}^2}{2} + \tfrac{N+1}{2}(\widehat{\mu}^2 + \widehat{z}^2) - nF.$$

$\rightarrow L(n) = E_B E_P(W^*) - \overline{W}$

Proposition 4. *The optimal number of representatives is 1 plus the integer part of,*

$$n^{**} = \widehat{\sigma}\sqrt{\frac{N+1}{2F}}.$$

Empirical Assessment on Political Data

♣ Proposition 4's formula suggests an econometric model of the form:

$$\log(n) = (1/2)\log(N+1)+\log(\sigma)-(1/2)\log(F)+\epsilon,$$

♣ To empirically predict the size of representative political institutions, we have assembled a data set for a sample of 111 countries which possess representative assemblies.

The Square-Root Model with World Data

- $REPRE$: total number of representatives in numbers of individuals (Europa World Year Book 1995).

- POP: population size in million (EWYB95).

-GNP: per capita in dollars 1996 (WB 98).

-$TAXREV$: central government tax revenue as a percentage of GDP 1996 (WB 98).

-$DENSITY$: population per square kilometer divided by 10,000 (WB 98).

-ELF: Ethno-linguistic fragmentation (Easterly and Levine Data set WB 03).

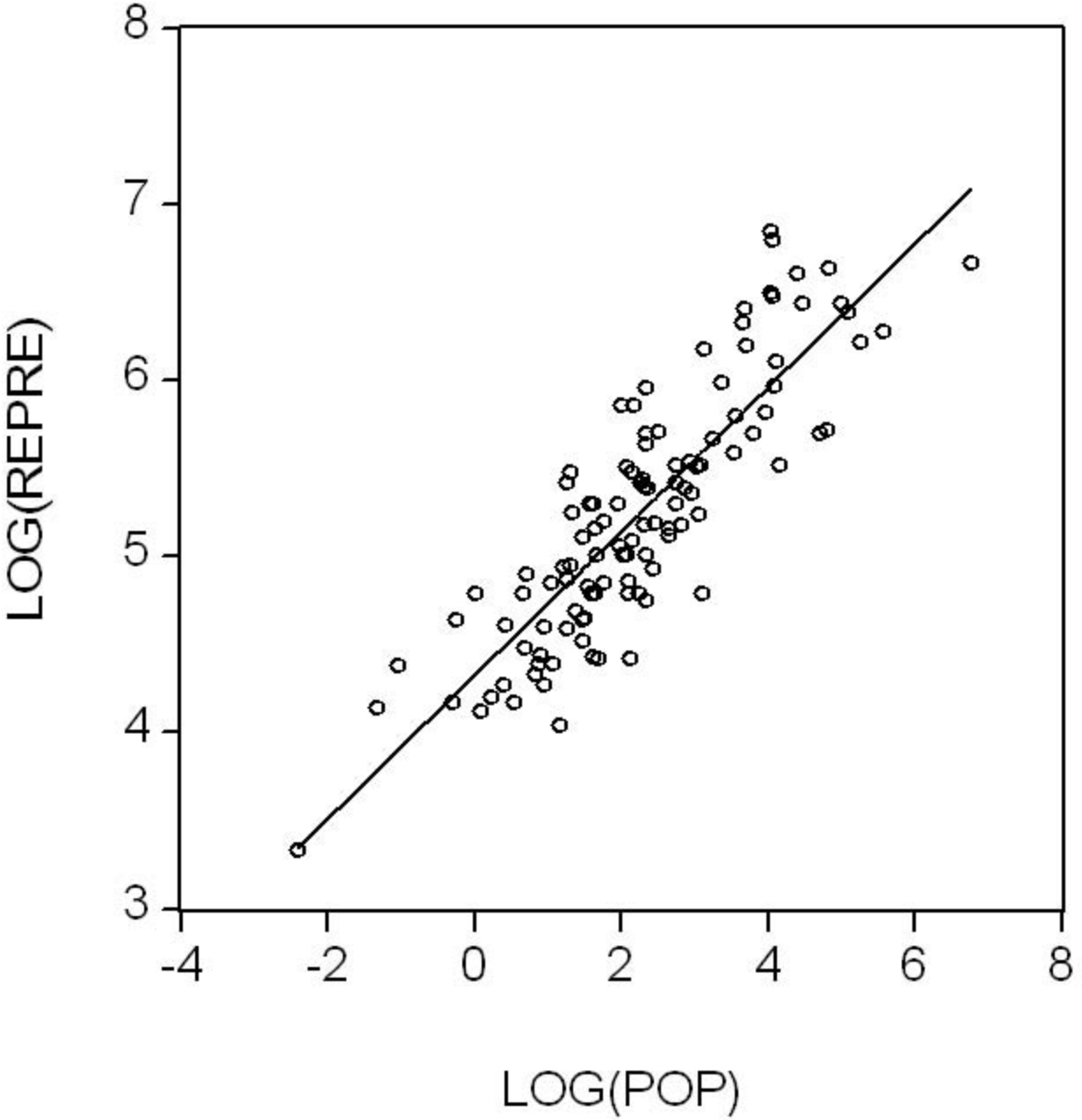
-$GINI$: Gini index divided by 100 (WB98).

Dependent Variable: $log(REPRE)$

	(1)	(2)	(3)	(4)	(5)
Constant	4.32	2.98	5.46	4.28	3.98
	(75.26)***	(7.9)***	(27.16)***	(10.0)***	(6.55)***
log(POP)	0.41	0.44	0.4	0.41	0.44
	(17.63)***	(14.71)***	(16.12)***	(16.88)***	(12.9)***
log(GNP)		0.04		0.04	0.02
		(1.21)		(1.53)	(0.55)
log(TAXREV)		0.34		0.17	0.24
		(2.98)***		(1.96)*	(1.76)*
log(GOVWAGE)		-0.12			-0.18
		(-2.08)**			(-1.82)*
DENSITY			-0.0001	-0.0001	- 8.5×10^{-5}
			(-2.48)**	(-2.96)***	(2.26)**
GINI			-2.53	-1.79	-1.1
			(-5.48)***	(-3.79)***	(-1.73)*
ELF			-1.74	-1.25	-1.37
			(-3.85)***	(-3.07)***	(-2.41)**
GINI×ELF			3.54	2.69	2.95
			(3.49)***	(3.06)***	(2.43)**
No. Obs.	111	62	93	93	55
Adjusted R^2	0.74	0.82	0.82	0.84	0.83

Columns (1)–(5) were estimated by ordinary least squares. White heteroskedastic-consistent standard errors are used to calculate t-statistics, which are reported in parentheses. Significance is denoted by *** (1%); ** (5%); * (10%).

Figure 1



Robustness Check: Controls to the basic regression

- $AFRICA$: dummy that equals 1 if the country is on the African continent.

- $OECD$: dummy that equals 1 if a country belongs to the OECD organization.

- $TRANS$: dummy that equals 1 if it is in transition.

- $DEM46$: dummy that equals 1 if the country has been democratic in all 46 years (1950-95) (Treisman 2000).

- $FEDERAL$: dummy that equals 1 if the country has a federal structure (T00).

- $FORMBRITCOL$: dummy that equals 1 if the country is a former British colony or is the UK (T00).

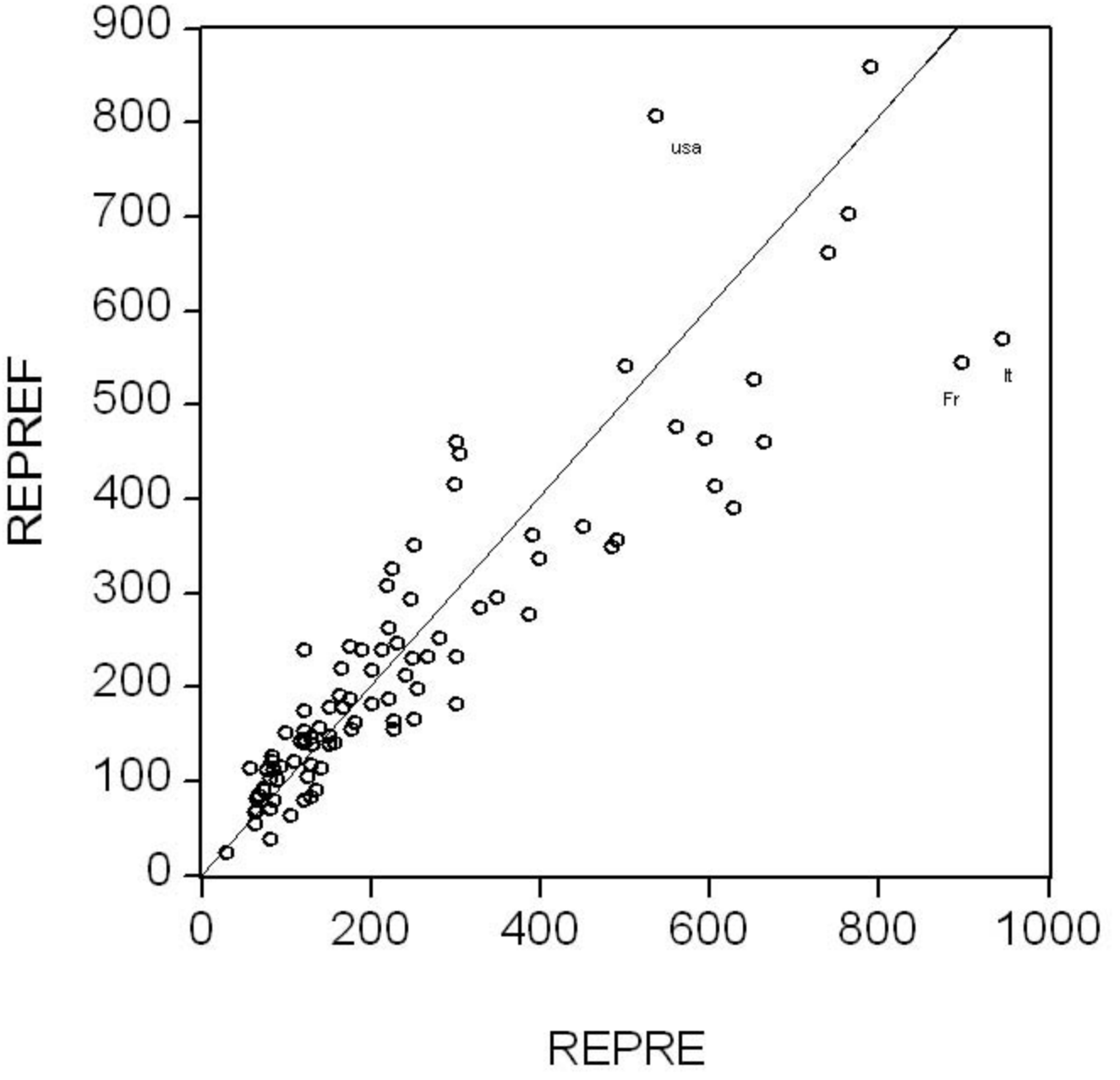
- $COMLAW$: dummy that equals 1 if the country has a common law system (T00).

- $PERCENTPROT$: percentage of protestant in the total population (T00).

Dependent Variable: $log(REPRE)$

	(1)	(2)	(3)	(4)	(5)
Constant	3.4	4.21	4.03	4.22	4.02
	(7.16)***	(9.97)***	(9.34)***	(10.21)***	(11.67)***
log(POP)	0.45	0.41	0.4	0.4	0.39
	(15.27)***	(16.79)***	(16.66)***	(16.16)***	(18.32)***
log(GNP)	0.08	0.07	0.06	0.04	0.02
	(2.11)**	(2.35)**	(1.69)*	(1.09)	(0.76)
log(TAXREV)	0.3	0.16	0.12	0.16	0.15
	(3.19)***	(1.94)*	(1.37)	(1.94)*	(1.91)*
DENSITY	-0.0001	-0.0001	-0.0001	- 0.0001	- 0.0001
	(-2.27)**	(-3.44)***	(-2.82)***	(-2.67)***	(-2.9)***
GINI	-1.63	-2.1	-1.37	-1.57	-0.71
	(-3.49)***	(-4.32)***	(-2.24)**	(-2.72)***	(-2.21)**
ELF	-1.22	-1.31	-0.83	-0.98	
	(-2.67)***	(-3.15)***	(-1.74)*	(-2.07)**	
GINI×ELF	2.7	2.89	1.69	2.18	
	(2.74)***	(3.26)***	(1.6)	(2.16)**	
DEM46	-0.19	-0.21	-0.22	-0.24	-0.2
	(-1.27)	(-1.92)*	(-1.82)*	(-2.31)**	(-1.96)*
PERCENTPROT	-0.0006				
	(-0.34)				
FEDERAL	0.009				
	(0.085)				
COMLAW	-0.13				
	(-1.1)				
FORMBRITCOL	0.06				
	(0.54)				
OECD			0.24	0.26	0.31
			(2.08)**	(2.13)**	(3.11)***
TRANS			0.10		
			(0.85)		
AFRICA			0.14		
			(1.28)		
N. of Obs.	71	93	93	93	111

Figure 2


900
800
700
600
500
400
300
200
100
0
REPREF
usa
Fr
It
0
200
400
600
800
1000
REPRE

Size of the Legislature and Economic Liberalism

The dependent variables are:

-$FREEOPEN$: measure of free trade openness (Barro-Lee 94)

- $SUNKCOST$: direct cost as a fraction of 1999 GDP per capita associated with meeting government requirements to open a new business (Djankov et al. 2002).

-$STATEINTERF$: Index based on a survey of executives in top and middle management in relevant countries which measure whether state interference does hinder the development of business in respondents' country (1=no interference ; 6=strong interference). The data are equal to 6 minus the indexes presented in Treisman 2000.

- $TISCORE$: 1996 Transparency International scores, by default Transparency International score 1998.

Dependent Variable:	FREEOP			
	(1)	(2)	(3)	(4)
Constant	0.09	0.22	0.11	0.14
	(1.3)	(3.8)***	(3.33)***	(5.39)***
log(n)	0.01	-0.024		
	(0.83)	(-2.39)**		
log(N)	-0.02			-0.02
	(-3.09)***			(-3.09)***
log(GNP)	0.015	0.017	0.02	0.02
	(3.76)***	(4.65)***	(3.05)***	4.48***
LAND	-1.6×10^{-5}	-1.89×10^{-5}	-2.19×10^{-5}	-1.65×10^{-5}
	(-5.3)***	(-5.31)***	(-5.5)***	(-5.37)***
DEM46	0.05	0.054	0.048	0.05
	(2.51)**	(2.53)***	(2.53)**	(2.69)***
log(n)-log($\widehat{n}$)			0.01	
			(0.49)	
No. Obs.	67	67	67	67
R^2	0.65	0.61	0.57	0.64
Adjusted R^2	0.62	0.58	0.54	0.62
Sum squared Resid	0.13	0.15	0.16	0.13

All columns were estimated by ordinary least squares. White heteroskedastic-consistent standard errors are used to calculate t-statistics, which are reported in parentheses. Significance is denoted by *** (1%); ** (5%); * (10%).

Dependent Variable:	STATEINTERF	STATEINTERF	SUNKCOST	SUNKCOST
	(1a)	(1b)	(2a)	(2b)
Constant	1.58	4.65	0.56	2.32
	(1.45)	(6.22)***	(1.18)	(3.38)***
log(n)	0.47		0.45	
	(2.2)**		(2.26)**	
log(N)	-0.08		-0.25	
	(-0.63)		(-2.29)**	
log(GNP)	-0.19	-0.26	-0.23	-0.23
	(-2.31)**	(-3.0)***	(-3.05)***	(-3.07)***
DEM46	-0.4	-0.38		
	(-2.27)**	(-1.81)*		
TRANS			-0.34	-0.3
			(-2.16)**	(-2.07)**
log(n)-log($\widehat{n}$)		0.51		0.42
		(2.08)**		(2.22)**
No. Obs.	45	45	71	71
R^2	0.43	0.38	0.3	0.28
Adjusted R^2	0.37	0.34	0.25	0.25
Sum squared Resid	12.32	13.32	20.05	20.46

All columns were estimated by ordinary least squares. White heteroskedastic-consistent standard errors are used to calculate t-statistics, which are reported in parentheses. Significance is denoted by *** (1%); ** (5%); * (10%).

Dependent Variable: TISCORE

	(1)	(2)	(3)	(4)
Constant	14.94	13.69	13.82	12.17
	(13.28)***	(11.31)***	(10.14)***	(10.21)***
REPRE	0.0018	0.0017	0.002	0.0023
	(2.61)**	(2.07)**	(1.93)*	(3.51)***
COMLAW	-0.18	-0.17	-0.12	
	(-0.45)	(-0.46)	(-0.32)	
FORMBRITCOL	-0.72	-0.56	-0.64	
	(-1.61)	(-1.28)	(-1.46)	
PERCENTPROT	-0.026	-0.014	-0.01	-0.01
	(-3.85)***	(-2.15)**	(-1.77)*	(-3.22)***
ELF	-0.24	-0.2	-0.42	
	(-0.33)	(-0.31)	(-0.57)	
MAT1	-0.008	-0.013	-0.01	
	(-0.69)	(-1.19)	(-1.16)	
log(GNP)	-1.29	-1.02	-1.03	-0.81
	(-9.24)***	-6.8)***	(-5.85)***	(-5.08)***
FEDERAL		0.95	0.93	
		(2.35)**	(2.06)**	
DEM46		-1.53	-1.72	-1.63
		(-3.34)***	(-3.57)***	(-4.73)***
FREEOP			0.3	
			(0.13)	
AFRICA				-1.12
				(-2.81)***
OECD				-0.76
				(-1.95)*
No. Obs.	69	69	56	79
Adjusted R^2	0.67	0.71	0.71	0.75

Columns (1) to (4) were estimated by ordinary least squares. White heteroskedastic-consistent standard errors are used to calculate t-statistics, which are reported in parentheses. Significance is denoted by *** (1%); ** (5%); * (10%).

Regression with 50 parliament of the US States

♣ Data from McCormick and Turner (2001), for the 50 US state legislatures in 1996 (state senators + representatives $=n$)

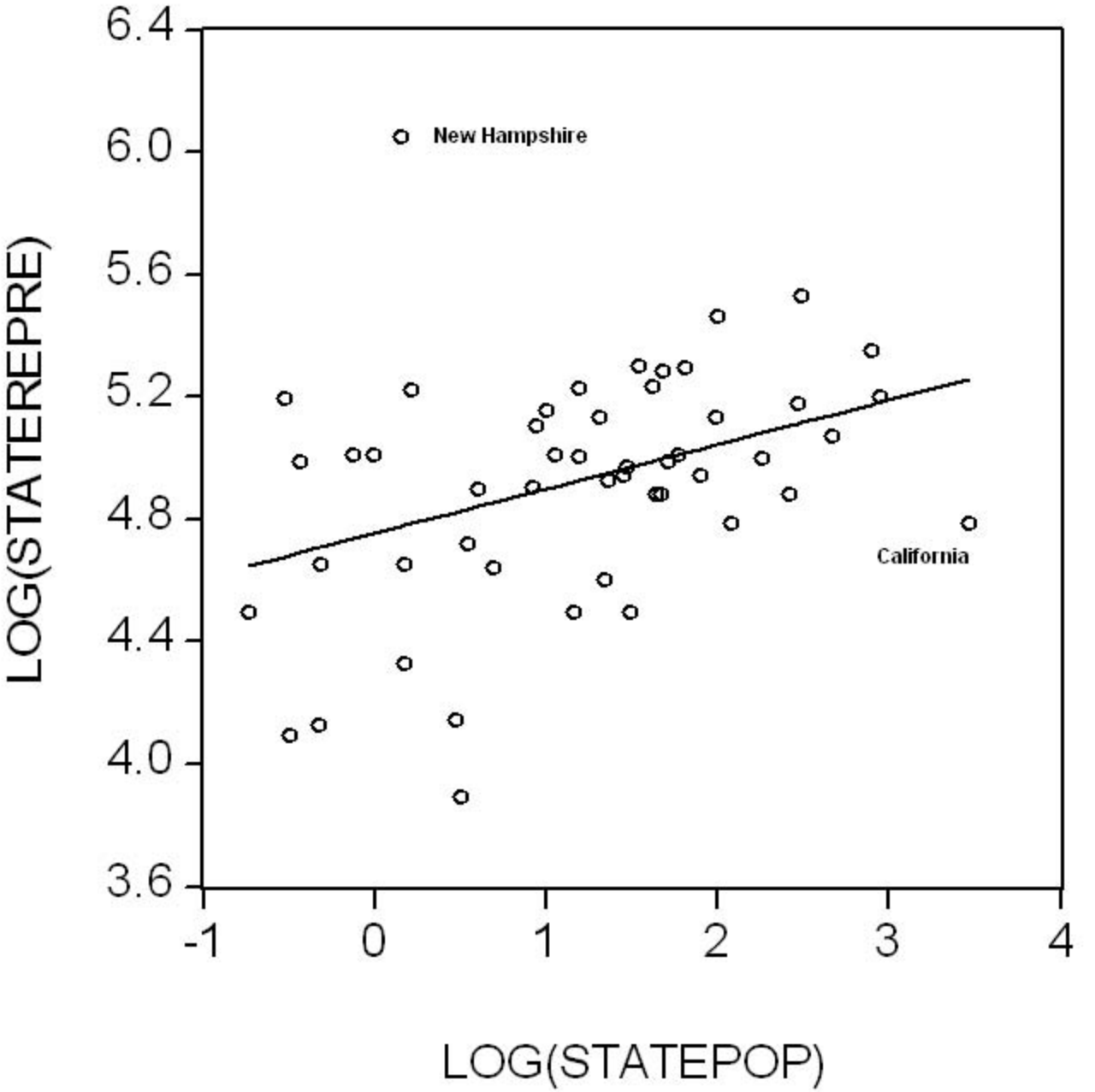
$$\log(n) = \underset{(52.35)}{4.696} + \underset{(3.32)}{0.172}\log(N).$$

Adjusted R^2 is equal to .21, and the global $F = 14.16$, with exactly 49 observations (t-statistics in parentheses).

$$\log(n) = \underset{(41.13)}{4.723} + \underset{(3.09)}{0.218}\log(N) + \underset{(2.47)}{5.11}(10^{-4})Density - \underset{(-2.03)}{0.218}(10^{-6})Salary.$$

Adjusted $R^2 = .35$, the global $F = 4.78$ (significant at 3%)

Figure 3


6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
-1
0
1
2
3
4
LOG(STATEREPRE)
LOG(STATEPOP)
New Hampshire
California

CONCLUSION

♣ A public decision-making problem under decentralized knowledge and costly representation $\Rightarrow$ sampling Groves mechanisms;

♣ Robustness to opportunistic manipulation by the Principal forbids the use of a priori knowledge in the public decision procedures.

♣ Theory of the determinants of the size of representative samples: trade-off between the unit cost of sampling and estimation error of unknown parameter μ

♣ Why elections instead of random sampling?

♣ The relationships of the sampling model with models in which representatives are chosen through elections should be explored.