Is the Science of Comparative Politics Possible?*

Adam Przeworski
Department of Politics
New York University

August 4, 2005

Abstract

Many research problems in comparative politics involve assessing the impact of institutions, policies, or events on some performance, outcome, or result. While such evaluations are relatively unproblematic when data can be generated by the researcher, they are subject to several biases when data are produced by history. The chapter is an overview of issues entailed in causal inference and an introduction to alternative research strategies.

Prepared for publication in Carles Boix and Susan C. Stokes (eds.), Oxford Handbook of Comparative Politics.

*With apologies to MacIntyre (1972) for stealing the title. I appreciate comments by Neal Beck, Gösta Esping-Anderson, and the editors of this volume..

1 Introduction

The chapter is an overview of issues entailed in making causal inferences when the data are generated by processes that are not under the control of the researcher. As all overviews, this one is just an introduction to issues that have been studied in greater depth by others.

Many research questions in comparative politics concern the impact of some institution, policy, or event on some outcome, result, or performance. I will generically refer to the former as "the (potential) cause" and to the latter as "the effect." Examples include:

(1) The impact of political institutions on economic development.

(2) The impact of political regimes on the initiation of wars.

(3) The impact of electoral systems on the number of parties.

(4) The impact of trade strategies on economic performance.

(5) The impact of signing particular international treaties on some performance, say of signing the Kyoto protocol on carbon emissions.

(6) The impact of revolutions on subsequent social change.

(7) The impact of peace-keeping missions on peace.

The list is endless: I just wanted to emphasize that the causes may include institutions, policies, and events. Moreover, problems of this form are not limited to "comparative politics" or even to the cross-national level. For example, in the aftermath of the 2004 elections, some people noted that the Bush vote was higher than expected in those Florida counties that used electronic machines while it was about what one would expect in those countries that used traditional punch card machines. Did the kind of voting machine affect the outcome? This is again a question about the effect of a cause.

These are then the kinds of questions that will be considered. We will find, however, that at least in one view such questions cannot be answered without inquiring where the causes come from. To identify the impact of political regimes on growth, we need to learn how political regimes come about and die. To identify the impact of peace-keeping missions on peace, we need to know when are such missions undertaken. To identify the impact of voting machines on the Bush vote, we need to know under what conditions different voting systems were in place. Hence, we need to study *causes of effects as well as effects of causes*. And this means that we will face almost all problems generic to comparative politics.

2 The Problem

Suppose we have a data matrix that looks like this.[1] T stands for the (potential) cause, where $T = 1$ indicates "treatment" and $T = 0$ "control" (or a different treatment).[2] X and V are "covariates," that is, traits of an individual unit prior to the application of the treatment. X is the vector of covariates observed by the researcher, V are covariates not observed. N.A. stands for "not available." $Y = \{Y_0, Y_1\}$ is the variable subject to the potential effect of the cause, where Y_0 stands for states of the units not exposed to treatment and Y_1 of those exposed to treatment, so that for each unit i we observe either Y_1 or Y_0 :

$$Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}. \tag{1}$$

A "unit" is an opportunity for the cause to operate. It may be an individual, a country, or what not. Moreover, it may be the same individual or a country in a different state: say Sweden in 1950 and in 1951. Hence, the "unit" is a full set of observable and unobservable covariates: i is coextensive with the vector of "background conditions" $(\mathbf{x}_i, \mathbf{v}_i)$.

*** Table 1 here ***

Now, let U stand for the effect of V on Y and assume linear separability. Then

$$E(Y|X, V) = E(Y|X) + U. \tag{2}$$

Substituting into (1) (and dropping the i subscript) yields

$$\begin{aligned} Y &= E(Y_0|X) + T[E(Y_1 - Y_0|X)] + \{T(U_1 - U_0) + U_0\} = \\ &\quad \beta_0(X) + \beta(X)T + U, \end{aligned} \tag{3}$$

[1] A comment on notation. As conventional in this literature, capital letters denote variables; small letters particular values. Bold letters represent vectors. $E()$ is to be read as "the expected value," $Y|X$ as "the value of Y conditional on the value of X," so that $E(Y|X)$ is "the expected value of Y given X." To simplify the notation, I implicitly use throughout the law of iterated expectations.

[2] Although for simplicity I assume that the cause is a binary variable, everything said here holds for any discrete or continuous values of T.

where $\beta(X) = E(Y_1 - Y_0|X)$ is the average causal effect, discussed further below, and $U = T(U_1 - U_0) + U_0$.

I introduce all this notation abruptly just to show the basic concern in identifying causal effects, namely, whether $E(U) = 0$. Whether it does equal 0 is unknowable in general, but there are different identifying assumptions that imply this property. These assumptions, in turn, are not testable but we can intelligently argue whether they are reasonable. I will let the reader decide whether what we practice is science or art.

3 Kinds of Data

3.1 Experimental Studies

To fix ideas, assume that we do know that $E(U) = 0$. This will be true if the treatment is randomly assigned to units.

The most important aspect of random assignment is that it matches on unobserved as well as on observed covariates. Note that in Table 1 the observed covariates are exactly "*matched*," meaning that for each vector $\mathbf{x}$ there is an equal number of observations with $T = 0$ and $T = 1$, so that the means of each X are identical for the treatment and control groups. Most of what will be said below, however, also holds when the observations are sufficiently well "*balanced*," meaning that the means of each X are sufficiently similar for units with $T = 0$ and $T = 1$. Under random assignment the unobserved covariates should be balanced as well. And since random assignment means that the value of the causal variable is independent of all characteristics of a unit, T is independent of U or $E(U) = 0$.

Now, what is the causal effect of treatment on the particular "unit" i, the Individual Treatment Effect? This effect is *defined* as the difference between the states of an individual unit when it is subjected and not subjected to the operation of the cause, say the intensity of Joe Smith's headache after he did and did not take aspirin or the extent of social change in France as a result of the revolution of 1789 and without it. Formally,

$$ITE_i = y_{1i} - y_{0i} \equiv \beta_i \tag{4}$$

But for all odd numbered units in Table 1 we observe only their state under control and for all even numbered only under treatment. Hence, even

under random assignment, this question cannot be answered without making some assumptions about hypothetical situations that would have occurred had an individual who did not get treatment (had not been exposed to the potential causes) received it or had an individual who did receive treatment not received it. Since these states did not occur, they are contrary to fact, *counterfactual*.[3] And since counterfactuals cannot be observed, assumptions about counterfactuals cannot be directly tested.[4] Hence, we arrive at the first conclusion.

Conclusion 1: *The effect of a cause on an individual unit cannot be determined without making assumptions about counterfactuals. These assumptions cannot be tested.*

What assumption would identify the individual treatment effect under random assignment?

Assumption 1: *Unit homogeneity (Holland 1986).*

For any $i, j \in N$,

$$if\ (\mathbf{x}_i, \mathbf{v}_i) = (\mathbf{x}_j, \mathbf{v}_j), then\ y_{0i} = y_{0j}\ and\ y_{1_i} = y_{1j}.$$

This assumption says that if any two individuals have the same values of covariates, they would have the same states under control and the same states under treatment. When this assumption is true, the process of selection can be ignored: it does not matter which of two identical units is subject to treatment and which serves as control.

This assumption *identifies* the causal effect of treatment. Applying the homogeneity assumption to the $(i+1)st$ unit in Table 1 yields

$$ITE_{i+1} = y_{1,i+1} - y_{0,i+1} = y_{1,i+1} - y_{0,i},$$

where now both $y_{1,i+1}$ and $y_{0,i}$ are observed.

[3] The idea of counterfactuals goes back to Pascal (1669, sec. 162): "Le nez de Cléopâtre: s'il eût été plus court, toute la face de la terre aurait changé." On the distinctions among different types of conditional propositions, see Edgington (2001). On the logical problems with counterfactuals, see Quine (1953), Lewis (1973), Mackie (2002 [1973]), Goodman (1979), and Stalnaker (1987).

[4] For a statistical view of causality without counterfactuals, see Dawid (2000), who rejects them as metaphysical.

What does "*identify*" mean? While econometrics textbooks use this term technically in many contexts, intuitively "to identify" is to be able to infer relations among variables (or the parameters of multivariate distribution) on the basis of all the possible observations (Koopmans 1949; in Manski 1995: 6). But very often this is possible only by assuming something that may or may not be testable. As Manski (1995: 18) observed, "Theories are testable where they are least needed, and are not testable where they are most needed. Theories are least needed to determine conditional distributions $P(y|x)$ on the support of $P(x)$. They are most needed to determine these distributions off the support." We have seen that since each unit can be observed only in one state at one time it is not possible to identify the individual causal effect without making some assumptions. Hence, we need *identifying assumptions*, such as unit homogeneity. This assumption is not testable. But it seems reasonable.

Now we can ask about the Average Treatment Effect, ATE. Specifically, under what assumptions

$$\beta_{ATE} = E(Y_1 - Y_0|X) = E(\beta|X) = \bar{y}_1 - \bar{y}_0 = \bar{\beta},$$

so that the observed mean difference identifies the *average* treatment effect? The answer is "conditional mean independence":

Assumption 2: *Conditional Mean Independence.*

$$E(Y_1|X, T = 1) = E(Y_1|X, T = 0) = E(Y_1|X)$$

$$E(Y_0|X, T = 0) = E(Y_0|X, T = 1) = E(Y_0|X)$$

This assumption says that conditional on observed covariates we can expect the units not exposed to treatment to react to it identically to those observed under treatment and the units exposed to treatment not to differ in their control state from those observed under control.[5] Under random as-

[5] Again, to help with the notation, $E(Y_1|T = 1)$ is to be read as "the expected value of the outcome under treatment, given that the units have been observed as treated," while $E(Y_0|T = 0)$ as "the expected value of the outcome under treatment, given that the units have been observed as not treated."

signment this assumption is trivially true. And it implies that the observed difference identifies the average causal effect:[6]

$$\bar{\beta} = E(Y_1|X, T = 1) - E(Y_0|X, T = 0) = E(Y_1 - Y_0|X).$$

Conclusion 2: *If the assignment to treatment is random, then the difference of the observed means identifies the average causal effect of treatment.*

3.2 Observational vs Historical Studies

Suppose now that the data in Table 1 were generated by some process unknown to the researcher. Note that "unknown" does not preclude random assignment: even if the researcher did not randomly assign treatments, history may have.[7] If it is possible to make a plausible case that assignment by history was in fact random, we have a "natural experiment" and everything established above applies, specifically the observed difference identifies the average causal effect.[8] But suppose that a claim of randomization cannot be supported. History generated observations by some process and all

[6] According to a theorem by Rosenbaum and Rubin (1983), if the conditional mean independence holds in the form specified in the text, then it also holds in the form in which $p(X) = \Pr(T = 1|X)$ is substituted for X, where $p(X)$ is the "propensity score."

[7] The distinction between randomization by the researcher and by nature goes back to Haavelmo (1944; cited in Angrist and Krueger 2001: 80), who drew an analogy between the experiments "we should like to make" and "the stream of experiments that nature is steadily turning out from her own enormous laboratory, and which we merely watch as possive observers."

[8] To see what is entailed, consider a beautiful study by Banerjee and Iyer (2002). When the British were conquering India, they implanted different tributary systems in different areas: during one period they delegated tax collection to landlords, during another they either charged tax collection to the village as a community or collected taxes themselves from individual peasants. Since these tributary systems depended on the date of conquest, rather than on the characteristics of particular districts, these institutions were exogenous with regard to local climate, endowments, and presumably the unobserved characteristics of the districts. The identification strategy adopted by Banerjee and Iyer was to construct a restricted sample of districts that are geographical neighbors, but which happened to have different tax systems. They observed that "Our strategy might give biased results if the British decision of which land tenure system to adopt depended on other characteristics of the area in systematic ways." (p.10-11). But using this strategy allowed them to assume that "there is no reason to think that the choice of land tenure system at the district level was closely tied to the characteristics of the district.... It is therefore probably reasonable to assume that when two districts lying directly across from each other on either side

the researcher did was to rearrange the data into the form of Table 1, perhaps dropping observations without an exact or a close match (or including only the common support). Some people would want to qualify such data structures as "quasi-experiments," which "lack random assignment of units to conditions but otherwise have similar purposes and structural attributes to randomized experiments" (Shadish, Cook, and Campbell 2002: 104). Yet quasi-experiments are not natural experiments. *Even if units can be perfectly matched on the observed covariates, there is nothing to guarantee that they are also randomized with regard to unobserved ones.*

In most studies in the social sciences, and generically in comparative politics, the researcher cannot control the assignment of causes to units. We cannot randomly assign political regimes, trade policies, revolutions, or civil wars to countries. Such studies, in which the investigator cannot control the assignment of potential causes to background characteristics, are generically referred to as "observational studies". Yet the type of research that bears this label, reviewed in Rosenbaum's (2002) magisterial treatise, is characterized not only by the sources of data but also by a methodological approach. The method of medical research on the causes of disease is to emulate experiments by treating the data as quasi-experimental. The assignment of causes to units is generated by nature: people are given to the researcher with the number of cigarettes they smoke and with their covariates. Observational studies seek to balance the smokers and non-smokers (or smokers of different intensity) on their observed characteristics. Having reached a satisfactory balance, they then invoke mean independence assumption, thus assuming in fact either that balancing on the observed covariates is sufficient to balance on the unobserved ones or that unobserved factors do not affect the outcome, say the incidence of lung disease. Needless to say, this is a venerated research design in comparative research, going back to J.S. Mill's "method of only difference," and dubbed "the most similar systems design" by Przeworski and Teune (1970).

Yet, following Heckman (2004)[9], I think that analogies with experiments

of the boundary between two settlement regions ended up with different types of tenure systems, it was for reasons mostly unrelated to their innate differences." Since institutions were exogenous with regard to background conditions and since (until independence) they remained the same regardless of the consequences they generated, the observed differences in development can be attributed to institutions.

[9]Heckman refers to what I call "observational" studies as "statistical" and juxtaposes them to "scientific" or "econometric." While my terminology is more neutral, the sub-

are misleading. The crux of the matter is how to identify causal effects *in the absence of random assignment*. Even if they look "quasi-experimental," treating historical data as experiments "but for random assignment" is to hide the central problem under the proverbial rug. The most inane example I recently read was a study which found that women who do not work are more likely to become sick. Suppose that we have matched the working and non-working women on all the observed covariates. Might it still not be true that women who are more prone to sickness are less likely to work? Observational studies treat causes of effects as given, while as social scientists we must consider the possibility that they are endogenous, either directly with regard to their effects or with regard to unobserved background characteristics.

While students of history, which is what we are, share with experimentalists the goal of identifying causal effects and the conceptual apparatus relying on counterfactuals to define them, they need to justify the counterfactual hypotheses. Experimental justification of counterfactuals is that the value of treatment is generated by a mechanism that *could have* assigned a different value to the same unit. The properties of this mechanism – randomization – are known, as is the probability that a particular value of treatment is assigned to any set of covariates. But the observations given to us in comparative politics, the *data*, are generated by some obscure processes, to which we agnostically refer as "history." And if we invoke counterfactuals, we must be assuming that history could have generated a world different from the one in which we live, that realizations of history other than the actual one are possible.

How can we justify such assumptions? They must be somehow disciplined; otherwise we could fantasize in any way we please. As Hathworn (1991: 168) posed the issue, "Are the alternatives to any actual given only by the facts of that actual, or by possibilities that were canvassed at the time, or by very close comparisons? Are there not also theories to suggest possibilities that we would not otherwise have been able to see?" While there is no general answer to such questions, the point is that to engage in practical counterfactual inferences we need some systematic criteria. Without such criteria, "the possibilities we would be entertaining would be possibilities not for an actual, but for what would itself be merely a possible. And at the point, our History or social science would have dissolved into a literature of the imagination" (Hawthorn 1991: 167; see also Kundera 2003).

stance of the distinction is the same.

If we are to be guided by the "facts of the actual," we need to use the world we observe to identify the mechanisms by which history produces observations, specifically, by which it assigns causes to covariates. "Historical studies" are studies that analyze data generated by history and, as distinct from observational studies, inquire into the causes of effects as well as the effects of causes.

4 Potential Biases

We have seen that to identify causal effects, we need to ensure that

$$U = T(U_1 - U_0)] + U_0 = 0,$$

where U is the impact of unobserved factors in $Y = \beta_0(X) + \beta(X)T + U$ and $\beta(X)$ is the average causal effect conditional on X. There are various reasons why this condition may be violated.

4.1 Baseline Bias

To see the potential sources of bias, note first that the causal effect of interest need not be the effect on the average unit but on those units that are actually observed as treated.[10] This *estimand* is typically referred to as the Average effect of Treatment on the Treated, ATT, defined as

$$\beta_{ATT} = E(Y_1 - Y_0|X, T = 1). \tag{5}$$

The value of this parameter tells us how the treatment changes the outcome for those unit that were observed as treated. Note that $E(Y_1|T = 1)$ is observed, while $E(Y_0|T = 1)$ is the missing counterfactual. Now consider the bias of the observed difference, $\bar{\beta}$, as an estimator of β_{ATT}:

$$\begin{aligned} \bar{\beta} - \beta_{ATT} &= E(Y_1|X, T = 1) - E(Y_0|X, T = 0) - E(Y_1 - Y_0|X, T = 1) \\ &= E(Y_0|X, T = 1) - E(Y_0|X, T = 0) \\ &= E(U_0|T = 1) - E(U_0|T = 0), \end{aligned} \tag{6}$$

[10] This effect is of particular interest in remedial policy programs. As Heckman repeatedly points out, it makes no sense to ask what would be the effect of manpower training program on millionaires. In turn, we want to know the effectiveness of such programs for the people who need them and get them.

where the last expression is the difference in the control state between those units that were treated and those that were not, typically referred to as the "baseline bias." Suppose, for example, that an omitted variable, say human capital, H, is correlated with the treatment and it affects the development prospects of a country, so that $E(U_0|H = high, T = 1) > E(U_0|H = low, T = 0)$. Since countries observed under $T = 1$ would have developed faster under $T = 0$ than those actually observed under $T = 0$, the observed difference overestimates the causal effect of T. This bias is sometimes referred to as "the" selection bias, but we will see that there are other potential selection biases than the baseline bias.

4.2 Self-Selection Bias

Now, return to ATE. The bias of $\bar{\beta}$ as the estimator of β_{ATE} is

$$\bar{\beta} - \beta_{ATE} = E(Y_1|X, T = 1) - E(Y_0|X, T = 0) - E(Y_1 - Y_0|X). \quad (7)$$

Adding and subtracting $E(Y_0|T = 1)$ yields

$$\begin{aligned} \bar{\beta} - \beta_{ATE} &= \{E(Y_0|X, T = 1) - E(Y_0|X, T = 0)\} + \\ \{E(Y_1 - Y_0|X, T &= 1) - E(Y_1 - Y_0|X)\} = \\ \{E(U_0|T &= 1) - E(U_0|T = 0)\} \\ +\{E(U_1 - U_0|T &= 1) - E(U_1 - U_0)\}. \end{aligned} \quad (8)$$

The term in the first curly brackets is the by now familiar baseline bias. The term in the second brackets, in turn, is best thought of as "self-selection" bias. This term is the difference between the effect of treatment on those who were actually treated and on the average unit. But why would the effect of the treatment on the treated differ from its effect on those who are not? One reason is that recruitment to treatment depends on something unobserved by the researcher but anticipated by the unit. This will occur if individuals seek treatment for some reasons other than the $X's$ observed by the researcher or if they comply differently with the treatment depending on the $X's$. Suppose – I am not asking you to believe it – that political elites which opt for democracy also know how to make the country develop faster. Then the effect of democracy on development for the countries observed as democracies will differ from the effect on the average country: a self-selection bias.

4.3 Post-treatment Bias: "Manipulability" and "Attributes"

Thus far we have assumed that the $X's$ and the $V's$, called here "covariates," do not change with treatment. The assumption was that causes can be manipulated one-at-a-time. But suppose that some of the covariates – call this subset A for "attributes" – change as the effect of treatment: this is called "post-treatment effect" by King and Zeng (2002). Now the treatment may have two effects: a direct one and an indirect via A. We need some identification assumptions to tell these two effects apart.

Can we always make such assumptions? Here we enter into a complex and subtle issue. According to Holland (1986), to qualify as a potential cause, the particular variable must be vulnerable to (potential) *manipulation.* The critical feature of the notion of cause is that different values of the cause can be realized under the same background conditions. This is why attributes, such race or gender, cannot be causes. "Causes," Holland says, "are only those things that could, in principle, be treatments in experiments" (1986: 954). What distinguishes statistical association from causation is (potential) manipulability: "the schooling a student receives can be a cause, in our sense, of the student's performance on a test, whereas the student's race or gender cannot." It makes no sense to say "Joe earns \$500 less than Jim *because* Joe is black," since skin color (called "race" in the United States) cannot be manipulated. Causal inference is concerned with the effect of causes under specific background conditions ("on specific units") and attributes cannot be manipulated without changing these conditions.

Note that this arguments confounds two propositions: (1) T cannot be manipulated and (2) T cannot be manipulated without changing A. The first one says that we cannot change the skin color of an individual. The second says that we can change it but if we change it, we will also change other characteristics of this individual (or the treatment of this individual by others). The confusion becomes apparent when we read that "An attribute cannot be a cause in an experiment, because the notion of *potential exposability* does not apply to it. The only way for an attribute to change its value [so it can be changed!] is for the unit to change in some way and no longer be the same unit" (Holland 1986: 954). Now, if (1) is true, it may still be true that there are other units that have the same background conditions but a different value of T and we can use the conditional mean independence assumption

to identify the causal effect. Only if (2) is true, does identification become impossible.

Consider an example closer to our practice: the location of a country in Africa, which in many analyses appears to affect civil strife and economic growth. Does it make sense to say that "the effect of Africa on growth is β"? "Africa" is clearly an attribute by Holland's definition, a set of related unobserved characteristics. If history had placed Zimbabwe in Latin America, it would have no longer been Zimbabwe: it would differ in various ways that make Africa distinct from Latin America. Hence, relying on the Africa dummy to generate counterfactuals would generate a "post-treatment bias."

King and Zeng (2002: 21) emphasize that controlling (matching) for variables that are endogenous with regard to treatment generates bias. This can be seen as follows. For simplicity, suppose that assignment is random, so that there is no baseline or self-selection bias, but $X_1 = X_0 + \delta T$. Then conditioning on X,

$$E(Y_1 - Y_0|X) = E(Y_1|X_0+\delta T) - E(Y_0|X_0) = E(Y_1 - Y_0|X_0) + \{E(Y_1|X_0+\delta T) - E(Y_1|X_0)\}, \tag{9}$$

where the last term is the "post-treatment bias." For example, Przeworski et al. (2000) found that labor force, a source of economic growth, increases faster under dictatorships. Conditioning on the growth of labor force would then generate post-treatment bias.

4.4 Non-independence Bias: "SUTVA"

One final implicit assumption concerns independence of the Y variables. This assumption of independence is called SUTVA, for "stable unit treatment value." Suppose that the units are individuals and that they learn from one another, so that $y_i = f(y_j)$. This means that the performance of the treated may affect the performance of the untreated, or vice versa. In Lucas's (1988) growth model, a young plumber learns from the experienced one. Hence, if we take the difference in their productivity as the effect of experience, it will be underestimated because of the externality. Or take T to be "export-oriented" strategy. South Korea adopted this strategy early and had high growth rates. Brazil adopted it late. But suppose that Brazil had adopted it early: would the growth rate of Korea been the same? If it would not have been the same, the values observed for Korea under treatment depend on the

realization of the treatment variable for Brazil: hence the Korean values are not "stable." One needs some kind of an equilibrium model to identify the causal effect when this assumption is violated.

5 Historical Studies

5.1 An Example

Let us look at an example, concerning the effect of political regimes, dichotomized as democracies ($T = 0$) and dictatorships ($T = 1$), on economic development between 1950 and 2000.[11]

*** Table 2 here ***

Here the cause is the political regime. The observed covariate X is GDP/cap lagged one year. The unobserved variable V is the "quality of leadership." The outcome (performance) variable Y is the rate of growth of total GDP.

The observations are sorted from the poorest to the wealthiest country-year. Note that nine observations with the lowest GDP/cap are dictatorships. Indeed, there are only four democratic years until the $155-th$ poorest observation: they are listed in the table. In turn, 82 observations with the highest GDP/cap are all democracies without an autocratic match: the wealthiest dictatorship, Singapore in 1996, ranked $5079-th$. The wealthiest dictatorship outside Singapore was in Taiwan in 1995: it ranked $4589-th$. Between Uganda in 1982, with per capita income of \$630, and Thailand in 1995 with income of \$14,036, there are dictatorships as well as democracies, but their distribution is still not the same. Dictatorships are heavily bunched in poor countries, democracies are frequent in rich ones. Indeed, 90% of observations of dictatorships are below \$6,000, while only 42% of democracies are below this level. Figure 1 shows the density of per capita incomes for the two regimes.

[11] The economic data are combined from Penn World Tables Release 5.6 and Release 6.1 They are in 1995 purchasing power parity dollars. Regime classification is due to Cheibub and Gandhi (2004). Six Middle Eastern oil countries are excluded.

*** Figure 1 here ***

We see in Table 2 that $\bar{y}_0 = 3.68$ and $\bar{y}_1 = 4.27$, so that $\bar{\beta} = 0.61$. At these rates, total income doubles in 16.2 years under dictatorships and in 18.8 years under democracy. Hence, it looks like dictatorships grow faster.

5.2 Types of Estimators

How can we identify causal effects when the data are generated by history?[12] Basically, we can adopt two approaches: drop the observations that are not "comparable," restricting identification of causal effects to those that are, or keep all the observations and generate hypothetical matches for each of them. Matching procedures would eliminate (or give almost zero weights) all the observations in Table 2 that do not have close matches, while procedures generating hypothetical counterfactuals would fill all the growth cells for which numbers are not available.

5.2.1 Matching

One way to proceed is to *match on observables.*[13] Say we want to examine the effect of guaranteed income programs on labor supply. We observe some wealthy countries with such programs (*Revenue minimum d'insertion* in France) and many countries, rich and poor, without them. We would not want to match the wealthy treatment cases with controls from poor countries. Hence, we use as controls countries with comparable per capita income, and restrict our causal inference to such countries.

Matching takes the assignment of causes as given and calculates causal effects conditional on the assignment of causes realized by history, relying on the conditional mean independence assumption

$$E(Y_j|X, T = j) = E(Y_j|X) \forall j, \tag{10}$$

[12] For overviews of estimators see Angrist and Krueger (1999), Berk (2004, Chapter 5), Du- o (2002), Persson and Tabelini (2003, Chapter 5), or Winship and Morgan (1999). For reasons of space, I do not discuss difference-in-difference estimators, for which see Woolridge (2002) and Bertrand, Duflo, Mullainathan (2004).

[13] On matching estimators, see Rosenbaum (2002), Imbens (2002), Becker and Ichino (2002), and, more critically, Heckman (2004).

which says that the value of Y in any state j does not depend on the state T in which a unit is observed once it is conditioned on the observed covariates. This is the same assumption as conditional mean independence introduced above, but written more generally to emphasize that the cause may assume any set of values.

Matching estimators are vulnerable to two problems:

(1) Droping observations reduces the scope of generality. Sometimes, as in the example of minimum income programs, this is not a loss. It is not a loss because the probability that a poor country would institute these programs is zero: poor countries cannot afford such programs, so that the question how these programs would affect labor supply in poor countries is moot. But how should we proceed when this probability is positive under all conditions, yet very differently distributed with regard to these conditions, as in the case of political regimes? What to do with observations without a close match? You know from Table 2 that there are poor dictatorships without a close democratic match and rich democracies without a close autocratic match. We can throw these observations out. Alternatively, and almost equivalently, we can keep them in but assign them a very low weight.[14] In either case, we have to worry whether the causal effect is the same for those observations with close matches and those without them. If we are matching on GDP/cap, we have to be concerned about the shape of the function that relates this covariate to the effect variable, the rate of growth. And observe that this relation is non-linear:

*** Figure 2 here ***

Consider the upper tail of distribution of GDP/cap. There are only 10 country-years of dictatorships with incomes above $\$14,036$, all from Singapore which grew at the spectacular rate of 7.86, while 562 observations of democracies in this range have mean growth of 2.82. Moreover, there are 82 observations of democracies that are wealthier than the wealthiest dictatorship: Singapore in 1996. Are we willing to believe that dictatorships grow

[14] Depending on the algorithm, matching estimators treat differently observations that cannot be matched exactly. When matching is restricted to common support or when it is confined to balanced strata, observations without a match are ignored. When some kind of distance measure is employed, distant matches obtain weights approaching zero.

faster in this range? As King and Zeng (2002) emphasize, extrapolations out of range of common support are highly sensitive to the form of the function.

(2) We can match on observables. But should we not worry about unobservables? Suppose that leaders of some countries go to study in Cambridges, where they absorb the ideals of democracy and learn how to promote growth. Leaders of other countries, however, go to the School for the Americas, where they learn how to repress and nothing about economics. Dictatorships will then generate lower growth because of the quality of the leadership, which is "Not Available" in Table 2. Since this is a variable we could not observe systematically, we cannot match on it. And it may matter. Conditional mean independence – the assumption that unobserved factors do not matter – is very strong, and likely to be often false in cross-national research.

All that was said about matching applies to regression models that control for the observables. Matching is just a non-parametric regression: both generate means of Y conditional on X and T. Moreover, as observed respectively by Manski (1995) and Achen (1986), both matching and parametric regressions that control for observables may in fact exacerbate the biases due to selection on unobservables.

Both matching and parametric regression estimates can be subjected to sensitivity analysis. Given assumptions about the unobservables, one can calculate the range of estimates that are compatible with the observed data (Manski 1995). Rosenbaum (2002, Chapter 4) presents methods for quantifying the sensitivity of the estimates of causal effects under different assumptions. Obviously, the more plausible the assumption and the narrower the bounds, the more credible is the estimate.

5.2.2 Instrumental Variables

Instrumental variables estimator is based on the assumption of conditional mean independence in the form:

$$E(Y_j|X, Z, T = j) = E(Y_j|X, Z)\forall j. \tag{11}$$

The idea is the following. Suppose that after conditioning on X, Y_j still depends on T, in other terms that $cov(T, U) \neq 0$. Now, suppose that there is a variable Z, called an "instrument," such that

$$cov(Z, T) \neq 0 \tag{12}$$

and

$$cov(Z, U) = 0. \tag{13}$$

Then conditioning on X and Z satisfies (11). Thinking in regression terms, let $\hat{Y} = f(Z)$ and $\hat{T} = g(Z)$. Then, by assumption (13), β in $\hat{Y} = \beta\hat{T}$ is that part of the causal effect of T on Y which is independent of U.

To qualify as an instrument, a variable must be related to the cause and only to the cause, so that its entire effect is transmitted by the cause. Note that while the assumption that the instrument is related to the cause (conditional on all exogenous variables) can be and should be tested, the assumption that it is independent of the conditions that also shape the effect is not testable.

Instruments must be correlated with the cause. Weak instruments (those weakly correlated with the treatment) can generate biased estimates even with very large samples. But instruments cannot be too strongly correlated with the cause. In the limit, if the instrument and the cause are the same, the instrument is as endogenous as the cause: this is "the curse of strong instruments." The causal effect cannot be identified, because it is impossible to separate the impact of the cause from that of the conditions that give rise to it.

In turn, the "exclusion restriction" (13) requires that the instrument have no effect that is not mediated by the cause. Moreover, given that $U = T(U_1 - U_0) + U_0$,

$$cov(Z, U) = cov(Z, U_0) + cov(Z, T(U_1 - U_0)). \tag{14}$$

Hence, the exclusion restriction has two parts, and Heckman (1996, 2004) repeatedly makes the point that, even if $cov(Z, U_0) = 0$, in the presence of unobserved self-selection the second covariance will not be zero.

Since the exclusion restriction is not testable, it necessitates conjuring and dismissing stories about rival channels through which the instrument may affect the outcome. For example, Acemoglou, Johnson, and Robinson (2002), who use settler mortality at the time of colonialization as the instrument for institutions, have to argue that the natives were not vulnerable to the same sources of mortality as the settlers: otherwise the causal impact of settler mortality would be transmitted by the productivity of the natives in addition to the path via institutions. Yet Djankov et al. (2003) pointed

out that settler mortality does not qualify as an instrument since it has an impact on economic performance via the human capital of the settlers. (See also Glaeser et al. 2004).

Justifying instruments is an art: one has to tell a story and it better be a good one. Identification is sometimes aided by the structure of the data. But proponents of instrumental variables often overstate their case. Angrist and Krueger (2001: Table 1), for example, distinguish between "natural experiments," which they never define, and "randomized experiments" as aids in identification. In turn, according to Woolridge (2002: 88), "A natural experiment occurs when some (often unintended) feature of the setup we are studying produces exogenous variation in an otherwise endogenous explanatory variable." Yet as long as the assignment is not random, at best we have "quasi-experiments," in which the units are matched on observables, but with no guarantees about unobservables. Finding such data makes the story better, but it is still a story.

5.2.3 Selection on Unobservables

Both matching and instrumental variables estimators condition on observed covariates and both are vulnerable to the influence of unobserved variables that are correlated with the treatment. Another approach conditions on unobserved as well as on the observed covariates. One way to think of these estimators is that they emulate experiments, but differently than matching does: not by eliminating observations that do not have an observed match but by creating observations to match all the observed values. The assumption is that if the conditioning is correct, then the resulting data have the same structure as if history had performed a random experiment assigning different values of treatment to each set of observed and unobserved covariates. Since the conditional mean independence of the form

$$E(Y_j|X, Z, V, T = j) = E(Y_j|X, Z, V) \forall j \tag{15}$$

holds whenever assignment is random, the only issue with regard to these estimators is whether they correctly emulated random assignment.

The basic idea is the following. We first describe the process by which the observed assignment of causes was generated by history:

$$T^* = Z\alpha + V, T = 1(T^* > 0), V \sim (0, 1). \tag{16}$$

This equation says that the propensity toward being observed under treatment depends on observable variables Z and unobserved factors V and that we observe $T = 1$ if $T^* > 0$. Secondly, we exploit the possibility that $cov(V, U) \neq 0$, by expressing $E(U_j|T = j)$ in

$$E(Y_j|X, T = j) = E(Y_j|X) + E(U_j|T = j), \tag{17}$$

as

$$E(U_j|T = j) = \theta_j E(V|T = j), \tag{18}$$

where the latter expectation can be estimated from (16). Finally, we substitute, to obtain

$$E(Y_j|X, T = j) = E(Y_j|X) + \theta_j E(V|T = j), \tag{19}$$

which can be now estimated by least squares. The OLS coefficients of $E(Y_j|X) = X\beta_j$ can be then used to generate counterfactual values of Y_j for the cases in which it is not observed, thus filling all the missing values in Table 2. Finally, for $j = 1, 0$,

$$\hat{\beta}_{ATE} = E(Y_1|X) - E(Y_0|X) = (\hat{\beta}_1 - \hat{\beta}_0)X,$$

is the estimator of the average causal effect.

Note that we still have to be concerned about strong endogeneity of treatment. In principle, it has to be true that $0 < \Pr(T = 1|Z) < 1 \forall Z$. Otherwise, the counterfactuals cannot be realized given the mechanism by which history assigns treatments, so that the entire exercise is moot. The main vulnerability of this class of estimators stems from the untestable assumption about the joint distribution (V, U_1, U_0).

5.3 Back to the Example

To illustrate these methods, let us return to our example, arbitrarily taking the "treatment" to be dictatorship. We will test the robustness of the estimators with regard to (1) the specification of the selection mechanism and (2) the functional form of the outcome equation. The model to be estimated is

$$p = \Pr(REGIME = 1) = \Pr(Z\alpha + V > 0) = F(Z\alpha). \tag{22a}$$

$$GROWTH = f(GDP/cap) + \beta * REGIME + U. \qquad (22b)$$

Table 3 contains all the results.

*** Table 3 here ***

We specify the selection mechanism (22a) in two ways. The "static" specification includes three variables: lagged per capita income, lagged proportion of countries in the world that were democratic in a particular year, and lagged number of completed spells of democracy in the history of a country. The "dynamic" specification adds to this list the lagged regime and its interactions with the three covariates. The static version assumes that regimes are generated each year anew according to the values of the covariates. The dynamic version presupposes that regimes are generated by a Markov process in which the transition probabilities depend on the covariates (Przeworski 2004a). In both cases, we estimate by probit the probability p that a regime is a dictatorship conditional on the covariates.

The static version fits reasonably well: all the variables are highly significant and pseudo $R^2 = 0.33$. Moreover, there are relatively few observations for which the probability of dictatorship is lower than 0.05 or higher than 0.95. In other words, under most conditions, as characterized by the values of the covariates, almost all countries have some reasonable probability of having both a democratic and a dictatorial regime during a particular year. This is not true under the dynamic specification, which predicts regimes much better: pseudo $R^2 = 0.86$. Now there are many observations for which the probability of dictatorship is almost zero or almost one. We are thus back to a philosophical question: should we entertain counterfactuals when the mechanism by which history assigns causes to covariates is almost deterministic? Obviously regimes are highly endogenous. Yet as long as $0 < p < 1$, history may have realized regimes that have a very low probability: in fact, even among the observations that are almost certain to be dictatorships, we observe some democracies (India!).

As Table 3A shows, for the matching (Imbens's nnmatch in Stata, with one match) and the Heckman two-step estimators (but not OLS and IV) the

specification of the selection mechanism makes some difference for the estimate of the causal effect of regimes. Under the static specification, dictatorships appear to grow somewhat faster, while the difference between regimes is lower under the dynamic specification.

Both the matching and the Heckman estimators are also sensitive to the form of the function that relates per capita income to the rate of growth. Note first that a non-parametric regression (lowess smooth) of the rates of growth on lagged per capita income, shown in Figure 2, suggests that the function is cubic, with a maximum around $\$1,500$ and a minimum around $\$23,000$. Columns 3 and 4 of Table 3A show estimates of the average causal effect with a cubic specification of the function. It is apparent that the estimates are higher when higher order terms are introduced.

*** Figures 3 here ***

Why would it be so? Inspect Figure 3, which shows the same lowess smooth separately for the two regimes and recall that the sharply upward segment of the dictatorial line is due almost exclusively to Singapore. Now, albeit in different ways, all the estimators compare the developed democracies to these observations of Singapore, which is their closest match. They predict that if the developed countries which are observed as democracies were to become dictatorships, they would grow like Singapore.[15] The effect of fast growing wealthy dictatorships is evident when we consider separately the effect of dictatorship on the cases actually observed as dictatorships (ATT) and those actually observed as democracies (ATC). Table 3B, which details the dynamic cubic specification, shows that if the cases observed as democracies had been dictatorships, they would have grown much faster, while if the cases observed as dictatorships had been democracies they would have grown somewhat slower (Imbens) or at about the same rate (Heckman).

I present this exercise to show the basic issues entailed in estimating causal effects. Estimates of causal effects are likely to be sensitive to assumptions

[15] Note, however, the different estimators do it differently. Consider the wealthiest observation in the data set: Luxembourg in 2000. Matching will assign it to it the rate of growth of the wealthiest dictatorship, Singapore in 1996. But Heckman, which estimates the parameters separately for the two regimes, will multiply the income of Luxembourg by the cubic coefficient of the dictatorial regression, hence generating a larger difference.

used to identify the models and to correct the potential biases. Specifying better the determinants of causes affects the estimates of their effects. The non-linearity of the relation between GDP/cap and the growth rates plays havoc even when we match for the observables or generate counterfactuals by studying selection on unobservables.

6 Conclusion

When we cannot control the assignment of the potential causes, we are at the mercy of history. The information we can squeeze from the data is a matter of luck. And luck may vary from context to context. History may be very kind and in fact randomize the unobserved, as well as the observed, background characteristics, thus generating a "natural experiment." Unfortunately, most historical data may have the structure illustrated by our example, where dictatorships were more likely to occur in poor countries and democracies in wealthy ones. What this means is that political regimes are endogenous with regard to the level of development. Suppose that this relation were perfect: that high income were a necessary and sufficient condition for a country to be democratic. Now all dictatorships would be poor and all democracies rich, so that we could never tell whether their rates of growth were due to their income or their political institutions. Even worse, suppose that political institutions survive only if they generate development, so that they are endogenous with regard to growth. Identification would not be possible. Hence, endogeneity makes identification difficult. And there are good reasons to think that institutions, policies, and events are endogenous.

The motor of history is endogeneity (Przeworski 2004a). From some initial circumstances and under some invariant conditions ("geography"), wealth, its distribution, and political institutions are mutually interdependent and evolve together. Since we can never completely specify this process, we observe some randomness. Indeed, we exploit this randomness to identify the particular models of this process: for identification, we need to observe different values of causes under the same background conditions. And here we face a paradox. The better we specify our models, the more endogenous loops we consider, the more difficult it becomes to identify their causal structure. As Mariscal and Sokoloff (2000: 198) correctly lament, "When variables are mutually reinforcing or simultaneously determined, discerning what is exogenous and what is endogenous is not transparent."

The difficulty presented by endogeneity is to distinguish the effects of causes from the effects of conditions under which they operate. Do democracies grow slower because they are democracies or because they tend to occur under conditions under which economies grow slower regardless of political institutions? Did the French revolution generate little social change, as Tocqueville (1964 [1856]) would have it, because revolutions result in little change or because they occur only in countries resistant to change?[16]

A necessary condition for identification is path *independence*: situations where historical paths diverged at some time from the same background conditions.[17] If different values of causes are to be found under the same background conditions, at some time the paths of causes, say political institutions, must diverge. In India, random assignment of different tributary systems to identical underlying conditions resulted from the ignorance about these conditions by the colonizing force. The political institutions of Costa Rica and Guatemala, which according to Yashar (1997) shared almost identical historical conditions until the 1940s, diverged as the result of policies adopted during democratization of the late 1940s-early 1950s.

Suppose that history is perfectly path dependent. From some initial conditions $\{X_0, U_0, T_0, Y_0\}$, all these variables evolve over time in a unique way. This means that X, U, and T vary together and there is no way to isolate the effect of T on Y. We can still engage in descriptions: we could say, for example, that in those areas where the colonizers found few natives ($X_0 = low$), they established good institutions ($T_0 = good$), and generated development ($Y = high$), while in those area where there were natives to exploit ($X_0 = high$), they established bad institutions ($T_0 = bad$), and generated stagnation ($Y = low$). This is a description, in the sense that all we can say is that all these features come together but cannot isolate the effect of T independent of X (and perhaps U).[18] This may be the most we can do, but I suspect that the temptation to entertain counterfactuals is irresistible: we do want to know what would have happened had good institutions been established in areas where there were natives to exploit. We may be hurling ourselves against the impossible, but hurl we do and hurl we will.

[16] As argued by Fearon (1991), small N does not change the logic of inference.

[17] Note that "path dependence" is not, as some would have it, an approach but a historical fact, to be determined.

[18] The language of "qualitative" vs. "quantitative" studies to describe this difference is neither here nor there: it makes no difference whether the variables range over numerals or string of words.

The question posed in the title is not rhetorical. To identify causal effects we need assumptions and some of these assumptions are untestable. In Heckman's (2004: 51) words, "There is no assumption-free method of causal inference." The reason is that even if we observe the marginal distributions of outcomes separately under different values of the potential cause, by construction we cannot observe their joint distribution. Moreover, no single estimator can correct for all the potential biases. And since each estimator invokes different assumptions, we cannot be certain that the conclusions would be robust.

What, then, can we do in the presence of endogeneity? All we can do in my view is to try different assumptions and hope that the results do not differ: Persson's and Tabelini's (2003) study is exemplary in this respect. If they do not differ, we know that the conclusions are at least robust with regard to different assumptions about the sources of bias. If they do differ, all we can do is to throw our hands up in the air. Where history was kind enough to have generated different causes under the same conditions we will know more and know better. But history may deviously generate causes endogenously and this would make our task next to impossible.

7 References

Acemoglou, Daron, Simon Johnson, and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *The Quarterly Journal of Economics 117:* 1231-1294.

Achen, Christopher. 1986. *The Statistical Analysis of Quasi-Experiments.* Berkeley, CA: University of California Press.

Amemyia, Takeshi. 1994. *Introduction to Statistics and Econometrics.* Cambridge, MA: Harvard University Press.

Angrist, Joshua D. and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." Chapter 23 in O. Ashenfelter and D. Card (eds.), *The Handbook of Labor Economics*, vol III. North Holland.

Angrist, Joshua D. and Alan B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives 15:* 69-85.

Banerjee, Abhijit and Lakshmi Iyer. 2002. "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in

India." Ms. Department of Economics, MIT.

Becker, Sascha O. and Andrea Ichino. 2002. "Estimation of average treatment effects based on propensity scores." *The Stata Journal 7:* 1-19.

Berk, Richard A. 2004. *Regression Analysis: A Constructive Critique.* Thousand Oaks: Sage. Chapter 5.

Bertrand, Marianne, Esther Duflo, and Sendhill Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics 119:* 249-275.

Campell, D.T. and Stanley J.C. 1963. *Experimental and quasi-experimental designs for research.* Chicago: RandMcNally.

Cheibub, José Antonio and Jennifer Gandhi. 2004. "Classifying Political Regimes." Ms. Department of Political Science, Yale University.

Dawid, A.P. 2000. "Causal Inference without Counterfactuals." *Journal of the American Statistical Association 95:* 407-424.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "The New Comparative Economics." *Journal of Comparative Economics 31:* 595-619.

Duflo, Esther. 2002. "Empirical Methods." Class notes. Department of Economics, MIT.

Edgington, Dorothy. 2001. "Conditionals." In Lou Goble (ed.), *The Blackwell Guide to Philosophical Logic.* Oxford: Blackwell. Pages 385-414.

Fearon, James. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics 43:* 169-95.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Political Institutions and Human Capital in Economic Development." Ms. Department of Economics, Harvard University.

Goodman, Nelson. 1979. *Fact, Fiction, and Forecast.* 4th edition. Cambridge, MA: Harvard University Press.

Hawthorn, Geoffrey. 1991. *Plausible Worlds: Possibility and understanding in history and the social sciences.* Cambridge: Cambridge University Press.

Heckman, James J. 1996. "Instrumental Variables: A Cautionary Tale." Technical Working Paper No. 185. Cambridge, MA: National Bureau of Economic Research.

Heckman, James J. 1997. "Instrumental Variables: A Study in Implicit Behavioral Assumptions Used in Making Program Evaluations." *The Journal of Human Resources 32:* 441-462.

Heckman, James J. 1992. "Randomization and Social Policy Evaluation," in C. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs.* Cambridge, MA: Harvard University Press.

Heckman, James J. 2004. "The Scientific Model of Causality." Working Paper. Department of Economics, University of Chicago.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association 81:* 945-960.

Imbens, Guido W. 2002. "Semiparametric Estimation of Average Treatment Effect under Exogeneity: A Review." Working Paper. Department of Economics, University of California at Berkeley.

King, Gary, and Langche Zeng. 2002. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference." http://GKing.Harvard.Edu

Kundera, Milan. 2003. *The Art of the Novel.* New York: Perennial.

Lewis, David. 1973. *Counterfactuals.* Cambridge, MA: Harvard University Press.

Lucas, Robert E. Jr. 1988. "On the mechanics of economic development." *Journal of Monetary Economics 22:* 3-42..

MacIntyre, Alasdair. 1972. "Is a Science of Comparative Politics Possible?" In Peter Laslett, W.G. Runciman, and Quentin Skinner (eds.), *Philosophy, Politics and Society.* Oxford: Basil Blackwell. Pages 8-26.

Mackie, J.L. 2002 [1973]. "The Logic of Conditionals." In Yuri Balashov and Alex Rosenberg (eds.). 2002. *Philosophy of Science: Contemporary Readings.* London: Routledge. Pages 106-114.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences.* Cambridge, MA.: Harvard University Press.

Mariscal and Sokoloff, 2000:

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Persson, Torsten and Guido Tabelini. 2003. *The Economic Effects of Constitutions.* Cambridge: MIT.

Przeworski, Adam, José Antonio Cheibub, Fernando Limongi, and Michael E. Alvarez. 2000. *Democracy and Development.* New York: Cambridge University Press.

Przeworski, Adam. 2004a. "The Last Instance? Are Institutions the Primary Cause of Economic Development?" *European Journal of Sociology 15:* 165-188.

Przeworski, Adam. 2004b. "Economic Development and the Transitions to Democracy." Ms. Department of Politics, New York University.

Przeworski, Adam and Henry Teune. 1970. *The Logic of Comparative Inquiry.* New York: John Wiley and Sons.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development.* New York: Cambridge University Press.

Rosenbaum, Paul R. 2002. *Observational Studies.* New York: Springer-Verlag. 2nd edition.

Rosenbaum, Paul R. and D.B. Rubin. 1983. "The central role of the propensity score in observational studies." *Biometrika 70:* 41-55.

Quine, W.V. 1953. *From the Logical Point of View.* Cambridge, MA: Harvard University Press.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mi- in.

Stalnaker, Robert C. 1987. *Inquiry.* Cambridge, MA: MIT Press.

Tocqueville, Alexis de. 1964 {1856]. *L'ancien régime et la Révolution.* Paris: Galimard.

Winship, Christopher, and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology 25:* 659-707.

Woolridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT.

Yashar, Deborah J. 1997. *Demanding Democracy. Reform and Reaction in Costa Rica and Guatemala, 1879s-1950s.* Stanford, CA: Stanford University Press.

Table 1: Experiments and Quasi-experiments

i	T	X_1	X_2	...	X_k	V_1	V_2	Y_C	Y_T
1	0	1	1	1	1	NA	NA	y_{C1}	NA
2	1	1	1	1	1	NA	NA	NA	y_{T2}
3	0	1	1	1	1	NA	NA	y_{C3}	NA
4	1	1	1	1	1	NA	NA	NA	y_{T4}
5	0	3	5	1	6	NA	NA	y_{C5}	NA
6	1	3	5	1	6	NA	NA	NA	y_{T6}
...	...	...	...	...	...	NA	NA	...	...
...	...	...	...	...	...	NA	NA	...	...
$N-1$	0	17	14	6	9	NA	NA	$y_{C,N-1}$	NA
N	1	17	14	6	9	NA	NA	NA	$y_{T,N}$

Table 2: Fragments of Data Concerning Political Regimes and Economic Development

$n-th\ lowest\ GDP/cap\ n$	$Country$	$Year$	GDP/cap	$Quality$	$Regime$	$Growth\ under\ DEM$	$Growth\ under\ DIC$
1	*Zaire*	1997	310	*N.A.*	*DIC*	*N.A.*	−5.90
...	...	...	...	...	*DIC*	*N.A.*	...
10	*Uganda*	1981	443	*N.A.*	*DEM*	44.36	*N.A.*
13	*Uganda*	1980	451	*N.A.*	*DEM*	0.47.	*N.A.*
69	*Malawi*	1995	545	*N.A.*	*DEM*	26.38	*N.A.*
155	*Uganda*	1982	630	*N.A.*	*DEM*	6.90	...
...	...	...	...	*N.A.*	*DEM*	...	*N.A.*
...	...	...	...	*N.A.*	*DIC*	*N.A.*	...
4589	*Taiwan*	1995	14036	*N.A.*	*DIC*	*N.A.*	6.19.
...	...	...	...	*N.A.*	*DEM*	...	*N.A.*
...	*Singapore*	...	...	*N.A.*	*DIC*	*N.A.*	...
...	...	...	...	*N.A.*	*DEM*	...	*N.A.*
5079	*Singapore*	1996	22642	*N.A.*	*DIC*	*N.A.*	14.22
...	...	...	...	...	*DEM*	...	*N.A.*
5161	*Luxembourg*	2000	41354	*N.A.*	*DEM*	7.68	*N.A.*
Average						3.68	4.27
N						2459	2702

Table 3A: Estimates of Causal Effects of Regimes on the Rate of Growth of Total Income

Estimator	*Static Linear*	*Dynamic Linear*	*Static Cubic*	*Dynamic Cubic*
OLS	−0.20	−0.24	−0.22	−0.23
	(0.60)	(0.60)	(0.60)	(0.60)
Match	0.63	0.44	0.74	0.66
	(0.39)	(1.06)	(0.39)	(1.08)
IV(*pscore*)	0.31	0.33	0.33	0.35
	(0.23)	(0.23)	(0.24)	(0.24)
Heckman	0.63	0.59	1.11	1.04
	(0.01)	(0.01)	(0.07)	(0.07)

Note: Match is Imbens nnmatch with one match. IV with the probability of dictatorship (pscore) as the instrument. (2SLS with separate instruments generates almost identical results). Heckman is the Heckman two-step estimator, with separate regressions for each regime. In the static specification, the probit used to generate pscore and to estimate the first stage of Heckman procedure uses GDP/cap (and its higher powers, as indicated), ODWPlag (Proportion of other countries that are democracies in a given year), and STRAlag (The number of completed spells of democracy in the history of the country). These variables are used as controls in OLS and as instruments in IV. The dynamic specification adds to this list the lagged regime and its interactions with all the exogenous variables. Standard errors in parentheses.

Table 3B: Detailed Estimates of Causal Effects of Regimes, Dynamic Cubic Specification

		Hypothetical as					
		Dic	*Dem*		*Heckman*	*Match*	*N*
Observed as	*Dic*	4.39	4.25	*ATT*	0.14(0.01)	−0.84(1.42)	2702
	Dem	5.74	3.72	*ATC*	2.02(0.14)	2.01(1.46)	2459
	All	5.04	4.00	*ATE*	1.04(0.07)	0.69(1.08)	5161

Note: The cell entries are the rates of growth predicted by the second stage of the Heckman estimator.

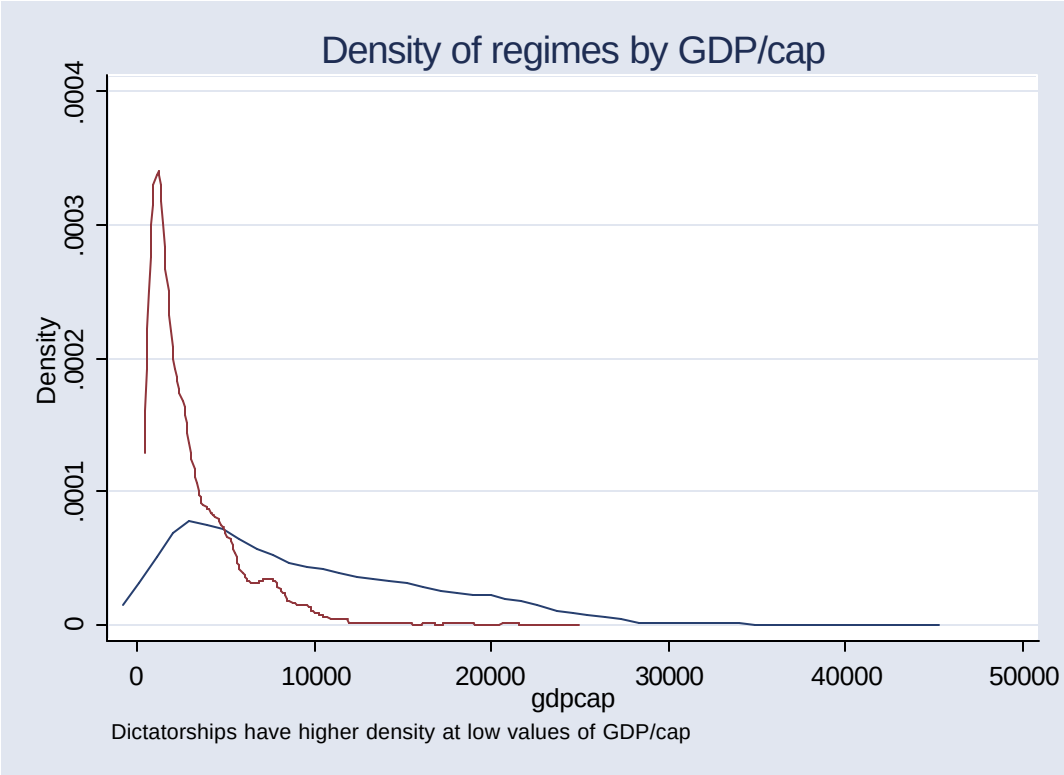
Density of regimes by GDP/cap
Density
0
.0001
.0002
.0003
.0004
0
10000
20000
30000
40000
50000
gdpcap
Dictatorships have higher density at low values of GDP/cap

Figure 1:

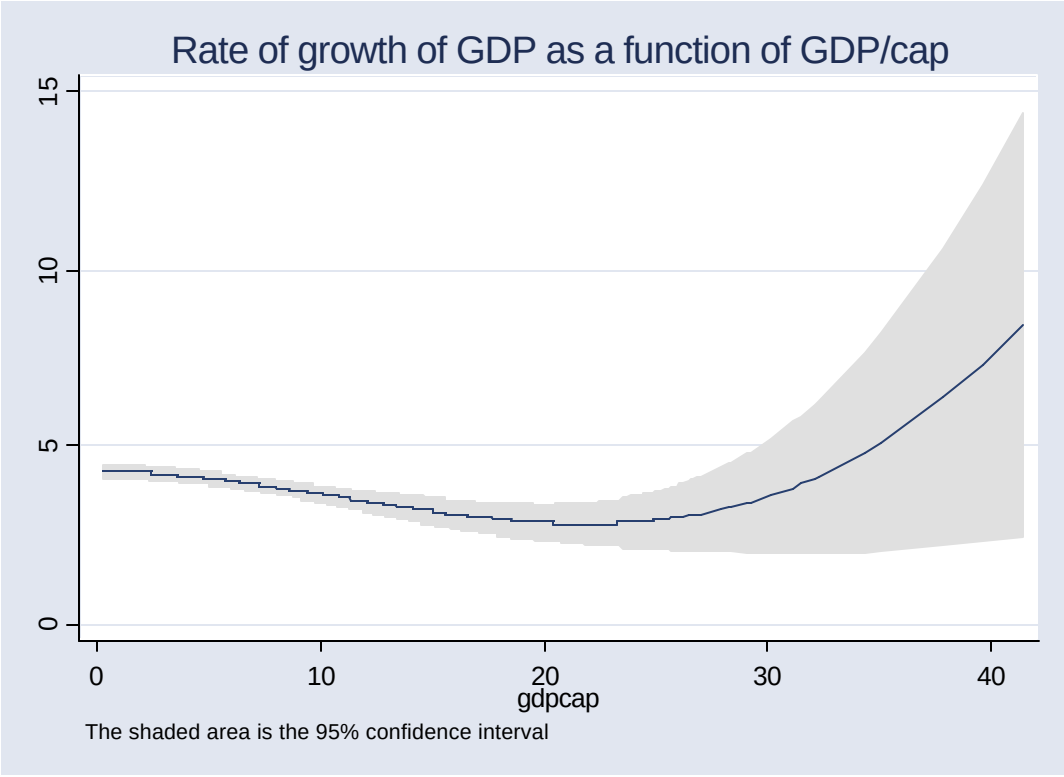
Rate of growth of GDP as a function of GDP/cap
15
10
5
0
0
10
20
30
40
gdpcap
The shaded area is the 95% confidence interval

Figure 2:

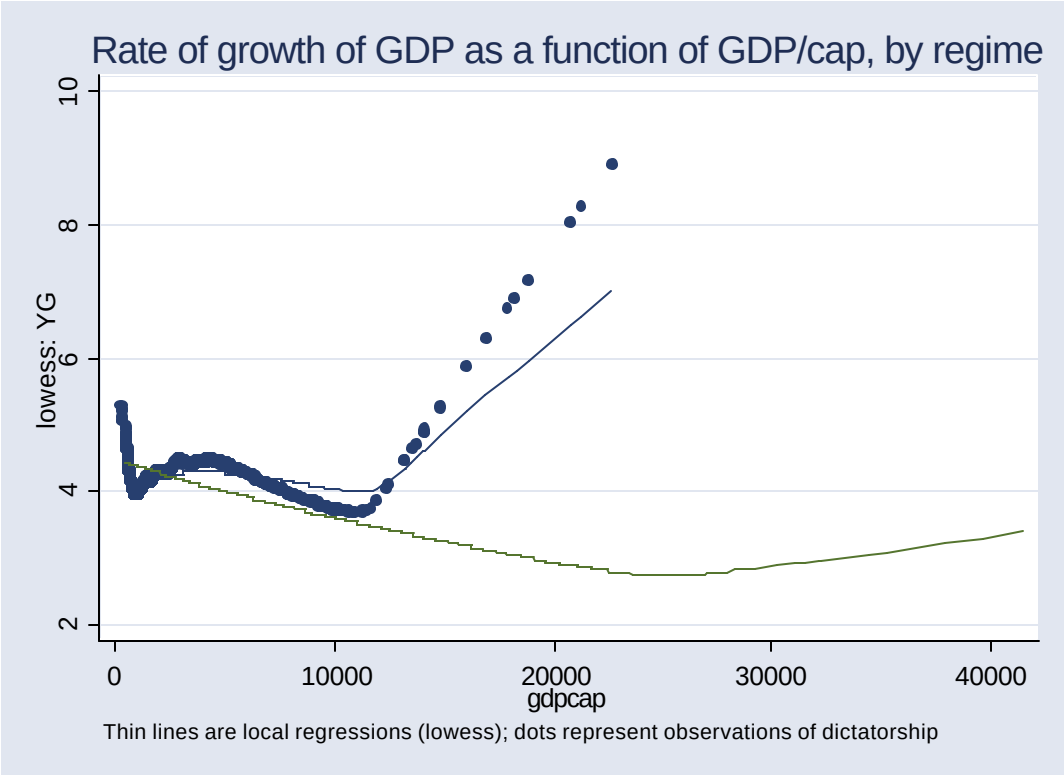
Rate of growth of GDP as a function of GDP/cap, by regime
lowess: YG
2
4
6
8
10
0
10000
20000
30000
40000
gdpcap
Thin lines are local regressions (lowess); dots represent observations of dictatorship

Figure 3: