Chapter 2

The Experimental Ideal

> It is an important and popular fact that things are not always what they seem. For instance, on the planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much—the wheel, New York, wars and so on—while all the dolphins had ever done was muck about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man–for precisely the same reasons. In fact there was only one species on the planet more intelligent than dolphins, and they spent a lot of their time in behavioral research laboratories running round inside wheels and conducting frighteningly elegant and subtle experiments on man. The fact that once again man completely misinterpreted this relationship was entirely according to these creatures' plans.
>
> Douglas Adams, *The Hitchhiker's Guide to the Galaxy* (1979)

The most credible and influential research designs use random assignment. A case in point is the Perry preschool project, a 1962 randomized experiment designed to asses the effects of an early-intervention program involving 123 Black preschoolers in Ypsilanti (Michigan). The Perry treatment group was randomly assigned to an intensive intervention that included preschool education and home visits. It's hard to exaggerate the impact of the small but well-designed Perry experiment, which generated follow-up data through 1993 on the participants at age 27. Dozens of academic studies cite or use the Perry findings (see, e.g., Barnett, 1992). Most importantly, the Perry project provided the intellectual basis for the massive Head Start pre-school program, begun in 1964, which ultimately served (and continues to serve) millions of American children.[1]

[1] The Perry data continue to get attention, particular as policy-interest has returned to early education. A recent re-analysis by Michael Anderson (2006) confirms many of the findings from the original Perry study, though Anderson also shows that the overall positive effects of Perry are driven entirely by the impact on girls. The Perry intervention seems to have done nothing for boys.

2.1 The Selection Problem

We take a brief time-out for a more formal discussion of the role experiments play in uncovering causal effects. Suppose you are interested in a causal "if-then" question. To be concrete, consider a simple example: Do hospitals make people healthier? For our purposes, this question is allegorical, but it is surprisingly close to the sort of causal question health economists care about. To make this question more realistic, imagine we're studying a poor elderly population that uses hospital emergency rooms for primary care. Some of these patients are admitted to the hospital. This sort of care is expensive, crowds hospital facilities, and is, perhaps, not very effective (see, e.g., Grumbach, Keane, and Bindman, 1993). In fact, exposure to other sick patients by those who are themselves vulnerable might have a net negative impact on their health.

Since those admitted to the hospital get many valuable services, the answer to the hospital-effectiveness question still seems likely to be "yes". But will the data back this up? The natural approach for an empirically-minded person is to compare the health status of those who have been to the hospital to the health of those who have not. The National Health Interview Survey (NHIS) contains the information needed to make this comparison. Specifically, it includes a question "During the past 12 months, was the respondent a patient in a hospital overnight?" which we can use to identify recent hospital visitors. The NHIS also asks "Would you say your health in general is excellent, very good, good, fair, poor?" The following table displays the mean health status (assigning a 1 to excellent health and a 5 to poor health) among those who have been hospitalized and those who have not (tabulated from the 2005 NHIS):

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003

The difference in the means is 0.71, a large and highly significant contrast in favor of the *non-hospitalized*, with a t-statistic of 58.9.

Taken at face value, this result suggests that going to the hospital makes people sicker. It's not impossible this is the right answer: hospitals are full of other sick people who might infect us, and dangerous machines and chemicals that might hurt us. Still, it's easy to see why this comparison should not be taken at face value: people who go to the hospital are probably less healthy to begin with. Moreover, even after hospitalization people who have sought medical care are not as healthy, on average, as those who never get hospitalized in the first place, though they may well be better than they otherwise would have been.

To describe this problem more precisely, think about hospital treatment as described by a binary random variable, $\text{D}_i = \{0, 1\}$. The outcome of interest, a measure of health status, is denoted by Y_i. The question is whether Y_i is *affected* by hospital care. To address this question, we assume we can imagine what might have happened to someone who went to the hospital if they had not gone and vice versa. Hence, for any individual there are two potential health variables:

$$\textit{potential outcome} = \begin{cases} \mathrm{Y}_{1i} & \text{if } \mathrm{D}_i = 1 \\ \mathrm{Y}_{0i} & \text{if } \mathrm{D}_i = 0 \end{cases}.$$

In other words, Y_{0i} is the health status of an individual had he not gone to the hospital, irrespective of whether he actually went, while Y_{1i} is the individual's health status if he goes. We would like to know the difference between Y_{1i} and Y_{0i}, which can be said to be the causal effect of going to the hospital for individual i. This is what we would measure if we could go back in time and change a person's treatment status.[2]

The observed outcome, Y_i, can be written in terms of potential outcomes as

$$\begin{aligned} \mathrm{Y}_i &= \begin{cases} \mathrm{Y}_{1i} & \text{if } \mathrm{D}_i = 1 \\ \mathrm{Y}_{0i} & \text{if } \mathrm{D}_i = 0 \end{cases} \\ &= \mathrm{Y}_{0i} + (\mathrm{Y}_{1i} - \mathrm{Y}_{0i})\mathrm{D}_i. \end{aligned} \tag{2.1.1}$$

This notation is useful because $\mathrm{Y}_{1i} - \mathrm{Y}_{0i}$ is the causal effect of hospitalization for an individual. In general, there is likely to be a distribution of both Y_{1i} and Y_{0i} in the population, so the treatment effect can be different for different people. But because we never see both potential outcomes for any one person, we must learn about the effects of hospitalization by comparing the average health of those who were and were not hospitalized.

A naive comparison of averages by hospitalization status tells us something about potential outcomes, though not necessarily what we want to know. The comparison of average health conditional on hospitalization status is formally linked to the average causal effect by the equation below:

$$\underbrace{E\left[\mathrm{Y}_i|\mathrm{D}_i = 1\right] - E[\mathrm{Y}_i|\mathrm{D}_i = 0]}_{\text{Observed difference in average health}} = \underbrace{E\left[\mathrm{Y}_{1i}|\mathrm{D}_i = 1\right] - E[\mathrm{Y}_{0i}|\mathrm{D}_i = 1]}_{\text{average treatment effect on the treated}} + \underbrace{E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 0\right]}_{\text{selection bias}}$$

The term

$$E[\mathrm{Y}_{1i}|\mathrm{D}_i = 1] - E[\mathrm{Y}_{0i}|\mathrm{D}_i = 1] = E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i}|\mathrm{D}_i = 1]$$

is the *average causal effect of hospitalization on those who were hospitalized*. This term captures the averages difference between the health of the hospitalized, $E[\mathrm{Y}_{1i}|\mathrm{D}_i = 1]$, and what would have happened to *them* had they not been hospitalized, $E[\mathrm{Y}_{0i}|\mathrm{D}_i = 1]$. The observed difference in health status however, adds to this causal effect a term called *selection bias*. This term is the difference in average Y_{0i} between those who

[2] The potential outcomes idea is a fundamental building block in modern research on causal effects. Important references developing this idea are Rubin (1974, 1977), and Holland (1986), who refers to a causal framework involving potential outcomes as the Rubin Causal Model.

were and were not hospitalized. Because the sick are more likely than the healthy to seek treatment, those who were hospitalized have worse Y_{0i}'s, making selection bias negative in this example. The selection bias may be so large (in absolute value) that it completely masks a positive treatment effect. The goal of most empirical economic research is to overcome selection bias, and therefore to say something about the causal effect of a variable like D_i.

2.2 Random Assignment Solves the Selection Problem

Random assignment of D_i solves the selection problem because random assignment makes D_i independent of potential outcomes. To see this, note that

$$\begin{aligned} E[\text{Y}_i|\text{D}_i = 1] - E[\text{Y}_i|\text{D}_i = 0] &= E[\text{Y}_{1i}|\text{D}_i = 1] - E[\text{Y}_{0i}|\text{D}_i = 0] \\ &= E[\text{Y}_{1i}|\text{D}_i = 1] - E[\text{Y}_{0i}|\text{D}_i = 1], \end{aligned}$$

where the independence of Y_{0i} and D_i allows us to swap $E[\text{Y}_{0i}|\text{D}_i = 1]$ for $E[\text{Y}_{0i}|\text{D}_i = 0]$ in the second line. In fact, given random assignment, this simplifies further to

$$\begin{aligned} E\left[\text{Y}_{1i}|\text{D}_i = 1\right] - E\left[\text{Y}_{0i}|\text{D}_i = 1\right] &= E\left[\text{Y}_{1i} - \text{Y}_{0i}|\text{D}_i = 1\right] \\ &= E\left[\text{Y}_{1i} - \text{Y}_{0i}\right]. \end{aligned}$$

The effect of randomly-assigned hospitalization on the hospitalized is the same as the effect of hospitalization on a randomly chosen patient. The main thing, however, is that random assignment of D_i eliminates selection bias. This does not mean that randomized trials are problem-free, but in principle they solve the most important problem that arises in empirical research.

How relevant is our hospitalization allegory? Experiments often reveal things that are not what they seem on the basis of naive comparisons alone. A recent example from medicine is the evaluation of hormone replacement therapy (HRT). This is a medical intervention that was recommended for middle-aged women to reduce menopausal symptoms. Evidence from the Nurses Health Study, a large and influential non-experimental survey of nurses, showed better health among the HRT users. In contrast, the results of a recently completed randomized trial shows few benefits of HRT. What's worse, the randomized trial revealed serious side effects that were not apparent in the non-experimental data (see, e.g., Women's Health Initiative [WHI], Hsia, *et al.*, 2006).

An iconic example from our own field of labor economics is the evaluation of government-subsidized training programs. These are programs that provide a combination of classroom instruction and on-the-job training for groups of disadvantaged workers such as the long-term unemployed, drug addicts, and ex-offenders. The idea is to increase employment and earnings. Paradoxically, studies based on non-

experimental comparisons of participants and non-participants often show that after training, the trainees earn less than plausible comparison groups (see, e.g., Ashenfelter, 1978; Ashenfelter and Card, 1985; Lalonde 1995). Here too, selection bias is a natural concern since subsidized training programs are meant to serve men and women with low earnings potential. Not surprisingly, therefore, simple comparisons of program participants with non-participants often show lower earnings for the participants. In contrast, evidence from randomized evaluations of training programs generate mostly positive effects (see, e.g., Lalonde, 1986; Orr, et al, 1996).

Randomized trials are not yet as common in social science as in medicine but they are becoming more prevalent. One area where the importance of random assignment is growing rapidly is education research (Angrist, 2004). The 2002 Education Sciences Reform Act passed by the U.S. Congress mandates the use of rigorous experimental or quasi-experimental research designs for all federally-funded education studies. We can therefore expect to see many more randomized trials in education research in the years to come. A pioneering randomized study from the field of education is the Tennessee STAR experiment designed to estimate the effects of smaller classes in primary school.

Labor economists and others have a long tradition of trying to establish causal links between features of the classroom environment and children's learning, an area of investigation that we call "education production." This terminology reflects the fact that we think of features of the school environment as inputs that cost money, while the output that schools produce is student learning. A key question in research on education production is which inputs produce the most learning given their costs. One of the most expensive inputs is class size - since smaller classes can only be had by hiring more teachers. It is therefore important to know whether the expense of smaller classes has a payoff in terms of higher student achievement. The STAR experiment was meant to answer this question.

Many studies of education production using non-experimental data suggest there is little or no link between class size and student learning. So perhaps school systems can save money by hiring fewer teachers with no consequent reduction in achievement. The observed relation between class size and student achievement should not be taken at face value, however, since weaker students are often deliberately grouped into smaller classes. A randomized trial overcomes this problem by ensuring that we are comparing apples to apples, i.e., that the students assigned to classes of different sizes are otherwise comparable. Results from the Tennessee STAR experiment point to a strong and lasting payoff to smaller classes (see Finn and Achilles, 1990, for the original study, and Krueger, 1999, for an econometric analysis of the STAR data).

The STAR experiment was unusually ambitious and influential, and therefore worth describing in some detail. It cost about $12 million and was implemented for a cohort of kindergartners in 1985/86. The study ran for four years, i.e. until the original cohort of kindergartners was in third grade, and involved about 11,600 children. The average class size in regular Tennessee classes in 1985/86 was about 22.3. The experiment assigned students to one of three treatments: small classes with 13-17 children, regular classes

with 22-25 children and a part-time teacher's aide, or regular classes with a full time teacher's aide. Schools with at least three classes in each grade could choose to participate in the experiment.

The first question to ask about a randomized experiment is whether the randomization successfully balanced subject's characteristics across the different treatment groups. To assess this, it's common to compare pre-treatment outcomes or other covariates across groups. Unfortunately, the STAR data fail to include any pre-treatment test scores, though it is possible to look at characteristics of children such as race and age. Table 2.2.1, reproduced from Krueger (1999), compares the means of these variables. The student

Table 2.2.1: Comparison of treatment and control characteristics in the Tennessee STAR experiment

	Students who entered STAR in kindergarten				
	Variable	Small	Regular	Regular/Aide	Joint P-value
1.	Free lunch	.47	.48	.50	.09
2.	White/Asian	.68	.67	.66	.26
3.	Age in 1985	5.44	5.43	5.42	.32
4.	Attrition rate	.49	.52	.53	.02
5.	Class size in kindergarten	15.10	22.40	22.80	.00
6.	Percentile score in kindergarten	54.70	48.90	50.00	.00

Notes: Adapted from Krueger (1999), Table 1. The table shows means of variables by treatment status. The P-value in the last column is for the F-test of equality of variable means across all three groups. All variables except attrition are for the first year a student is observed, The free lunch variable is the fraction receiving a free lunch. The percentile score is the average percentile score on three Stanford Achievement Tests. The attrition rate is the proportion lost to follow up before completing third grade.

characteristics in the table are a free lunch variable, student race, and student age. Free lunch status is a good measure of family income, since only poor children qualify for a free school lunch. Differences in these characteristics across the three class types are small and none are significantly different from zero. This suggests the random assignment worked as intended.

Table 2.2.1 also presents information on average class size, the attrition rate, and test scores, measured here on a percentile scale. The attrition rate was lower in small kindergarten classrooms. This is potential a problem, at least in principle.[3] Class sizes are significantly lower in the assigned-to-be-small class rooms, which means that the experiment succeeded in creating the desired variation. If many of the parents of children assigned to regular classes had effectively lobbied teachers and principals to get their children assigned to small classes, the gap in class size across groups would be much smaller.

[3]Krueger (1999) devotes considerable attention to the attrition problem. Differences in attrition rates across groups may result in a sample of students in higher grades that is not randomly distributed across class types. The kindergarten results, which were unaffected by attrition, are therefore the most reliable.

Because randomization eliminates selection bias, the difference in outcomes across treatment groups captures the average causal effect of class size (relative to regular classes with a part-time aide). In practice, the difference in means between treatment and control groups can be obtained from a regression of test scores on dummies for each treatment group, a point we expand on below. The estimated treatment-control differences for kindergartners, reported in Table 2.2.2 (derived from Krueger, 1999, Table 5), show a small-class effect of about 5 to 6 percentile points. The effect size is about $.2\sigma$, where σ is the standard deviation of the percentile score in kindergarten. The small-class effect is significantly different from zero, while the

Table 2.2.2: Experimental estimates of the effect of class-size assignment on test scores

Explanatory variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes)	–	–	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	–	–	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	–	–	-13.15	-13.07
			(.77)	(.77)
White teacher	–	–	–	-.57
				(2.10)
Teacher experience	–	–	–	.26
				(.10)
Master's degree	–	–	–	-0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Note: Adapted from Krueger (1999), Table 5. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors that allow for correlated residuals within classes are shown in parentheses. The sample size is 5681.

regular/aide effect is small and insignificant.

The STAR study, an exemplary randomized trial in the annals of social science, also highlights the logistical difficulty, long duration, and potentially high cost of randomized trials. In many cases, such trials are impractical.[4] In other cases, we would like an answer sooner rather than later. Much of the research

[4] Randomized trials are never perfect and STAR is no exception. Pupils who repeated or skipped a grade left the experiment. Students who entered an experimental school one grade later were added to the experiment and randomly assigned to one of the classes. One unfortunate aspect of the experiment is that students in the regular and regular/aide classes were reassigned after the kindergarten year, possibly due to protests of the parents with children in the regular classrooms. There was also some switching of children after the kindergarten year. Despite these problems, the STAR experiment seems to have been an

we do, therefore, attempts to exploit cheaper and more readily available sources of variation. We hope to find natural or quasi-experiments that mimic a randomized trial by changing the variable of interest while other factors are kept balanced. Can we always find a convincing natural experiment? Of course not. Nevertheless, we take the position that a notional randomized trial is our benchmark. Not all researchers share this view, but many do. We heard it first from our teacher and thesis advisor, Orley Ashenfelter, a pioneering proponent of experiments and quasi-experimental research designs in social science. Here is Ashenfelter (1991) assessing the credibility of the observational studies linking schooling and income:

> How convincing is the evidence linking education and income? Here is my answer: Pretty convincing. If I had to bet on what an ideal experiment would indicate, I bet that it would show that better educated workers earn more.

The quasi-experimental study of class size by Angrist and Lavy (1999) illustrates the manner in which non-experimental data can be analyzed in an experimental spirit. The Angrist and Lavy study relies on the fact that in Israel, class size is capped at 40. Therefore, a child in a fifth grade cohort of 40 students ends up in a class of 40 while a child in fifth grade cohort of 41 students ends up in a class only half as large because the cohort is split. Since students in cohorts of size 40 and 41 are likely to be similar on other dimensions such as ability and family background, we can think of the difference between 40 and 41 students enrolled as being "as good as randomly assigned."

The Angrist-Lavy study compares students in grades with enrollments above and below the class-size cutoffs to construct well-controlled estimates of the effects of a sharp change in class size without the benefit of a real experiment. As in Tennessee STAR, the Angrist and Lavy (1999) results point to a strong link between class size and achievement. This is in marked contrast with naive analyses, also reported by Angrist and Lavy, based on simple comparisons between those enrolled in larger and smaller classes. These comparisons show students in smaller classes doing worse on standardized tests. The hospital allegory of selection bias would therefore seem to apply to the class-size question as well.[5]

2.3 Regression Analysis of Experiments

Regression is a useful tool for the study of causal questions, including the analysis of data from experiments. Suppose (for now) that the treatment effect is the same for everyone, say $\mathrm{Y}_{1i} - \mathrm{Y}_{0i} = \rho$, a constant. With

extremely well implemented randomized trial. Krueger's (1999) analysis suggests that none of these implementation problems affected the main conclusions of the study.

[5] The Angrist-Lavy (1999) results turn up again in Chapter 6, as an illustration of the quasi-experimental regression-discontinuity research design.

constant treatment effects, we can rewrite equation (2.1.1) in the form

$$\begin{array}{ccccccc} \mathrm{Y}_i = & \alpha & + & \rho & \mathrm{D}_i & + & \eta_i, \\ & \shortparallel & & \shortparallel & & & \shortparallel \\ & E(\mathrm{Y}_{0i}) & & (\mathrm{Y}_{1i} - \mathrm{Y}_{0i}) & & & \mathrm{Y}_{0i} - E(\mathrm{Y}_{0i}) \end{array} \tag{2.3.1}$$

where η_i is the random part of Y_{0i}. Evaluating the conditional expectation of this equation with treatment status switched off and on gives

$$\begin{aligned} E\left[\mathrm{Y}_i|\mathrm{D}_i = 1\right] &= \alpha + \rho + E\left[\eta_i|\mathrm{D}_i = 1\right] \\ E\left[\mathrm{Y}_i|\mathrm{D}_i = 0\right] &= \alpha + E\left[\eta_i|\mathrm{D}_i = 0\right], \end{aligned}$$

so that,

$$E[\mathrm{Y}_i|\mathrm{D}_i = 1] - E[\mathrm{Y}_i|\mathrm{D}_i = 0] = \underbrace{\rho}_{\text{treatment effect}} + \underbrace{E[\eta_i|\mathrm{D}_i = 1] - E[\eta_i|\mathrm{D}_i = 0]}_{\text{selection bias}}.$$

Thus, selection bias amounts to correlation between the regression error term, η_i, and the regressor, D_i. Since

$$E\left[\eta_i|\mathrm{D}_i = 1\right] - E\left[\eta_i|\mathrm{D}_i = 0\right] = E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_{0i}|\mathrm{D}_i = 0\right],$$

this correlation reflects the difference in (no-treatment) potential outcomes between those who get treated and those who don't. In the hospital allegory, those who were treated had poorer health outcomes in the no-treatment state, while in the Angrist and Lavy (1999) study, students in smaller classes tend to have intrinsically lower test scores.

In the STAR experiment, where D_i is randomly assigned, the selection term disappears, and a regression of Y_i on D_i estimates the causal effect of interest, ρ. The remainder of Table 2.2.2 shows different regression specifications, some of which include covariates other than the random assignment indicator, D_i. Covariates play two roles in regression analyses of experimental data. First, the STAR experimental design used conditional random assignment. In particular, assignment to classes of different sizes was random within schools, but not across schools. Students attending schools of different types (say, urban versus rural) were a bit more or less likely to be assigned to a small class. The comparison in column 1 of Table 2.2.2, which makes no adjustment for this, might therefore be contaminated by differences in achievement in schools of different types. To adjust for this, some of Krueger's regression models include school fixed effects, i.e., a separate intercept for each school in the STAR data. In practice, the consequences of adjusting for school

fixed effects is rather minor, but we wouldn't know this without taking a look. We will have more to say about regression models with fixed effects in Chapter 5.

The other controls in Krueger's table describe student characteristics such as race, age, and free lunch status. We saw before that these individual characteristics are balanced across class types, i.e. they are not systematically related to the class-size assignment of the student. If these controls, call them X_i, are uncorrelated with the treatment D_i, then they will not affect the estimate of ρ. In other words, estimates of ρ in the long regression,

$$\mathrm{Y}_i = \alpha + \rho \mathrm{D}_i + \mathrm{X}_i'\gamma + \eta_i \tag{2.3.2}$$

will be close to estimates of ρ in the short regression, (2.3.1). This is a point we expand on in Chapter 3.

Nevertheless, inclusion of the variables X_i may generate more precise estimates of the causal effect of interest. Notice that the standard error of the estimated treatment effects in column 3 is smaller than the corresponding standard error in column 2. Although the control variables, X_i, are uncorrelated with D_i, they have substantial explanatory power for Y_i. Including these control variables therefore reduces the residual variance, which in turn lowers the standard error of the regression estimates. Similarly, the standard errors of the estimates of ρ are reduced by the inclusion of school fixed effects because these too explain an important part of the variance in student performance. The last column adds teacher characteristics. Because teachers were randomly assigned to classes, and teacher characteristics appear to have little to do with student achievement in these data, both the estimated effect of small classes and it's standard error are unchanged by the addition of teacher variables.

Regression plays an exceptionally important role in empirical economic research. Some regressions are simply descriptive tools, as in much of the research on earnings inequality. As we've seen in this chapter, regression is well-suited to the analysis of experimental data. In some cases, regression can also be used to approximate experiments in the absence of random assignment. But before we can get into the important question of when a regression is likely to have a causal interpretation, it is useful to review a number of fundamental regression facts and properties. These facts and properties are reliably true for any regression, regardless of your purpose in running it.

Part II

The Core

Chapter 3

Making Regression Make Sense

> 'Let us think the unthinkable, let us do the undoable.
> Let us prepare to grapple with the ineffable itself,
> and see if we may not eff it after all.'
> Douglas Adams, *Dirk Gently's Holistic Detective Agency* (1990)

Angrist recounts:

> I ran my first regression in the summer of 1979 between my freshman and sophomore years as a student at Oberlin College. I was working as a research assistant for Allan Meltzer and Scott Richard, faculty members at Carnegie-Mellon University, near my house in Pittsburgh. I was still mostly interested in a career in special education, and had planned to go back to work as an orderly in a state mental hospital, my previous summer job. But Econ 101 had got me thinking, and I could also see that at the same wage rate, a research assistant's hours and working conditions were better than those of a hospital orderly. My research assistant duties included data collection and regression analysis, though I did not understand regression or even statistics at the time.
>
> The paper I was working on that summer (Meltzer and Richard, 1983), is an attempt to link the size of governments in democracies, measured as government expenditure over GDP, to income inequality. Most income distributions have a long right tail, which means that average income tends to be way above the median. When inequality grows, more voters find themselves with below-average incomes. Annoyed by this, those with incomes between the median and the average may join those with incomes below the median in voting for fiscal policies which - following Robin Hood - take from the rich and give to the poor. The size of government consequently increases.
>
> I absorbed the basic theory behind the Meltzer and Richards project, though I didn't find it