European Economic Review 46 (2002) 908–918
EUROPEAN ECONOMIC REVIEW
www.elsevier.com/locate/econbase

Do constitutions cause large governments? Quasi-experimental evidence

Torsten Persson[a,b,c,d,*], Guido Tabellini[c,e,f]

[a] *Institute for International Economic Studies, Stockholm University, S-10691 Stockholm, Sweden*
[b] *LSE, London, UK*
[c] *Centre for Economic Policy Research, London, UK*
[d] *National Bureau of Economic Research, Cambridge, MA, USA*
[e] *IGIER, Bocconi University, Milan, Italy*
[f] *CES-IFO, Munich, Germany*

Abstract

How do constitutional rules for elections and legislation affect the size of government? We ask this question in a new sample of about 80 countries in the 1990s. In addition to conventional regression methods, we use quasi-experimental, matching methods, which more convincingly address legitimate criticisms of causal inference from cross-country data. Both sets of estimates suggest that presidential regimes and majoritarian elections produce smaller governments.
© 2002 Elsevier Science B.V. All rights reserved.

JEL classification: D7; H1

Keywords: Fiscal policy; Size of government; Electoral rules; Goverment regimes; Matching methods

1. Introduction

How do the constitutional rules for elections and legislation shape the size of government? A recent literature has studied this question. In particular, Persson et al. (2000a) predict that presidential regimes lead to smaller governments, compared with parliamentary regimes. A larger literature has studied the effect of majoritarian vs. proportional elections on economic policy, but obtained less specific predictions for the size of government. Spurred by these theoretical results, a few empirical papers have explored cross-country and panel data. Countries ruled by presidential regimes indeed seem to have much smaller government spending, by 5–10% of GDP,

* Corresponding author. Tel.: +46-8-163066; fax: +46-8-164177.
E-mail addresses: torsten.persson@iies.su.se (T. Persson), guido.tabellini@uni.-bocconi.it (G. Tabellini).

0014-2921/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0014-2921(01)00224-0

depending on the sample (Persson and Tabellini, 1999, 2001). The correlation of the electoral rule with the size of government is less clear cut, though some papers have found an association of majoritarian elections with smaller governments (Persson and Tabellini, 1999, 2001; Milesi-Ferretti et al., 2000).[1]

But can these associations be interpreted as *causal*? For example, can we confidently infer that Italy's reform of its electoral rule from proportional towards majoritarian in the mid 1990s will cause a smaller size of government, and that the opposite will happen in New Zealand or Japan, where electoral reform went the other way? Among democracies, deep constitutional reforms are rare. Hence, reverse causation (from policies to institutions) is unlikely, but we mainly have to base our inference on cross-country variation. This raises the problem that institutions – and not just policies – are endogenous. Even a cursory look at the data reveals systematic patterns: Presidential countries are concentrated in the Americas, almost all former British colonies have majoritarian elections, most of continental Europe is parliamentary with proportional elections. Countries have thus not selected their constitutions randomly, but on the basis of historical, cultural or geographic determinants.[2] How do we know that these constitutional determinants are not also the ultimate cause of the observed size of government?

The goal of this paper is to address this inference problem by applying matching methods based on the propensity score. These quasi-experimental methods were developed long ago for application in medical sciences, and have recently become popular in labor economics. They lend themselves naturally to comparative politics, as they focus precisely on non-random selection.[3]

To see the nature of the problem, suppose we are interested in how constitutional reform, from rule $S=0$ to 1, say from proportional to majoritarian elections, affects the size of government, Y, conditional on a vector of exogenous variables, $\mathbf{X}$, say the country's socio-economic and historical characteristics. Causality is naturally defined by the *average* effect of reform on the outcome Y conditional on $\mathbf{X}$, namely $\mathsf{E}(Y^1 - Y^0 \mid \mathbf{X})$, where superscripts denote the constitutional state ($S=0,1$) and the E operator refers to expectations in the overall population of countries, conditional upon $\mathbf{X}$. Let P (possibly also a function of $\mathbf{X}$) denote the probability of observing the constitutional state corresponding to $S=1$ in a country drawn at random. Then we can write:

$$\begin{aligned}\alpha \equiv \mathsf{E}(Y^1 - Y^0 \mid \mathbf{X}) &= P \cdot [\mathrm{E}(Y^1|\mathbf{X}, S=1) - \mathrm{E}(Y^0 \mid \mathbf{X}, S=1)] \\ &\quad + (1-P)\cdot[\mathrm{E}(Y^1 \mid \mathbf{X}, S=0) - \mathrm{E}(Y^0 \mid \mathbf{X}, S=0)], \qquad (1)\end{aligned}$$

where $\mathrm{E}(\cdot \mid \mathbf{X}, S=I)$ is the expectations operator conditional on $\mathbf{X}$ and the state $S=I$. In the specific example, α is thus the weighted average of the effect of electoral reform

---

[1] This theoretical and empirical literature also deals with other aspects of economic policy, such as the composition of spending and corruption by elected officials.

[2] See Colomer (2001) for a very useful account of constitutional origins in existing democracies.

[3] Persson et al. (2000b) and Persson (2001) have applied this methodology to study the effect of electoral rules on corruption, and the effect of common currencies on bilateral trade, respectively. Persson and Tabellini (2002) use this methodology more extensively on the same general topic as this paper. King and Zeng (2001) discuss similar selection and inference problems, aiming at a political-science audience.

in the two groups of countries, those currently under majoritarian rule (the first square bracket) and those currently under proportional rule (the second square bracket), each weighted by its relative frequency. Note that these two effects are *not* necessarily symmetric if the selection of the constitution is not random and related to the outcome.[4]

How can we estimate α? In our data, we only observe $Y_i = S_i Y_i^1 + (1 - S_i) Y_i^0$ in country i. Hence, we can easily form unbiased estimates of P, $\mathrm{E}(Y^1 \mid \mathbf{X}, S = 1)$ and $\mathrm{E}(Y^0 \mid \mathbf{X}, S = 0)$. But the other terms, $\mathrm{E}(Y^1 \mid \mathbf{X}, S = 0)$ and $\mathrm{E}(Y^0 \mid \mathbf{X}, S = 1)$, are *unobservable counterfactuals*: Any country can only have one electoral rule at any given point in time. If the world had many countries and if constitutional choices were completely random (or made by controlled experiment), this would not be a problem as the observed distribution of Y_i in each group of countries would give an unbiased estimate of the counterfactual. Formally, we could safely assume $\mathrm{E}(Y^J \mid \mathbf{X}, S = I) = \mathrm{E}(Y^I \mid \mathbf{X}, S = I)$, $I \neq J$, and write $\alpha = \mathrm{E}(Y^1 \mid \mathbf{X}, S = 1) - \mathrm{E}(Y^0 \mid \mathbf{X}, S = 0)$. Such is not the world, however; countries do not select their constitutions at random. If the variables shaping constitutional choice also influence the size of government, we have to take this into account to avoid biasing the estimate of α. Alternative methodologies implicitly or explicitly replace the unobservable counterfactual in different ways, and are more or less robust to this prospective bias.

Section 2 describes the data and the sample of countries; it is a larger sample than those used in the existing literature, consisting of about 85 countries in the 1990s. Section 3 estimates the constitutional effect on the size of government by conventional regression methods, discussing the underlying identifying assumptions. Section 4 estimates the causal effect by different matching estimators based on the propensity score. Section 5 concludes.

2. Data

We could gather data for 86 democracies in the 1990s. A democracy is defined as having a Gastil score of political and civic liberties of less than 5 throughout the 1990s.[5] Two dummy variables, *MAJ* and *PRES*, classify electoral rules and regime types. Majoritarian countries ($MAJ = 1$) rely exclusively on plurality rule in electing the lower house – there are 35 such countries. The remaining 51 countries (mixed and strictly proportional) are lumped together with $MAJ = 0$. The 33 countries where the executive is not accountable to the legislature through a vote of confidence are coded as presidential ($PRES = 1$), the 53 where it is as parliamentary ($PRES = 0$).[6]

[4] Our measure α is a version of the *average treatment effect* used in the evaluation literature. It is a weighted average of the *average effect of treatment of the treated* and the *average effect of non-treatment on the non-treated*. Heckman et al. (1999) discuss these and other causal effects.

[5] Gastil scores, complied by Freedom House, range from 1 to 7, with lower values denoting better democracies. While many countries are true democracies (a score of 1 or 2), our threshold of 5 is quite high. The sample thus includes some dubious cases. Several former socialist countries are included.

[6] A few countries changed their constitution during the 1990s, so that the average value of *MAJ* and *PRES* is between 0 and 1. We have redefined *MAJ* and *PRES* as equal to 0 or 1, depending on whether its average is above or below 0.5. Persson and Tabellini (2001, 2002) discuss these measures and their relation to the underlying theory in more detail.

The size of government is measured by the ratio of central government spending (inclusive of social security) to GDP, expressed as a percentage (*CGEXP*). Data for this variable exist for 81 countries, often for the entire 1990–1998 period; we just compute the average over any available years. Finally, we include a number of conditioning variables likely to influence the size of government and/or constitutional choice. They are the log of real per capita income (*LYP*); openness (*TRADE*), defined as exports plus imports over GDP; the percentages of the population between 15 and 64 (*PROP1564*) and above 65 years (*PROP65*); the Gastil index of political and civic liberties (*GASTIL*); indicator variables distinguishing federal states (*FEDERAL*) and OECD countries (*OECD*); indicator variables measuring geographic location and colonial origin (discounted to the present from the year when the colony first became independent). [7]

3. Linear regression estimates

Let the size of government in country i and constitutional state S be: $Y_i^S = F^S(\mathbf{X}_i, \varepsilon_i^S)$, $S=0,1$, where ε^S is an unobserved random variable (i.e., each country has two potential constitutional states). The constitutional state is determined by: $S_i = 1$ as $G(\mathbf{Z}_i, \eta_i) \geqslant 0$, and $S_i = 0$, otherwise, where $\mathbf{Z}$ is a set of observable constitutional determinants, possibly overlapping or coinciding with $\mathbf{X}$, and η an unobserved random variable. Linear regressions give an unbiased estimate of the parameter α in (1) under two assumptions.

A first assumption concerns functional form, namely $F^1(\mathbf{X}_i, \varepsilon_i^1)$ and $F^0(\mathbf{X}_i, \varepsilon_i^0)$ are *linear* and differ only by an intercept term. Specifically, $F^1(\mathbf{X}_i, \varepsilon_i^1) = \alpha^1 + \beta\mathbf{X}_i + \varepsilon_i^1$, while $F^0(\mathbf{X}_i, \varepsilon_i^0) = \alpha^0 + \beta\mathbf{X}_i + \varepsilon_i^0$, where the vector β of coefficients is the same in the two expressions. Moreover, the distribution of the error term ε^S is the same irrespective of S. A second assumption is *conditional independence* (or "selection on observables", or "no omitted variables", or "recursivity"): The error terms ε and η are orthogonal. Under these two assumptions, we can express the causal effect in (1) as $\alpha = \mathsf{E}(Y^1 - Y^0 \mid \mathbf{X}) = \mathrm{E}(Y^1 \mid \mathbf{X}) - \mathrm{E}(Y^0 \mid \mathbf{X})$. As $\alpha = \alpha^1 - \alpha^0$, we can estimate it by the coefficient on S in a linear regression of Y on $\mathbf{X}$ and S. [8] Conditional independence and linearity thus allow us to replace the unobservable counterfactuals in (1) by "holding constant" the X variables in an OLS regression.

We now apply this methodology, evaluating the coefficients on *MAJ* and *PRES*. We enter these constitutional variables one by one, as well as together. In the latter case, our coding convention makes the coefficient on *PRES* measure the difference between presidential – proportional countries and the default group of parliamentary – proportional countries ($MAJ = PRES = 0$), while the coefficient on *MAJ* measures differences

[7] All variables and their sources are described in Persson and Tabellini (2001, 2002). A description plus all the data used in the article are available at: http://www.iies.su.se/data/home/perssont/data.htm

[8] To interpret the estimated coefficient as a measure of the true causal effect of S on Y, we also need to assume that a change in S does not alter the covariates $\mathbf{X}$. As our main concern here is selection, we will make this assumption in the following. See Heckman et al. (1999) and King and Zeng (2001) for more discussion.

Table 1
OLS estimates

Majoritarian		−2.14	**−4.24**		
		(1.85)	**(2.14)**		
Presidential	**−6.50**	**−7.28**	**−5.04**		
	(2.07)	**(2.18)**	**(2.47)**		
Majoritarian – parliamentary				**−4.56**	−4.17
				(2.57)	(2.81)
Proportional – parliamentary				**−9.70**	−4.96
				(2.81)	(3.54)
Majoritarian – presidential				**−9.34**	**−9.28**
				(3.24)	**(3.42)**
Colonial origin	NO	NO	YES	NO	YES
Geographic	NO	NO	YES	NO	YES
# Obs.	81	81	81	81	81
Adj. R^2	0.58	0.59	0.64	0.59	0.64

Dependent variable is CGEXP. Standard errors in brackets. Boldface fonts denote significance at the 10% level. Controls LYP, TRADE, PROP1564, PROP65, GASTIL, FEDERAL, OECD always included (see Section 2 of the text for variable names). Geographic dummy variables (for Latin America, Asia, Africa, Europe–Middle East) and colonial origin dummy variables (UK, Spanish, French or Other) included as indicated. All colonial origin dummy variables multiplied by (1-TINDEP/250), where TINDEP denotes years since date of independence.

between majoritarian – parliamentary and the default group. As a further check, we replace *PRES* and *MAJ* with three indicator variables, partitioning the countries more finely into *MAJPRES*, *MAJPAR* and *PROPRES* (obviously defined) relative to the same proportional – parliamentary default group.

Table 1 reports different specifications. The most parsimonious one follows the existing literature. We then add indicator variables for a country's location and/or colonial origin. In this larger and more recent sample, we confirm most earlier empirical results. As expected from theory, *PRES* has a negative coefficient, ranging from −7% to −5% of GDP. This estimate is significantly different from zero, whatever is done to the specification. The coefficient on *MAJ* is also negative, though smaller and statistically significant only when continental and colonial dummies are included. [9] Results based on the finer partition of the constitution, displayed in the lower part of the table, suggest that the effects of *MAJ* and *PRES* are indeed additive. As the four groups are quite small and group membership is correlated with the colonial and continental dummies, however, standard errors are higher.

How convincing are the independence and linearity assumptions? A central concern is selection on unobservables; i.e., some omitted variables drive constitutional forms as well as outcomes. Classic methods of dealing with this problem rely on finding sources

[9] Milesi-Ferretti et al. (2000) obtain this result in OECD data on about 20 countries from 1960 to 1995, whereas Persson and Tabellini (2001) obtain weaker results in a broader data set of about 60 countries from 1960 to 1998. These papers conduct sensitivity analyses of similar linear regressions in various dimensions.

of exogenous variation in constitutional rules (through time variation or instrumental variables). Unfortunately, we lack time variation and have not found any instruments for constitutional choices that could be claimed not to influence the size of government. Thus, some conditional-independence assumption is necessary for inference. [10]

What about linearity? It is easy to think about plausible non-linearities in the relations underlying our empirical work. In particular, the constitutional effect may interact with some of the controls. For instance, how the electoral rule shapes the welfare state and hence the size of government may co-vary with the age composition of the population; the effect of presidentialism may be different in Latin America and Europe; or the effect of formal constitutional rules may be less important in less developed democracies and economies. If the distribution of the **X** variables is very similar among countries in different constitutional states, such non-linearities may be neglected. But non-random selection on observables implies that the distributions of the observable attributes could be very different across constitutional groups. Rather than a convenient local approximation, the linearity assumption may then be a source of considerable bias. The bias may arise in two ways: groups may have non-overlapping distributions of **X**, so that we compare incomparable observations, or different densities on the overlapping part of the distribution, so that we weigh the observations incorrectly (see e.g., Heckman et al., 1999 for a precise decomposition of the prospective bias).

As argued above, non-random selection is an important feature of our data. To make the point formally, consider the 14 conditioning variables (including geographical and continental dummies) used in our regression analysis. Testing equality of means across different regimes (*PRES* equal to 1 or 0) we reject (at the 5% level) in 9 cases; across different electoral rules (*MAJ* 1 or 0) we reject in 7 cases. Differences across regimes are more pronounced; not only do we reject more often but also more decisively. Presidential countries are poorer, worse democracies, more often located in Latin America (and thus of Spanish–Portuguese colonial origin), and have younger populations. Countries with majoritarian elections are more often former UK colonies and have younger populations

In sum, the prospective bias from non-random selection and non-linearity seems a problem worth taking seriously. The matching methods considered next relax the functional-form assumption and directly address distributional differences across country groups.

4. Matching estimates

The central idea in matching is to approach the evaluation of causal effects as in a controlled experiment. Data are split into one group of "treated" observations (corresponding say to $S=1$) and another group of "non-treated" or "control" observations

[10] Another possibility would be to include a Heckman-style correction for selection on unobservables in the outcome regression. Absent valid instruments for institutions, however, identification in the correction procedure becomes fragile, hinging entirely on the functional-form assumptions on the joint distribution for η and ε. Under the assumption of bivariate normality, the coefficients displayed in Table 1 become more strongly negative, while maintaining statistical significance.

($S = 0$). The unobservable counterfactual outcome for a specific treated observation is then estimated from the outcome among controls with similar observable attributes. When we compare similar countries, the selection into different constitutions is largely random, as in an experiment. In fact, the estimate can be made non-parametrically, i.e., without any assumption on the functional form of the constitutional effect. Successful matching thus removes the bias due to systematic selection and interaction terms or other potential non-linearities. [11]

This methodology has a difficulty, however, which is easy to see in our application. We have already stressed that countries differ in many attributes that may correlate with observed policy outcomes as well as observed constitutional states; i.e., the dimension of $\mathbf{X}$ is high. Comparing similar countries under different constitutional rules would therefore rapidly exhaust available data. But an important result due to Rosenbaum and Rubin (1983) provides a way out. It implies that matching countries with the same *probability of selecting* a specific constitutional rule, given the relevant controls $\mathbf{X}$, is equivalent to matching directly on $\mathbf{X}$. This probability is called the *propensity score*.

Formally, let $p_i = p(\mathbf{X}_i)$ be the propensity score that country i is selected into state 1 (rather than 0) and assume that the so-called *common support condition* $0 < p(\mathbf{X}_i) < 1$ holds for all $\mathbf{X}_i$. Rosenbaum and Rubin (1983) show that conditional independence implies: $\mathrm{E}(Y^0|p, S=1) = \mathrm{E}(Y^0|p, S=0) = \mathrm{E}(Y^0|p)$ and $\mathrm{E}(Y^1|1-p, S=0) = \mathrm{E}(Y^1|1-p, S=1) = \mathrm{E}(Y^1|p)$, where the expectation operator refers to the distribution of Y conditional upon the propensity score p and, where indicated, upon the constitutional state S. In other words, for countries with similar propensity scores, the constitutional state S is uncorrelated with the vector of potential outcomes (Y^1, Y^0), so conditioning on $\mathbf{X}$ or on $p(\mathbf{X})$ is equivalent. Moreover, by the law of iterated expectations, $\mathrm{E}(Y^J|S=I) = \mathsf{E}\{\mathrm{E}(Y^J|p)|S=I)\}$, where the inner expectation E is over the distribution of Y conditional on p, while the outer expectation E is over the distribution of p, conditional on $S = I$. Using these results, we can rewrite (1) as follows:

$$\begin{aligned}\alpha = {} & P \cdot \mathsf{E}\{[\mathrm{E}(Y^1|p) - \mathrm{E}(Y^0|p)]|S=1\} \\ & + (1-P) \cdot \mathsf{E}\{[\mathrm{E}(Y^1|p) - \mathrm{E}(Y^0|p)]|S=0\},\end{aligned} \tag{2}$$

where as before P is the probability of observing state $S = 1$.

The parameter α defined in (2) is directly relevant for estimating the causal effect of presidential regimes (*PRES* equal to 1 not 0), or of majoritarian electoral rules (*MAJ* equal to 1 not 0). An extension to multiple constitutional states is considered below. But estimating α by this formula raises a number of specific issues.

First, we must estimate the probability P. This is easily done by computing the relative frequency of presidential regimes and majoritarian rules, respectively.

Second, we have to estimate the propensity scores $p_i = p(\mathbf{X}_i)$ for both $PRES = 1$ and $MAJ = 1$ for each country. We can do this by a simple logit. But which

[11] Heckman et al. (1999) and Angrist and Kreuger (1999) discuss matching and its relation to other techniques, including linear regression.

Table 2
Matching estimates

Matching method	Nearest neighbor	Stratification	Kernel-based
Majoritarian	**−3.15**	−3.46	**−5.17**
	(3.89)	(4.23)	**(2.96)**
Presidential	**−5.60**	−4.11	**−6.09**
	(2.14)	(3.20)	**(1.96)**
Majoritarian – parliamentary			−3.70
			(2.75)
Proportional – presidential			**−8.94**
			(1.74)
Majoritarian – presidential			**−10.12**
			(2.27)

Dependent variable is CGEXP. Standard errors in brackets computed by bootstrapping (200 repetitions). Boldface fonts denote significance at the 10% level. Estimates in each of the first two rows rely on propensity scores estimated by logit, conditioning on LYP, PROP65, FEDERAL, GASTIL, UK colonial origin and Latin-America dummies. Estimates in each of the last three rows rely on pair-wise propensity scores obtained from unconditional probabilities estimated by a multinomial logit, including the same variables except the colonial origin and continent dummies.

variables should we include in **X**? A first concern is the conditional-independence assumption. To respect it, we should not omit any variable correlated with constitutional choices that might also influence the size of government. [12] A second concern is the common-support condition. If we explain constitutional choice "too well", we shrink the region of overlapping propensity scores in the treatment and control groups, so that matching becomes infeasible. Preserving enough randomness in the propensity scores thus speaks for a parsimonious logit specification. In practice, we resolve this trade-off by including a subset of the control and indicator variables listed in Table 1, namely those with the highest t-values in the regressions and the equal-means tests (see the end of Section 3). Table 2 details the specification.

Third, we have to evaluate the expected outcomes in the actual treated ($S = 1$) and non-treated ($S = 0$) groups, the first and fourth terms in (2), $\mathsf{E}[\mathrm{E}(Y^1|p)|S = 1]$ and $\mathsf{E}[\mathrm{E}(Y^0|p)|S = 0]$. A straightforward, non-parametric estimator is the average outcome in each group.

Fourth, we must select a specific matching method, i.e., a method for evaluating the remaining two counterfactual terms in (2). In the first expression, e.g., which control ($S=0$) countries have a propensity score close enough to the score of a specific treated ($S = 1$) country that we should include them in the matching of that country? As the small sample properties of different matching estimators are unknown, we use three alternative methods, each defining a specific estimator. The *nearest neighbor* method is simple and intuitive. For each treated country, we just find its "closest twin": the

[12] More precisely, our assumption of conditional (mean) independence says $\mathrm{E}(Y^0|p,S=1)=\mathrm{E}(Y^0|p,S=0)$. This allows for omitted variables, provided that they affect potential outcomes in the two states in similar ways.

non-treated country with the closest estimated value of p_i. Control countries are used several times, if they happen to be the closest match for several treated countries (at the price of a higher standard error). In a small sample, however, this estimator can be quite fragile: Small changes in the specification of the propensity score can change the ranking of countries, thereby switching the control observations more heavily used. To cope with this, we rely on two additional methods of matching. Under *stratification*, we rank the treated and non-treated countries according to their estimated propensity scores and group them into three different strata, corresponding to the intervals $(0, 0.33)$, $[0.33, 0.67]$, and $(0.67, 1.0)$. Each treated country is matched with an arithmetic average of all the non-treated countries belonging to the same stratum; thus, each stratum is weighted by the proportion of treated countries it contains. Our third method is *kernel-based*. Here, each treated country is matched with a weighted average of all non-treated countries within a certain propensity-score distance, with weights declining in that distance. Specifically, we use a radius of 0.25.

Finally, we must impose the common-support condition. Thus, the computations described above are only performed for the treated and non-treated countries that share a common support in their estimated propensity scores. Observations outside the common support are discarded as non-comparable in terms of observable attributes.

Matching on the propensity score and imposing the common support does indeed balance the underlying **X** variables considerably across groups, even for those components of **X** not included in the logit specification. Repeating the equal-means tests of Section 3 across the $PRES = 1$ and 0 groups for each of the three strata defined above, we reject equal means in 2, 5 and 1 cases, compared with 9 cases for the full sample. Across the $MAJ = 1$ and 0 groups, we reject equal means in 1, 0 and 0 cases, compared with 7 cases for the full sample. Furthermore, the remaining rejections are weaker (lower t-statistics) than in the full sample.

The resulting matching estimates and their standard errors are displayed in the upper part of Table 2. Clearly, the point estimates accord well with those in Table 1. All three estimators suggest that both presidential regimes and majoritarian elections induce smaller governments, with a slightly larger effect of the former – on the order of 5% rather than 4% of GDP. The similarity with the OLS estimates suggests that the linearity assumption is appropriate. [13]

Finally, what are the causal effects of the finer partitioning into four constitutional groups also considered in Section 3? Formally, this is analogous to the case of multiple treatments in the evaluation literature. Following Lechner (2000), we can generalize the above methodology with a slight reinterpretation. If we rewrite (2), everywhere replacing state 1 by state $S = 1, 2, 3$, the resulting expression defines an alternative average treatment effect: The expected effect of state S (vs. state 0) on Y for a country drawn at random among those currently in either 0 or S. Accordingly, the propensity

[13] Note that matching estimates are based on a smaller set of countries than OLS, because of the common support condition (31 countries when estimating the effect of *MAJ* and 63 for *PRES*, compared with 81 in Table 1). Given the small sample, standard errors have been computed by bootstrapping. They are generally larger than the OLS standard errors. This is not surprising, as the whole point of non-parametric matching is to reduce prospective bias at the cost of efficiency.

scores $p_i^S = p^S(\mathbf{X}_i)$ (indexed by S as there are more than two states) denote the probability of state S, conditional on being either in state 0 or S. We estimate these scores by running a multinomial logit on $\mathbf{X}$, obtaining unconditional probabilities q_i^S, and then computing the conditional probabilities $p_i^S = q_i^S/(q_i^S + q_i^0)$.

For comparability with Section 3, state 0 is the proportional – parliamentary group. Given that we estimate 3 parameters instead of 1 for each covariate in $\mathbf{X}$, we must choose an even more parsimonious specification of our multinomial logit, dropping continental and colonial dummies. Some of the four groups are very small, so that the nearest-neighbor method is quite sensitive to the specification. The stratification method also becomes less meaningful: Imposing the common-support condition leaves us with few usable observations in the strata of certain groups. Therefore, we only rely on the kernel-based method, widening the radius to 0.4 (from 0.25). Results appear in the lower part of Table 2. The point estimates for the size of government, are again in the same range as the regression estimates in Table 1.

5. Conclusion

Empirical work in comparative politics generally exploits cross-country variation: Time-series variation is often not available because deep constitutional reforms are so rare. But cross-country analysis is often criticized as unreliable. While some doubts concern omitted variables, others concern systematic selection ("isn't what you call S just Latin America?") or non-linearity ("don't you think that the effects of S vary with culture and history?"). Quasi-experimental methods are therefore a useful complement to conventional methods in economics and political science, because they directly focus attention on whether one is really trying to compare the incomparable.

Linear regressions and non-parametric matching are complementary statistical tools in comparative politics in the following sense. Systematic selection into a constitutional form is a strong feature of the data. Linear regressions enable us to control for many country attributes, and hence lend some credibility to the assumption of selection on observables. But strong selection on observables might seriously bias our estimates of causal effects, if linearity does not hold. Matching instead relaxes linearity by focusing on local comparisons. But in a small sample such as ours, we have to estimate the propensity score conditioning on fewer variables, thus straining the selection on observables assumption.

In this paper, our quasi-experimental, matching estimates do not differ much from our conventional regression estimates. This makes us more confident in a true causal interpretation of our findings that presidential regimes and majoritarian elections produce smaller governments.

Acknowledgements

We are grateful to the participants in seminars at the IIES and the LSE for comments, Alessandra Bonfiglioli and Davide Sala for research assistance, Christina Lönnblad for editorial assistance, and Barbara Sianesi for sharing her program codes with us. The research was supported by the European Commission (a TMR grant), MURST, Bocconi University and the Swedish Council for Research in the Humanities and Social Sciences.

References

Angrist, J., Kreuger, A., 1999. Empirical strategies in labor economics. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3c. North-Holland, Amsterdam.

Colomer, J., 2001. Political Institutions, Democracy and Social Choice. Oxford University Press, Oxford.

Heckman, J., Lalonde, R., Smith, J., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3c. North-Holland, Amsterdam.

King, G., Zeng, L., 2001. How factual is your counterfactual? Mimeo., Harvard University.

Lechner, M., 2000. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. Mimeo., University of St. Gallen.

Milesi-Ferretti, G-M., Perotti, R., Rostagno, M., 2000. Electoral systems and the composition of public spending. Mimeo., Columbia University.

Persson, T., 2001. Currency unions and trade: How large is the treatment effect? Economic Policy 33, 435–448.

Persson, T., Tabellini, G., 1999. The size and scope of government: Comparative politics with rational politicians, 1998 Alfred Marshall Lecture. European Economic Review 43, 699–735.

Persson, T., Tabellini, G., 2001. Political institutions and fiscal policy: What are the stylized facts? Mimeo., IGIER.

Persson, T., Tabellini, G., 2002. Economic Policy in Representative Democracies. Munich Lectures in Economics, MIT Press, Cambridge, MA, forthcoming.

Persson, T., Roland, G., Tabellini, G., 2000a. Comparative politics and public finance. Journal of Political Economy 108, 1121–1141.

Persson, T., Tabellini, G., Trebbi, F., 2000b. Electoral rules and corruption. NBER working paper no. 8154.

Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.