ABSTRACT

Decentralizing Equality of Opportunity and Issues Concerning the Equality of Educational Opportunity

Caterina Calsamiglia

2005

This dissertation explores issues concerning local versus global equality of opportunity, the presence of inequality for education attainment and the optimality of policies addressing such problems. The literature distinguishes between two different approaches to distributive justice: the global and the local approach. Local distributive justice analysis determines the allocation of rights, duties and resources concerning a subset of the aspects that determine individual welfare. On the other hand, global distributive justice analysis determines the allocation of rights, duties and resources concerning all aspects in a society that affect individuals welfare.

Global and local distributive justice notions have been analyzed independently in the literature and the focus has been more on global rather than local distributive justice. In the first chapter we interpret a collection of local distributive justice problems as decentralized global distributive justice problems. This permits global distributive justice analysis, mainly done by philosophers, to inform policymakers on what the local distributive justice goal should be, and also allows us to understand what the global distributive justice properties of a collection of local problems are. We then analyze the properties on local distributive justice that are necessary and sufficient to decentralize the global distributive justice notion of equality of opportunity. The second chapter provides evidence that school could potentially serve as an instrument to provide equality of educational opportunity. Using data from the UK, we find that the education process affects individuals final educational achievement, contrary to what part of the literature suggests. The last chapter is somewhat unrelated to

the earlier chapters. We provide necessary and sufficient conditions on demand data for the individual to behave as if she was maximizing quasilinear preferences subject to a budget constraint. In the case where such property fails, we provide the quasilinear utility that "better" approximates the data.

Decentralizing Equality of Opportunity and Issues Concerning the Equality of Educational Opportunity

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Caterina Calsamiglia

Dissertation Director: Professor Stephen Morris

May 2005

Contents

List of Tables

List of Figures

Acknowledgments

To Angel, a true believer, to whom I am forever indebted.

To my parents, Xavier and Mariona, for being unconditionally and wisely supportive.

To Maia, for being the love of our lives.

I want to thank Stephen Morris, John E. Roemer, Hanming Fang and Donald J. Brown for valuable advice. I also thank Dirk Bergemann, Xavier Calsamiglia, Stephen Coate, Gil Kalai, Humberto Llavador, Efe Ok, Ben Polak, Ariel Rubinstein, Andrew Lemon and seminar participants at Yale University for helpful comments and encouragement. Finally, I am grateful to Yale University and the Department of Economics, and especially to Pam o'Donnell for providing a wonderful environment for us to grow emotionally and intellectually.

Chapter 1

Introduction

This dissertation mainly explores issues concerning local versus global equality of opportunity and the presence of inequality for education attainment. The literature distinguishes between two different approaches to distributive justice: the global and the local approach. Local distributive justice analysis determines the allocation of rights, duties and resources concerning a subset of the aspects that determine individual welfare. On the other hand, global distributive justice analysis determines the allocation of rights, duties and resources concerning all aspects in a society that affect individuals welfare.

Global and local distributive justice notions have been analyzed independently in the literature and the focus has been more on global rather than local distributive justice. In the second chapter we interpret a collection of local distributive justice problems as decentralized global distributive justice problems. This permits global distributive justice analysis, mainly done by philosophers, to inform policymakers on what the local distributive justice goal should be, and also allows us to understand what the global distributive justice properties of a collection of local problems are. We then analyze the properties on local distributive justice that are necessary and sufficient to decentralize the global distributive justice notion of equality of opportunity. The third chapter provides evidence that school could potentially serve as an instrument to provide equality of educational opportunity. Using data from the UK, we find that the education process affects individuals final educational achievement, that is, that school matters, contrary to what part of the literature suggests. The last chapter is somewhat unrelated to the earlier chapters. We provide necessary and sufficient conditions on demand data for the individual to behave as if she was maximizing quasilinear preferences subject to a budget constraint. In the case where such property fails, we provide the quasilinear utility that better approximates the data.

The second chapter is entitled *Decentralizing Equality of Opportunity.* Equality of opportunity is an important welfare criterion in political philosophy and policy analysis. Philosophers define equality of opportunity as the requirement that an individual's well being be independent of his or her irrelevant characteristics. The difference among philosophers is mainly about which characteristics should be considered irrelevant. Policymakers, however, are often called upon to address more specific questions: how should admissions policies be designed so as to provide equal opportunities for college?' or how should tax schemes be designed so as to equalize opportunities for income? These are called local distributive justice problems, because each policymaker is in charge of achieving equality of opportunity in a specific issue. In the literature global and local distributive justice problems have been analyzed independently. We interpret a collection of local distributive justice problems as the decentralized global distributive justice problem. We show that the provision of equality of opportunity, which is not concerned about effort directly, is decentralized if the collection of local mechanisms provides local equality of rewards to effort, that is, equality of rewards to productive individual choices. Moreover, if the mechanism and the environment are rich, a collection of local mechanisms decentralizes equality of opportunity only if it provides local equality of rewards to effort.

The third chapter, *Is Education Attainment Determined at Age 7?*, provides empirical evidence that suggests that the education process of an individual may affect his final school attainment, that is, the number of years in school do affect the individuals achievement. This is opposed to the hypothesis that characteristics present at very early stages, such as family background or IQ, are the only determinants of the individuals later success in school. This work uses data from the National Child Development Survey (NCDS) of an entire cohort born in Britain in one week in 1958. We question the fact that test scores at 7 being correlated with test scores at 16 implies that the education process between 7 and 16 does not affect his final attainment. We find that including age 11 scores in the analysis eliminates the significance of age 7 scores, suggesting that the process between age 7 and age 11 does matter for the individuals final educational achievement.

The last chapter is entitled *Rationalizing and Curve-fitting Demand Data with Quasilinear Utilities*. In the empirical and theoretical microeconomic literature a consumer's utility function is often assumed to be quasilinear. In this paper we provide necessary and sufficient conditions for testing through demand data if the consumer acts as if she is maximizing a quasilinear utility function over her budget set. If the consumer's choices are inconsistent with maximizing a quasilinear utility function over her budget set, then we compute the best quasilinear rationalization of her choices.

Chapter 2

Decentralizing Equality of Opportunity

2.1 Introduction

Equality of opportunity occupies a prominent place in moral philosophy. For philosophers the central problem is to define what is meant by a just social order, i.e., the proper distribution of resources, rights, and duties in society at large. This is the global distributive justice problem. Here, equality of opportunity requires that an individual's success or welfare in life be independent of *irrelevant* characteristics, that is, of characteristics that the individual should not be responsible for. Such irrelevant characteristics may include, for example, parents' income. Arneson, Cohen, Dworkin and other political philosophers address this question.[1] The wide variety of interpretations that we see of equality of opportunity corresponds to a wide spectrum of beliefs regarding irrelevant characteristics. For example, egalitarians claim that all characteristics are irrelevant and that therefore welfare should be

[1] See Roemer (1998) for a discussion.

equalized across individuals. Libertarians, however, argue that the set of irrelevant characteristics is the set of non-productive characteristics, such as race, i.e., one's success in life should be independent of one's race.

In contrast to the moral philosophers' abstract problem of constructing a just social order, institutions and governments must decide what is meant by a just allocation of rights and resources for concrete problems. For example, companies must set the wages for their employees; schools have to decide what applicants to admit; and public agencies must determine who gets into a nursery school or the distribution of the tax burden among individuals in the society. When addressing these issues they face a local distributive justice problem.[2]

Although global and local distributive justice problems seem to be related, a gap exists between them. In the first chapter of his book *Equity*, Peyton Young describes this disconnect: "[t]heories of justice in the large do not tell us how to solve concrete, everyday distributive problems such as who should get into medical school or how much to charge for a subway ride." He argues that the reason for this is that "[i]ssues of local justice tend to be compartmentalized. Societies make no effort to coordinate distributive decisions across different domains." For example, kidney agencies do not give precedence to patients who failed to get into university and a worker's wage at a given job is not increased because the government located a waste dump near his home. He concludes by saying that "[T]here is no mechanism comparable to the invisible hand of the market for coordinating distributive justice at the micro into just outcomes at the macro level." He then argues that local

[2] Elster (1992) defines local justice as the study of how institutions allocate scarce resources.

distributive justice should be derived by inducing general rules from what we observe in different local distributive justice problems.

In this paper we interpret local distributive justice problems as essentially the decentralized global distributive justice problem. Therefore the question of interest is whether we can solve the global distributive justice problem by solving a series of local distributive justice problems. In particular, we characterize the set of local distributive justice rules that attain global equality of opportunity as defined by political philosophers.

When decentralizing equality of opportunity, a problem that arises is that of properly accounting for the cost of local effort, because it generally depends on information that the local policymaker does not have. For example, colleges may be interested in equalizing the probability of admission into college for students that only differ in their neighborhood characteristics. But if different individuals have invested a different number of hours of study, colleges want to compensate for the individual's cost of studying the extra hours by admitting them with a higher probability, even if college admissions do not intrinsically care about effort. The main problem in doing that will be that the cost of the extra hours of studying is unknown to the policymaker if it depends on how much effort or time the individual is spending on the other activities available to him, like training basketball or working in a paid job. If the cost of effort in school depends on the hours the individual is investing on the other activities, then the college, lacking knowledge on what the efforts are in the other activities, will not be able to compensate for the cost of effort. In this case, if the college wants to contribute to the provision of equality of opportunity, he may have to equalize the probability of being admitted for individuals doing the same effort, that is, equalize the probability of being admitted as a function of the hours studied.

The existing literature has not explicitly addressed the problem of the dispersion of information and interrelation of local environments just discussed. The approach has been to implicitly assume independence of local environments, that is, to assume separability in the utility individuals get from the different aspects determining their welfare, or to ignore the cost of effort. Then the analogous notion of equality of opportunity for the global problem is applied to the local problem. Specifically, local equality of opportunity has been interpreted as local outcomes or local utilities being independent of a set of irrelevant characteristics that are affecting the local outcome. For example, Benabou and Ok (2001) take the income distribution and measure inequality of opportunity by the differences between the expected distribution of individual income conditional on parental income, where parental income is taken to be the irrelevant characteristic determining individual income. They attribute the variance in the income distribution of individuals with similar parental income to relevant characteristics and measure inequality as the difference between those distributions. By measuring inequality by inequality of income they are ignoring the different cost of effort that may be imposed to individuals exerting different efforts to get a given income. Another example is Betts and Roemer (2002). They analyze the investment in education required to equalize the wage distributions across blacks and whites where race is considered the irrelevant characteristic.[3] We show that such notion of local equality of opportunity may not lead to global equality of opportunity.[4] Specifically, local independence of outcomes from

[3]See more examples in Llavador and Roemer (2001) or Roemer et al (2003).

[4]Cavanagh (2002) in the introduction to his book *Against Equality of Opportunity* criticizes equality of opportunity because of the apparent inconsistency when applied to the global and the local problem. He says that " [E]ven those of us who realize that these are two quite different ways of approaching the issue

irrelevant characteristics will not aggregate into independence of total welfare from irrelevant characteristics. Therefore, local independence of outcomes from irrelevant characteristics fails to provide global equality of opportunities. This paper shows that a collection of local mechanisms will decentralize the provision of equality of opportunity if it provides local equality of rewards to effort, where effort is a productive individual choice variable, for example time devoted to a given activity.[5] Moreover, we show that if the mechanisms and the environment are "rich", notion that will be made precise later in the paper, global equality of opportunity will be decentralized only if the collection of mechanisms provides local equality of rewards to effort. Thus, when measuring inequality, one has to also control for the effort that individuals have invested in obtaining the outcome.[6] Furthermore, when designing policies one should analyze the effect of the policy on the rewards to effort and not only on the outcome it will generate.

---

nevertheless tend to think it doesn't matter. We assume that we will end up in the same place whichever end we start—that is whether we start at the micro level, thinking about how each individual decision should be made, or whether we start at the macro level, thinking about what we want the resulting pattern to look like. But this is not quite right." He then presents examples where the general notion of equality of opportunity applied to the local or global problem leads to different policy recommendations. This leads him to conclude that equality of opportunity is not an adequate distributive justice notion.

[5] Roemer (1998) proposes an algorithm that equalizes outcomes for individuals with different irrelevant characteristics that exert the same *degree* of effort. In order to define the *degree* of effort, he partitions the set of individuals into a set of types defined by irrelevant characteristics. Individuals within a type have the same irrelevant characteristics. He then argues that the effort distribution for each type is just a result of the irrelevant characteristic defining the type, and therefore that individuals are only responsible for their position in the distribution of effort within his type. The *degree* of effort is then defined as the *percentile* in which the individual is located in the effort distribution *within* his type. This implies that he does not want to reward effort per se, but the relevant characteristics which are causing individuals to locate themselves in a given position in the effort distribution from his type, that is the relevant characteristics that are driving the individual to do more or less effort relatively to the people with the same irrelevant characteristics. His algorithm is therefore designed to reward the set of exogenous relevant characteristics, even though calling it degree of effort may lead you to think otherwise. In his analysis he does not distinguish between local and global equality of opportunity.

[6] In Benabou and Ok (2001), for example, this would involve controlling for labor hours spent to earn a given income and use the difference between rewards to labor hours as a measure of inequality of opportunity.

The local rule requires that outcomes be independent of irrelevant characteristics but that it does so through making rewards to effort independent of irrelevant characteristics. Implementing this may be hard because effort is usually unobservable. Therefore some may interpret this as a critique to the equal opportunities enterprise since we present a more difficult problem. Others, however, may interpret this result as a step toward understanding what the provision of global equal opportunities in society entails and view the local rule as a useful tool for policy design.

The structure of the remainder of the paper is as follows. Section 2 presents an example to illustrate the effects that drive the results. Section 3 presents the benchmark model and derives the main results. Section 4 extends the results to the case of heterogenous preferences. Section 5 presents some concluding remarks.

2.2 Local Equality of Opportunity

Education, health, sports or reading, among other things, generally contribute to one's welfare. The welfare derived from any one of these activities or attributes typically depends on an individual's characteristics (e.g., preferences, skills, or family characteristics) and the time and effort one devoted to them. We distinguish between two kinds of characteristics an individual has: (1) relevant characteristics which social debate has concluded the individual should be responsible for, such as an innate ability or preference for an activity and (2) irrelevant characteristics which social debate has concluded individual should not be responsible for, such as parents' income and education. In this framework, equality of opportunity requires that an individual's ultimate welfare be independent of his or her irrelevant characteristics. Specifically, we want two individuals with identical relevant characteristics to achieve the same welfare regardless of their irrelevant characteristics.

To accomplish this goal, a central planner knowing the joint distribution of relevant and irrelevant characteristics could compensate each individual for his advantages and disadvantages due to irrelevant characteristics. In reality we observe independent institutions providing "equality of opportunity" for specific issues by compensating individuals for their irrelevant characteristics affecting a specific activity, and ignoring other aspects that do not directly affect it. For example, colleges focus on compensating for family education and income, but ignore differentials in parents' success in sports. Also, health care policies compensate for family income and family health conditions but not parents' education. Thus equality of opportunity is provided in a decentralized manner.

The following example presents the main difficulties involving the decentralized approach to equality of opportunity and the effects driving the results in this paper.

2.2.1 Education and Basketball example

Paul and Richard are intelligent, highly motivated and hardworking teenagers, denoted by $i \in \{P, R\}$. Both would like to become professional basketball players and get a good general education. Paul and Richard are identical except that Richard's neighborhood has good public schools and a local library. In Paul's neighborhood there are poor public schools and no local library. Unlike education, Paul and Richard have identical resources to improve their basketball skills.

Paul and Richard have to decide how much time to spend on homework, denoted by e_S, and on playing basketball, denoted by e_B. Their ultimate objective is to get into college and the NBA.

Social debate has determined that some individual characteristics should be relevant in determining individual's welfare and some should not. The individual is characterized by a vector of relevant characteristics $r = (r_S, r_B)$ and a vector of irrelevant characteristics, $b = (b_S, b_B)$.

For this example we assume that society has determined that the individual relevant characteristics are education ability and basketball ability, and the irrelevant characteristics are neighborhood school resources and basketball facilities. We assume $(b_S^P, b_B^P) = (\frac{1}{2}, 1)$ and $(b_S^R, b_B^R) = (\frac{7}{10}, 1)$, and $(r_S^P, r_B^P) = (r_S^R, r_B^R) = (\frac{1}{3}, \frac{1}{5})$.

College admissions observe education-specific characteristics, r_S and b_S, and NBA recruiters observe basketball-specific characteristics, r_B and b_B.

Out of the characteristics and effort, Paul and Richard get a performance in education and basketball that can be measured by the outcome functions "normalized SAT score" and "normalized NBA test score". The normalized SAT score (percentile in the distribution of SAT scores in the population) can be represented by:

$$a_S = a_S(b_S, r_S, e_S) = b_S r_S (e_S)^{\frac{1}{2}}$$

College admissions therefore know that $a_S = \frac{1}{3} b_S (e_S)^{\frac{1}{2}}$. The normalized NBA test score (percentile in the distribution of NBA scores in the population) can be represented by:

$$a_B = a_B(b_B, r_B, e_B) = b_B r_B (e_B)^{\frac{1}{2}}$$

NBA recruiters therefore know that $a_B = \frac{1}{5}(e_B)^{\frac{1}{2}}$. The cost of training and studying to

Paul and Richard is represented by:

$$c(e_S, e_B) = \frac{1}{500}(e_S + e_B)^2$$

College admissions officers and NBA recruiters, with education-specific and basketball-specific information respectively, decide the probability of individuals being admitted into college and into the NBA, that is $p_S(b_S, r_S, e_S)$ and $p_B(b_B, r_B, e_B)$.[7] Paul and Richard's utility is determined by the probability of them being admitted into college or the NBA and by the cost of devoting time to either of the two activities:

$$U = p_S(b_S, r_S, e_S) + p_B(b_B, r_B, e_B) - c(e_S, e_B)$$

We say that college and NBA recruiters follow a *market policy* if they admit individuals with probability $p_j = a_j$. In that case Paul and Richard solve:

$$\max_{e_S, e_B} \frac{1}{3} b_S (e_S)^{\frac{1}{2}} + \frac{1}{5}(e_B)^{\frac{1}{2}} - \frac{1}{500}(e_S + e_B)^2$$

Paul will choose $(e_S^P, e_B^P) = (4.2, 6)$ and get $p_S^P = 0.34$, $p_B^P = 0.5$ and $U^P = 0.62$; and Richard will choose $(e_S^R, e_B^R) = (6.6, 4.8)$ and get $p_S^R = 0.6$, $p_B^R = 0.44$ and $U^R = 0.77$. As represented in Figure 1, Richard's productivity in school is higher than that of Paul, $\frac{\partial p_S^R}{\partial e_S} > \frac{\partial p_S^P}{\partial e_S}$, but his productivity is the same as Paul's in basketball, $\frac{\partial p_B^R}{\partial e_B} = \frac{\partial p_B^P}{\partial e_B}$. Their choices are determined by the points where marginal cost equals marginal increase in the probability of admission, that is, the points where the cost functions are tangent to the probability of

[7] In this example we are not modelling and admission policy that is allocating a fixed number of seats, but assuming that the admissions policy for Paul can be changed without Richard's being affected, which is somewhat unrealistic, but greatly simplifies presenting the point of the example.

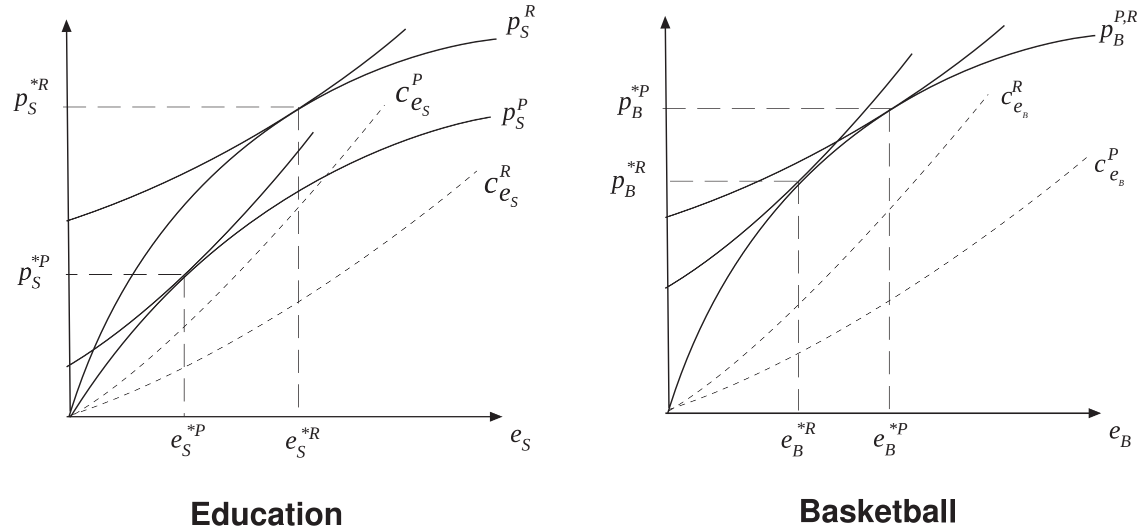

Figure 2.1: Choices and outcomes when college and NBA recruiters follow the market policy

admission . The marginal cost in one activity depends on the effort being spent at the other activity. Specifically, the more time one spends in one activity, the higher the marginal cost of effort for the other activity. Richard devotes relatively more time to school than Paul, given his higher productivity of effort, which leaves Richard with a higher marginal cost of playing basketball, leading him to spend relatively less time in basketball. Notice that equality of opportunity is violated because Richard has a higher welfare than Paul even though they only differ in irrelevant characteristics.

Equality of opportunity is a concern when irrelevant characteristics affecting individual's achievement differ. Notice that since irrelevant characteristics affecting basketball are the same for Paul and Richard, NBA recruiters need not worry about violating equality of opportunity. For them the market policy, for example, will satisfy equality of opportunity.

Suppose that college admissions had as a first priority the contribution to the provision of equality of opportunity through providing local equality of opportunity or equality of educational opportunity. It is not clear a priori what equality of educational opportunity would require. One possibility is that a consultant, having read some political philosophy or Roemer's book on Equality of Opportunity, tells them that since Paul and Richard are identical except for their irrelevant characteristic their goal should be to equalize the chances that they get admitted into college. This will lead college admissions not to evaluate Paul and Richard on their SAT scores alone, since Paul is getting lower SAT scores because of irrelevant characteristics. College admissions will then design an admissions policy that will ensure Paul and Richard being admitted with equal probability. There will generally be a large set of policies satisfying this property, for example admitting Paul and Richard conditional on their relevant characteristics alone, ignoring the SAT scores, which would not reward hours of study. College admissions could choose any of the policies in the set if that was its only concern. I want to point out two examples of policies that are often discussed and even applied in the political debate, that are in this set but do not have the property of providing global equality of opportunity, that is, that will not provide the same welfare for Paul and Richard. This will illustrate that the provision of local equality of opportunity will generally not lead to global equality of opportunity.

Example 1: equalizing local outcomes through Affirmative Action

Colleges could set the admissions probability according to a modified SAT score, $p_S = a_S + T(b_S)$, to evaluate the different applicants. The transfers $T(b_S)$ are such that the two individuals get the same modified SAT score. In particular it could give a positive transfer of $T(b_S^P) = a_S^R - a_S^P = 0.26$ to Paul and a zero transfer to Richard.[8] As seen in Figure 2, $p_S'^P$ corresponds to p_S^P shifted up until $p_S^{*P} = p_S^{*R}$. Given college admissions policy Paul and Richard will not change their hours devoted to education and basketball because of the lump sum nature of the transfer, i.e., the rewards to time into either activity are left unchanged. We observe that Paul is better off than Richard under this policy. Numerically their welfare is: $U^P = 0.88$ and $U^R = 0.77$.

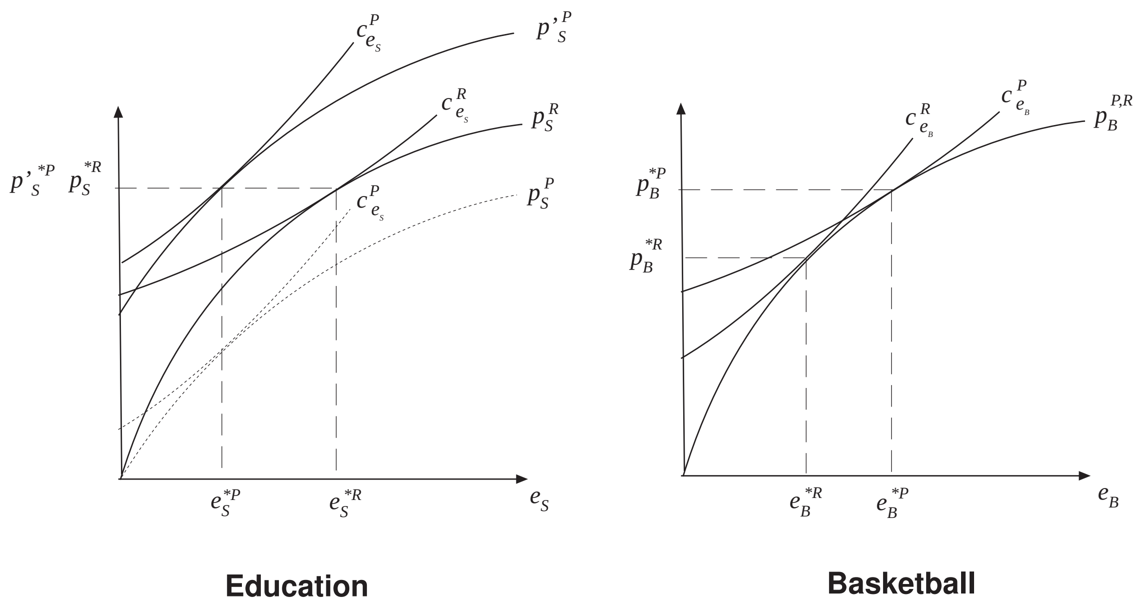

Figure 2.2: Effect of Affirmative Action consisting of adding a constant number of points to SAT scores

College admissions accept both Paul and Richard with equal probabilities. NBA recruiters accept Paul with higher probability than Richard because his performance is better—he chooses to devote more time to basketball.[9]

[8]Affirmative Action policies have been of this form in some universities. Disadvantaged minorities were given a fixed number of extra points in their evaluation procedure.

[9]Notice that we did not consider Paul sophisticated enough to know that the college admissions were equalizing outcomes. If that were the case Paul would have spent NO time in school because he'd know that would not affect his education degree relatively to Richard's.

This intervention gives Paul admission to college at a lower "price" in terms of effort than to Richard. Richard devotes more time to it and is equally rewarded with the same chances to get admitted. Paul can then spend more time training basketball at a lower marginal cost. Global equality of opportunity is violated since Paul's utility is higher than Richard's. Now Richard is disadvantaged because of neighborhood educational resources!

Example 2: equalizing local outcomes "paternalistically"

College admissions can paternalistically force Paul to study long enough (thorough having the teacher lock him up in school) so that he gets the same SAT score as Richard. That is, force Paul to study a certain amount of hours, $e_S^{\prime P}$, so that $a_S(\frac{1}{3}, e_S^{\prime P}, 0.5) = a_S(\frac{1}{3}, e_S^R, 0.7)$. If then the market policy is implemented the probability of being admitted is equalized. Figure 3 shows that Paul is now forced to study long enough so that he gets the same SAT score as Richard. Paul has to study longer hours, $e_S^{\prime P} = 13$, and therefore faces a larger marginal cost for training basketball that leads him to choose to train few hours, $e_B^{\prime P} = 2.6$. This lowers his utility to $U^P = 0.44$. Paul and Richard are equally likely to get into college, but Richard has higher chances to enter the NBA. Richard gets a utility of $U^R = 0.77$. He is now not only worse than Richard, but worse than himself under no intervention! Paul gets the same education outcome but by putting in a larger effort than Richard. He is, therefore, left with a higher marginal cost for basketball training. Here we also fail to provide global equality of opportunity since Paul is getting an even lower utility than Richard only because of neighborhood educational characteristics.

These two examples illustrate that independence of local outcomes to irrelevant characteristics does not guarantee global equality of opportunity, and that it may generate new

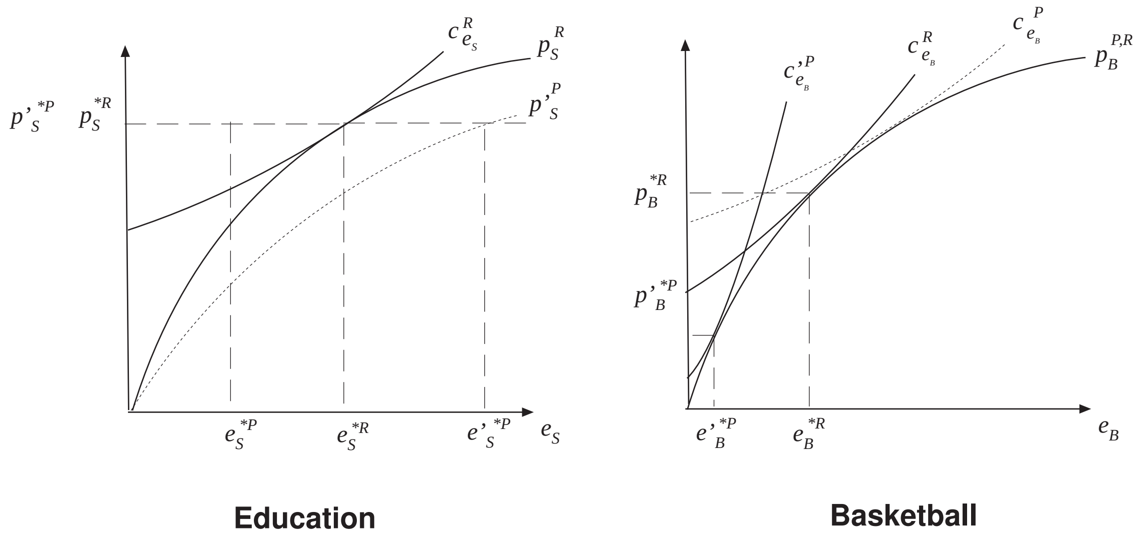

Figure 2.3: Effect of policy that forces Paul to stay in school long enough to get the same educational outcome as Richard

and stronger inequalities. Recall that for global equality of opportunity we need that Paul and Richard's welfare be independent of irrelevant characteristics, i.e., that their utility is equal since they only differ in irrelevant characteristics. In the first example Richard, originally advantaged, ends up with a lower utility than Paul. In the second example, Paul, originally disadvantaged, is made worse off.

This paper shows that the rule that decentralizes equality of opportunity is one that equalizes rewards to effort or time, that is, that it requires the same "price" in terms of effort for each unit of local outcome. In this case, the goal is that for every extra hour of studying, Paul and Richard get the same increase in their probability of being admitted. The following is an example of a policy that would accomplish such goal.

Example 3: equalizing rewards to effort

College admissions, understanding the bias generated by the neighborhood education not only on outcome but on Paul's choice of effort, bases his admissions decisions on the following modified SAT score: $a'_S(b_S) = T(b_S) \cdot a_S$ and sets $p_S = a'_S$. This would imply increasing the the probability of being admitted proportionally to the SAT achieved. In this case $T(b_S^R) = 1$ and $T(b_S^P) = \frac{b_S^P}{b_S^R} = \frac{7}{5}$ would make the marginal increase in the probability of being admitted for one more hour studied equal for Paul and Richard. As Figure 4 shows, under this policy Paul and Richard's choice sets are equalized, that is, the probabilities of entering college and the NBA for any chosen level of effort are independent of neighborhood education. This leaves Paul and Richard with the exact same choice problem:

$$\max_{e_S^R, e_B^R} \; T(b_S^R) \cdot \frac{1}{3} b_S^R e_S^{R\,1/2} + \frac{1}{5} e_B^{1/2} - \frac{1}{500}(e_S + e_B)^2 \;=\; \frac{7}{10} e_S^{1/2} + \frac{1}{5} e_B^{1/2} - \frac{1}{500}(e_S + e_B)^2$$

$$\max_{e_S^P, e_B^P} \; T(b_S^P) \cdot \frac{1}{3} b_S^P e_S^{P\,1/2} + e_B^{1/2} - \frac{1}{500}(e_S + e_B)^2 \;=\; \frac{7}{5}\frac{5}{10} \cdot e_S^{1/2} + \frac{1}{5} e_B^{1/2} - \frac{1}{500}(e_S + e_B)^2$$

The problems Paul and Richard face, the tradeoffs and payoffs from studying and training are exactly the same, and therefore their welfare will be equal. Paul and Richard would choose to study the same number of hours, $e_S^P = e_S^R = 6.6$, and to train the same number of hours, $e_B^P = e_B^R = 4.8$, getting the same chances to enter college and the NBA, $p_S^P = p_S^R = 0.6$ and $p_B^P = p_B^R = 0.44$ and getting equal utilities $U^P = U^R = 0.77$.

This example shows that equalizing rewards to effort will provide equality of opportunity, where effort is the hours studied and the hours trained. The following section formalizes and generalizes the result and shows not only that this criterion is guaranteed to decentralize the

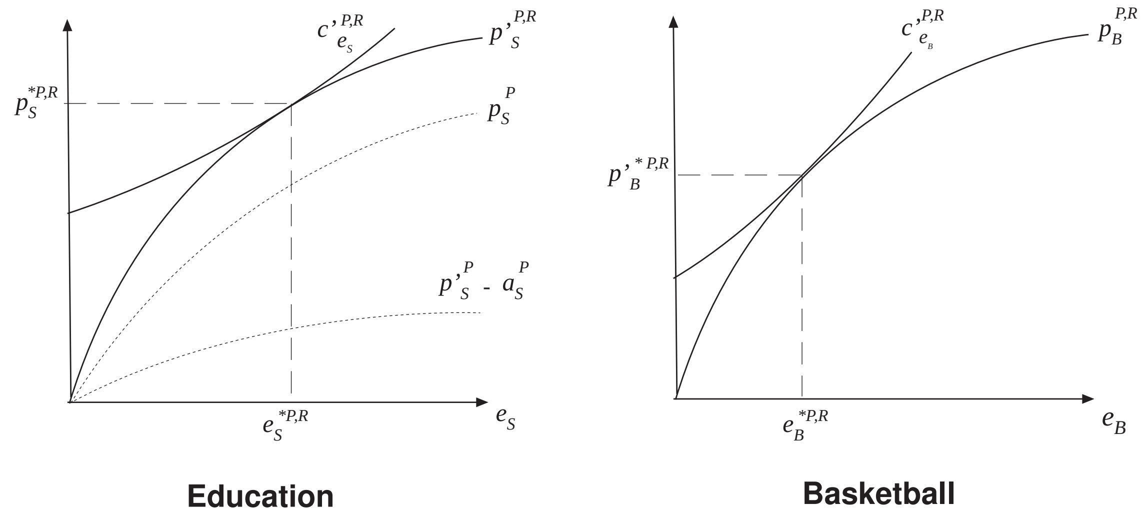

Figure 2.4: Effect of Affirmative Action consisting of adding a proportion of extra points to SAT scores

provision of equality of opportunities, but that if the mechanisms and the environment are "rich", a collection of local mechanisms decentralizes the provision of equality of opportunity only if it locally equalizes rewards to effort.

2.3 Equality of Opportunity and Effort

There is a continuum of individuals in an interval I and two utility-providing activities, $j \in \{1, 2\}$. Each individual $i \in I$ is characterized by two vectors of exogenous characteristics: *irrelevant characteristics*, $b = (b_1, b_2) \in I\!R^2$, which represent the characteristics that society has decided should not affect individual well-being; and *relevant characteristics*, $r = (r_1, r_2) \in I\!R^2$, which represent the characteristics that society has decided may affect individual well-being. Individuals choose the *effort* or time they devote to each of the activities, $e = (e_1, e_2) \in I\!R^2$. In the "Education and basketball" (EB) example, the vector of irrelevant characteristics contains neighborhood education and basketball facilities, (b_S, b_B);

the vector of relevant characteristics contains the education and basketball ability, (r_S, r_B); and the choice variables were hours studied and trained, (e_S, e_B). Individual characteristics are described by the joint distribution $F : I\!R^2 \times I\!R^2 \to I\!R$. Given the choice of effort of the individuals we can describe the set of individuals by the joint distribution, $G \in \Omega$, $G : I\!R^2 \times I\!R^2 \times I\!R^2 \to I\!R$, where $G(\hat{b}, \hat{r}, \hat{e}) = \Pr(\{(b,r)|b \leq \hat{b}, r \leq \hat{r} \text{ and } e \leq \hat{e}\})$. Similarly, $G_1 : I\!R \times I\!R \times I\!R \to I\!R$ and $G_2 : I\!R \times I\!R \times I\!R \to I\!R$ are the marginal distributions of characteristics and effort choices for each of the activities. That is, $G_1(\hat{b_1}, \hat{r_1}, \hat{e_1}) = \Pr(\{(b_1, r_1)|b_1 \leq \hat{b_1}, r_1 \leq \hat{r_1} \text{ and } e_1 \leq \hat{e_1}\})$ and $G_2(\hat{b_2}, \hat{r_2}, \hat{e_2}) = \Pr(\{(b_2, r_2)|b_2 \leq \hat{b_2}, r_2 \leq \hat{r_2} \text{ and } e_2 \leq \hat{e_2}\})$.

Outcome produced in the two activities depends on the distribution of relevant and irrelevant characteristics and the choices of effort. The outcome production function of individuals with characteristics and effort (b, r, e) is represented by a vector of strictly increasing functions:

$$a(b, r, e; G) = (a_1(b_1, r_1, e_1; G_1), a_2(b_2, r_2, e_2; G_2)) \in I\!R^2$$

where $a_j : I\!R \times I\!R \times I\!R \times \Omega_j \to I\!R$ represents the outcome produced by individual (b, r, e) in activity j when living in society with characteristics and efforts G_j. This allows for externalities between individual's productivity, but they are not required for the results presented to hold. In the EB example the outcome production functions where the normalized SAT score and the normalized NBA test score, $(a_S(b_S, r_S, e_S), a_B(b_B, r_B, e_B))$.

Individual's utility depends on the outcome consumed for each activity and the effort they chose to exert in each activity:

$$U(x, e) = u(x) - c(e)$$

where u is a strictly increasing function representing the utility out of the outcome vector $x = (x_1, x_2)$ and c is a *non separable* function in e_1 and e_2 that represents the cost of the effort vector, $e = (e_1, e_2)$.

We now introduce concepts that will allow us to formally distinguish between global and local distributive justice problems.

Definition 1: A *social mechanism* μ is a mapping that assigns for every distribution $G \in \Omega$ an outcome vector to each individual:

$$\mu : I\!R^2 \times I\!R^2 \times I\!R^2 \times \Omega \rightarrow I\!R \times I\!R$$

$$\mu \equiv (\mu_1, \mu_2) : (b, r, e; G) \rightarrow (x_1, x_2)$$

$$s.t. \int_{\boldsymbol{R}^2 \times \boldsymbol{R}^2 \times \boldsymbol{R}^2} \mu_j(b, r, e; G)\ dG \leq \int_{\boldsymbol{R}^2 \times \boldsymbol{R}^2 \times \boldsymbol{R}^2} a_j(b_j, r_j, e_j; G_j)\ dG \quad \text{for } j = 1, 2$$

That is, x_j is the outcome that the individual with characteristics (b, r) and effort e gets out of activity j when the distribution of individual characteristics and efforts is $G(b, r, e)$.[10] The social mechanism in a market economy, for example, assigns to each individual the outcome production function, $\mu(r, b, e; G) = a(r, b, e; G)$. Each individual gets what he produces.

Given a social mechanism an individual's effort vector will maximize his utility, given the efforts of others:

$$e^* = e^*(b, r; G^*) \equiv \arg\max_{e \in \boldsymbol{R}^2} u(\mu(b, r, e; G^*) - c(e)$$

[10]Note that individuals are fully characterized by their characteristics and choice of effort and that the mechanism is implicitly assumed to be symmetric, i.e., it assigns the same outcomes to individuals with the same characteristics and effort choices, given a distribution G.

Let $G^*(b, r, e^*) = G^*$ represent the distribution of characteristics and efforts in equilibrium.[11] The individual's utility in equilibrium is denoted by $u^*(b, r; G^*) = u(\mu(b, r, e^*; G^*)) - c(e^*)$. Let (b^i, r^i, e^i) denote the characteristics of individual $i \in I$.

Definition 2: A social mechanism μ provides *equality of opportunity* if for all G^*:

$$\forall i, k \in I, \;\; r^i = r^k \;\Rightarrow\; u^*(b^i, r^i; G^*) = u^*(b^k, r^k; G^*)$$

A social mechanism provides equality of opportunity if the welfare achieved by each individual varies with his relevant characteristics but not with his irrelevant characteristics. Notice that in this definition effort is not directly relevant. Political philosophers overlook the role of effort because they are only concerned about the indirect utility, which is ultimately a function of exogenous characteristics.

We have now defined the goal of equality of opportunity for the global distributive justice problem. If one policymaker had all information he could act as a social planner and guarantee global equality of opportunity. We are interested in solving the decentralized problem in which the information specific to each activity is known by a policymaker but no policymaker has all the information. We will therefore assume that there is one policymaker in charge of each activity. In the EB example the college admissions knew only the education-specific information while NBA recruiters knew only the basketball-specific information. Again, let G_1 and G_2 be the marginal distribution of G with respect to (b_1, r_1, e_1) and (b_2, r_2, e_2) respectively, which represent the information that each policy maker has.

[11] In the case where there is more than one equilibrium we assume the planner may choose one for individuals to play.

Definition 3: A *local mechanism* for policymaker j is a mapping ϕ_j that assigns to every marginal distribution G_j a local outcome vector to each individual:

$$\phi_j : I\!R \times I\!R \times I\!R \times \Omega_j \to I\!R$$

$$\phi_j : (b_j, r_j, e_j; G_j) \to x_j$$

$$\text{s.t.} \int_{\boldsymbol{R}\times\boldsymbol{R}\times\boldsymbol{R}} \phi_j(b_j, r_j, e_j; G_j)\, dG_j \leq \int_{\boldsymbol{R}\times\boldsymbol{R}\times\boldsymbol{R}} a_j(b_j, r_j, e_j; G_j)\, dG_j$$

A local mechanism assigns to every individual and outcome for a specific activity conditional on his characteristics and choices specific to that activity. In the EB example, the college admissions process was a local mechanism assigning a probability of being admitted into college conditional on the individual education-specific characteristics.

A pair of local mechanisms generates a social mechanism such that $\mu = (\mu_1, \mu_2)(b, r, e; G) = (\phi_1(b_1, r_1, e_1; G_1), \phi_2(b_2, r_2, e_2; G_2))$. Instead of the mapping μ assigning an outcome vector to each individual we have (ϕ_1, ϕ_2), with ϕ_j assigning the outcome for activity j.

Definition 4: A pair of local mechanism, $\{\phi_1, \phi_2\}$, provides *global equality of opportunity* if for all G^*, the social mechanism $\mu = (\phi_1, \phi_2)$ provides equality of opportunity.

We say that a pair of local mechanisms decentralizes the provision of equality of opportunity if it provides global equality of opportunity.

Let $\sup^e G$ be the projection of the support of G on the effort dimension, that is, $\text{supp}^e G = \{e | (b, r, e) \in \text{ supp} G\}$. Similarly let $\sup^{b,r} G$ be the projection of the support of G on the b and r dimensions, that is, $\text{supp}^{b,r} G = \{(b, r) | (b, r, e) \in \text{ supp} G\}$.

Definition 5: A local mechanism provides *local equality of rewards to effort* if for all G_j^*:

$$\forall e_j \in \text{supp}^e G_j^* \qquad \forall b_j, r_j \in \text{supp}^{b,r} G_j^*$$

$$r_j^i = r_j^k = r_j \text{ and } e_j^i = e_j^k = e_j \;\Rightarrow\; \phi_j(b_j^i, r_j, e_j; G_j^*) = \phi_j(b_j^k, r_j, e_j; G_j^*)$$

Equivalently if there exists w_j such that for all G_j^*:

$$\forall e_j \in \text{supp}^e G_j^*, \quad \forall b_j, r_j \in \text{supp}^{b,r} G_j^*, \quad \phi_j(b_j, r_j, e_j; G_j^*) = w_j(r_j, e_j; G_j^*)$$

A local mechanism providing local equality of rewards to effort allows rewards to effort only to depend on relevant characteristics. In other words, two individuals with different irrelevant characteristics but the same relevant characteristics should get the same if they chose to do the same effort. Notice that the rewards have to be the same only for the levels of effort that can be potentially chosen by some individual (it has to be in $\text{supp}^e G^*$). This implies that if a certain level of effort is never chosen by an individual for any profile of characteristics, the rewards there need not be equalized.

We now introduce notation that will allow us to describe the set of efforts in activity j that an individual with certain characteristics in activity j may choose. Given a pair of local mechanisms ϕ_1, ϕ_2 and the marginal distribution of characteristics in activity j, G_j, let $E_j(b_j, r_j; G_j)$ be the set of effort levels in activity j that an individual with characteristics (b_j, r_j) will choose for some $(b_{-j}, r_{-j}) \in I\!R \times I\!R$ under some distribution of characteristics in the other activity, $G_{-j}(b_{-j}, r_{-j}, e_{-j}) \in \Omega_{-j}$, that is:

$$\begin{aligned} E_1(b_1, r_1; G_1^*) &= \{e_1 \in I\!R | e_1 = e_1^*(b_1, b_2, r_1, r_2; G_1^*, G_2^*) \\ &\quad \text{for some } (b_2, r_2) \in I\!R \times I\!R \text{ and some } G_2^* \in \Omega_2 \\ E_2(b_2, r_2; G_2^*) &= \{e_2 \in I\!R | e_2 = e_2^*(b_2, b_1, r_2, r_1; G_1^*, G_2^*) \\ &\quad \text{for some } (b_1, r_1) \in I\!R \times I\!R \text{ and some } G_2^* \in \Omega_2\} \end{aligned}$$

Definition 6: The mechanisms and the environment are *rich* if for all $G_j^* \in \Omega_j$, $\text{supp}^e G_j^* \subseteq E_j(b_j, r_j; G_j^*)$ for all $(b_j, r_j) \in \text{supp}^{b,r} G_j^*$.

The mechanisms and the environment are rich whenever, a priori, the set of efforts in equilibrium, e_j^*, can be chosen by all individuals (b_j, r_j) under some circumstance in the other activity. For example, if schools observe that individuals put from 1 to 10 hours of study, then for each of these hours of study, every individual with given characteristics in school, (b_S, r_S), should be interested in choosing those hours for some circumstance that schools cannot observe. Therefore, whether you are in a good or bad school there should be circumstances outside of school that should make you interested in any of the 1 to 10 hours of study. For example, different levels of ability for basketball should lead you to choosing the range of hours 1 to 10 in school. This does not mean that independently of school characteristics individuals should be equally interested in studying the 1 to 10 hours, but that they should all be interested in studying them under some circumstance in basketball.[12] This is a condition on both the environment and the mechanisms since

[12] One circumstance where this property would never hold, independently of the space of individual characteristics or distributions is in the case where the utility and the cost function are separable in activity 1 and 2, since then the choice e_j is just a function of (b_j, r_j). In that case it is clearly not necessary to equalize

it requires enough heterogeneity in individual's characteristics and the outputs they get with those characteristics in each activity. If, for example, there is a large heterogeneity in characteristics for one sector, but the local mechanism for that activity assigns a fixed outcome to all individuals, then the local planner in the other sector can infer individual's choice of effort by only observing the individual's characteristics in his sector. In that case, therefore, the mechanism and the environment would not be rich.

Theorem 2.1. *If the pair of local mechanisms* (ϕ_1, ϕ_2) *provides local equality of rewards to effort then they provide global equality of opportunity.*

Moreover if the mechanisms and the environment are rich, and if the pair of local mechanisms (ϕ_1, ϕ_2) *is differentiable, then it provides global equality of opportunity only if it provides local equality of rewards.*

PROOF:

Part 1: If the pair of local mechanisms provide local equality of rewards to effort then it provides global equality of opportunity. The union of the local mechanisms gives rise to a social mechanism. Since the local mechanisms provide local equality of rewards to effort we know the effort chosen by the individual in equilibrium maximizes his utility given a social mechanism of the following form:

$$e^*(r; G^*) \equiv \arg\max_{e \in \boldsymbol{R}^2} u((w_1(r_1, e_1; G_1^*), w_2(r_2, e_2; G_2^*))) - c(e)$$

---

rewards to effort: equalizing partial utilities would be sufficient. But this assumes away any interaction between the activities, which is what we are trying to address.

So the final welfare of each individual is a function only of the relevant characteristics r:

$$u^*(b,r) = v(r) = u((w_1(r_1, e_1^*(r_1)), w_2(r_2, e_2^*(r_2))) - c(e_1^*(r_1), e_2^*(r_2))$$

Part 2:[13] *If a pair of local mechanisms provides global equality of opportunity then it must provide local equality of rewards.*

Given a pair of local mechanisms ϕ_1 and ϕ_2 a person (b_1, b_2, r_1, r_2) in equilibrium G^* chooses:

$$e^* = e^*(b,r;G^*) = \arg\max_{e \in \boldsymbol{R}^2} u(\phi_1(b_1, r_1, e_1; G_1^*), \phi_2(b_2, r_2, e_2; G_2^*)) - c(e_1, e_2)$$

Note that since each individual has zero mass, his actions will not change the distribution of efforts in equilibrium and his payoff will only depend on his effort. At the optimum then:

$$u^*(b,r) = u(\phi_1(b_1, r_1, e_1^*; G_1^*), \phi_2(b_2, r_2, e_2^*; G_2^*)) - c(e_1^*, e_2^*)$$

If the pair of local mechanisms provides global equality of opportunity then the following is true:

$$\forall G^* \in \Omega \; \forall b, r \in \text{supp}^{b,r} G^* \;\; \frac{\partial u^*(b,r)}{\partial b_1} = \frac{\partial u^*(b,r)}{\partial b_2} = 0$$

By the Envelope Theorem we know that $\frac{\partial u^*(b,r)}{\partial b_j} = \frac{\partial u}{\partial \phi_j}\frac{\partial \phi_j}{\partial b_j}$:

$$\frac{\partial u^*(b,r)}{\partial b_j} = 0 \Rightarrow \frac{\partial u}{\partial \phi_j}\frac{\partial \phi_j}{\partial b_j} = 0 \Rightarrow \frac{\partial \phi_j}{\partial b_j}(b_j, r_j, e_j^*(b,r;G^*)) = 0$$

For both local mechanisms this has to be true for all $e_j \in E_j(b_j, r_j; G_j^*)$, since from the

[13] I am grateful to John E. Roemer for providing the idea for this part of the proof.

point of view of activity j, E_j represents all the possible choices that individual b_j, r_j could want to choose under some circumstance in activity $-j$. Since the space is rich we know that $\text{supp}^e G_j^* \subseteq E_j(b_j, r_j; G_j^*)$. This implies that the local rewards to effort cannot change with b_j for any level of effort played in equilibrium, that is, for all $e_j \in \text{supp}G_j^*$, which is precisely what local equality of rewards to effort requires in the differentiable case.

2.4 Heterogeneous preferences

In this section we extend the framework to the case where the characteristics do not only affect the product in the activities, but also the utility of individuals. In particular, we introduce unobservable heterogeneity in the preferences. We allow for preferences to depend on unobservable individual characteristics, $t \in T$, that we assume are relevant,[14] for example laziness or taste for a given activity.[15] Note that the relevant characteristics affecting the productivity may be observable or unobservable. Since the policymaker can observe the irrelevant characteristics, effort, and the final output, he can infer the relevant characteristics. The point of this section is therefore to introduce unobservable relevant characteristics that the policymaker cannot infer. Individual characteristics are described by $F : T \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$. Given the choice of effort of the individuals we can describe individuals in this society by $G \in \Omega$, $G : T \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}$, where $G(\hat{t}, \hat{b}, \hat{r}, \hat{e}) = \Pr(\{(t, b, r) | t \leq \hat{t}, b \leq \hat{b}, r \leq \hat{r} \text{ and } e \leq \hat{e}\})$.

[14] If they were irrelevant and unobservable it would become impossible to compensate for any disadvantage and the provision of equality of opportunities would not be possible.

[15] Heterogeneity in preferences, when considered relevant, are commonly denoted *expensive tastes*.

Individual preferences can be represented by:

$$U^i = u(t,x) - c(t,e)$$

where u is strictly increasing and c *non separable* in e_1 and e_2.

We now redefine the main concepts to the extended framework.

Definition 1': A *social mechanism* μ is a mapping which assigns for every distribution $G \in \Omega$ an outcome vector to each individual:

$$\mu : T \times I\!R^2 \times I\!R^2 \times I\!R^2 \times \Omega \rightarrow I\!R \times I\!R$$

$$\mu \equiv (\mu_1, \mu_2) : (t,b,r,e,G) \rightarrow (x_1, x_2)$$

$$s.t. \int_{\boldsymbol{R}\times\boldsymbol{R}\times\boldsymbol{R}} \mu_j(t,b,r,e;G)\ dG \leq \int_{\boldsymbol{R}\times\boldsymbol{R}\times\boldsymbol{R}} a_j(b_j,r_j,e_j;G_j)\ dG \quad \text{for } j = 1,2$$

Definition 2': A social mechanism μ provides *equality of opportunity* if for all G^*:

$$\forall i,k \in I\ \ r^i = r^k \text{ and } t^i = t^k \ \Rightarrow\ u^*(t^i,b^i,r^i;G^*) = u^*(t^k,b^k,r^k;G^*)$$

In this context what the policymaker observes is exactly the same as in the previous section. Therefore G_j will be the marginal distribution of (b_j, r_j, e_j) (integrating out all characteristics unobserved to policymaker j, including t) and therefore a local mechanism will be defined as before.

Definition 3': a *local mechanism* is a function ϕ_j that assigns to every marginal distribution, G_j, of individual characteristics and choices specific to activity j a local outcome vector to each individual:

$$\phi_j : I\!R \times I\!R \times I\!R \times \Omega_j \to I\!R$$

$$\phi_j : (b_j, r_j, e_j; G_j) \to x_j$$

$$\text{s.t.} \int_{\boldsymbol{R}\times\boldsymbol{R}\times\boldsymbol{R}} \phi_j(b_j, r_j, e_j; G_j)\, dG_j \leq \int_{\boldsymbol{R}\times\boldsymbol{R}\times\boldsymbol{R}} a_j(b_j, r_j, e_j; G_j)\, dG_j$$

Equality of rewards to effort is only conditioned on the observables. The definition is therefore the same as before:

Definition 5': A local mechanism provides *local equality of rewards to effort* if for all G_j^*:

$$\forall e_j \in \text{supp}^e G_j^* \qquad \forall b_j, r_j \in \text{supp}^{b,r} G_j^*$$

$$r_j^i = r_j^k = r_j \text{ and } e_j^i = e_j^k = e_j \;\Rightarrow\; \phi_j(b_j^i, r_j, e_j; G_j^*) = \phi_j(b_j^k, r_j, e_j; G_j^*)$$

Equivalently if there exists w_j such that for all G_j^*:

$$\forall e_j \in \text{supp}^e G_j^* \qquad \forall b_j, r_j \in \text{supp}^{b,r} G_j^* \qquad \phi_j(b_j, r_j, e_j; G_j^*) = w_j(r_j, e_j; G_j^*)$$

Next theorem shows that equalizing rewards on the observed characteristics will also provide global equality of opportunity.

Theorem 2.2. *If the pair of local mechanisms (ϕ_1, ϕ_2) provides local equality of rewards to effort then it provides global equality of opportunity. Moreover if the environment is rich, and if the pair of local mechanisms (ϕ_1, ϕ_2) is differentiable then it provides global equality* *of opportunity only if it provides local equality of rewards.*

PROOF:

Part 1: If the pair of local mechanisms provides local equality of rewards to effort then it provides global equality of opportunity. The union of the local mechanisms gives rise to a social mechanism. Since the local mechanism provides local equality of rewards to effort we know the effort chosen by the individual maximizes his utility given a social mechanism of the following form:

$$e^*(t,r) \equiv \arg\max_{e \in \boldsymbol{R}^2} u((t, w_1(r_1, e_1; G_1^*), w_2(r_2, e_2; G_2^*))) - c(t,e)$$

The individual's utility in equilibrium is denoted by $u^*(t,b,r;G^*) = u(\mu(t,b,r,e^*;G^*)) - c(e^*)$. The final welfare of each individual is a function only of the relevant characteristics t and r:

$$u^*(t,b,r) = v(t,r)$$

$$v(t,r) = u((t, w_1(r_1, e_1^*(t,r); G_1^*), w_2(r_2, e_2^*(t,r); G_2^*))) - c(t, e_1^*(t,r), e_2^*(t,r))$$

Part 2: If a pair of local mechanisms provides global equality of opportunity then it must provide local equality of rewards. Same as for Theorem 1.

This shows that a pair of local mechanisms providing local equality of rewards to effort decentralizes equality of opportunity if individuals are heterogeneous in unobservable relevant characteristics affecting their preferences.

2.5 Concluding remarks

The allocation of rights and resources in a society is done in a decentralized manner, with each policy maker deciding how to allocate resources based on partial information. A policymaker concerned about the justice of his policy faces a local justice problem. Local justice problems have not been given as much attention as global distributive justice problems in the literature, and their solutions have been somewhat frustrating. Local justice has been studied by induction due to the lack of results showing how the collection of local problems may aggregate. The book *Local Justice* by J. Elster, for instance, may lead the reader to think that there is no general rule to be applied to every local justice problem guaranteeing the attainment of local justice. Consequently, local justice is bound to be case specific. This leads us to conclude that policymakers cannot be informed on local justice issues on a general basis.

This paper interprets local distributive justice as decentralized global distributive justice. Thinking of local justice as decentralized global justice provides a systematic approach to the problem and allows us to evaluate local interventions on the basis of their success in achieving global justice.

Equality of opportunity is of great concern in modern societies, but its implications for institutions when they design their local policies is not well understood. In this paper we analyze what local justice requires if we want to achieve the global justice goal of providing equality of opportunity. We show that under some conditions a collection of local mechanisms decentralizes the provision of equality of opportunity if and only if each mechanism provides local equality of rewards to effort. This implies that if a society is concerned about providing equality of opportunity, then its institutions should be concerned about equalizing rewards to effort, that is, the incentives individuals are given should not depend on irrelevant characteristics.

This paper, however, motivates but does not address important questions. First it does not address how restrictive equalizing rewards to effort is. Sometimes equalizing rewards to effort may be hard and involve informational and efficiency problems. But there may be circumstances where equalizing rewards to effort may just imply moving away from lump sum forms of assistance (like the lump sum nature of Affirmative Action in some universities) into proportional forms of assistance. A second important aspect that the paper does not address is how "costly" it is not to locally equalize rewards to effort and to naively equalize opportunities (by making local outcomes independent of irrelevant characteristics). Also, and related to the last point, we do not analyze how this goal may interact with other objectives that policymakers may have when designing their policies.

In addition, this paper does not explore how inequality of opportunity may be measured but it suggests that we may also change how we evaluate to what extent there is equality of opportunity for aspects like education or income earning. For example, if we are evaluating to what extent there is equality of opportunity for income earning in two countries, this paper suggests that income is not the relevant variable to be equalizing among advantaged and disadvantaged, but income per hour. Benabou and Ok (2001) conclude that USA seems to provide more equality of opportunity than Italy, but they do not control for hours worked. So the results could be driven by the fact that in the US disadvantaged groups work three times as hard as advantaged groups (by working three shifts), but not in Italy because of strong regulations limiting the amount of hours that workers can work. So the fact that the income distribution is more similar between advantaged and disadvantaged groups in the US may be mainly because the disadvantaged systematically work many more hours, making the gap in income smaller. If that was the case we would not necessarily conclude that there is more equality of opportunity in the US than in Italy. Controlling for hours worked may change the results.

Finally, another interesting idea that I do not address is whether people think of equality of opportunity as an important distributive justice criterion, but the results presented provide a rationale for the difference between the distributive justice notions that political philosophers propose, i.e., normative distributive justice, and those that people express when voting for welfare programs or when stating their opinions in surveys or experiments, i.e., positive notions of distributive justice. Political philosophers overlook the role of effort because it is endogenous and ultimately a function of exogenous characteristics. Instead the debate has centered on what exogenous characteristics should be rewarded.[16] Individuals in a society, however, seem to care about the effort of those on assistance programs in addition to exogenous characteristics. For example, Bowles and Gintis (1998) report that "[T]he welfare state is in trouble not because selfishness is rampant (it is not), but because many egalitarian programs no longer evoke, and sometimes offend, deeply held notions of fairness [...] In a 1995 CBS/NYT survey, for example, 89% supported and mandated work

[16]An example of a discussion on rewarding effort and justice is in Sen's article in *Meritocracy and economic inequality* where he states that "[A]n incentive argument is entirely "instrumental" and does not lead to any notion of intrinsic "desert". If paying a person more induces him or her to produce more desirable results, then an incentive argument may exist for that person's pay being greater. This is an instrumental and contingent justification—it does not assert that the person intrinsically "deserves" to get more[...]" Actions are rewarded for what they help to bring about but the rewarding is not valued in itself.

requirement for those on welfare."[17] Yet, Konow (2003) in a survey on positive distributive justice states "[a] common view is that differences owing to birth, luck and choice are all unfair and that only differences attributable to effort are fair. A common finding (and claim) among social scientists is that individual effort affects the perceived fairness of allocations". We believe this difference between political philosophers and individuals in society may be the result of philosophers examining global distributive justice while individuals examining local distributive justice. If global equality of opportunity is desirable, then, even though we can ignore effort—or more generally endogenous choice variables— in the global distributive justice problem, we must consider it in the local distributive justice problem.

[17] Bowles and Gintis argue that the reason people care about the effort recipients of welfare exert is due to strong reciprocity. "[H]*omo reciprocans* is not committed to the abstract goal of equal outcomes but to a rough balancing out of burdens and rewards."

Chapter 3

Is Education Attainment determined at age 7?

3.1 Introduction

The power of education as an instrument to allow for social mobility has been questioned by various empirical studies. The large correlation between the performance of an individual as a child and as an adult has been interpreted as the role of education as an instrument to allow for social mobility being small. Whatever determines an individual's performance as an adult is present when he is a child. There is some "ability" determined at early stages that characterizes the individual's attainment throughout his life. Several previous studies have tried to examine the relationship between age 7 test scores and future outcomes. For example, Hutchinson, Prosser and Wedge (1979) found that early test scores were strong predictors of future education attainment. Connoly, Mickelwright and Nickell (1992), and more recently Robertson and Symons (1996) and Harmon and Walker (1998) examine the relationship between early test scores and future earnings. All of these studies find that early test scores matter for future outcomes.

This paper provides empirical evidence consistent with the high correlation between early characteristics and later attainment, but also with the education process being important in explaining the final individual educational attainment. There are at least two factors explaining these high correlations. The first is that an individual that is better prepared in the initial stages will be more productive in later stages given that the knowledge in each stage is complements to the acquisition of knowledge in the next stage. But another factor that may be explaining this correlation and that would be consistent with the education process being responsible for student's attainment is if the characteristics affecting the initial performance of individuals was very correlated with the education process that individuals face. For example, a well educated family is more likely to give the appropriate tools to a child so that it performs well at initial stages. It is also more likely to ensure that the education process is the best, given the child's characteristics, by providing supplementary education and motivation to the kid. This will help maintain the correlation between early stage characteristics and later achievement. The idea is then that the education process tends to be best for individuals who are better endowed early on and that this may be an important factor driving the high correlations between early and later achievement. If this was not the case, that is, if early characteristics were the only determinants of later attainment, we would find that including performance at intermediate stages of and individual's education's process should not add explanatory power to the regression. This paper provides some evidence that this may not be the case and that there are crucial aspects in an individual's educational experience that make the explanatory power of age 11 scores much more relevant than those at age 7, contradicting the idea that the education process itself does not matter.

The study uses the National Child Development Survey (NCDS), which includes information about the entire cohort born in Great Britain during one week in March 1958. Information about these individuals has been collected from before birth until the present. This information includes family characteristics and school characteristics at ages 7, 11 and 16. The individuals were given math and reading tests at these three different ages. Section 2 describes in more detail the data used.

Section 3 replicates the results obtained by different papers on the correlation between age 7 and age 16 test scores, first without conditioning on any other variable, then including family background variables and last including both background and school variables.

Section 4 presents the main results of this paper. It includes a regression analysis of age 16 test scores on all of the variables included in the previous regression, but will control for test scores and school variables at age 11. Section 5 decomposes the different correlations in order to try to identify the different factors determining the dynamics of the education attainment process.

The last section provides two simple tests for determining whether the results we observe in the fourth section can be a result of test scores being a worse measure of ability at 7 then at 11. One of the reasons for age 7 tests scores to have a higher measurement error is because some kids have to adapt to the specific way in which there asked to show their ability. Sociological studies show that diferent sociological groups present different ways of learning and of providing evidence of their skills. If advantaged children are exposed at home to similar questions and problems as the exams, for them test scores will be a better measure of ability than for disadvantaged children at age 7. But at age 11 disadvantaged children may have learned to take tests through having been exposed to the type of questioning. Hence test scores will become a better measure of the ability of groups with less educated parents. We test for this increase measurement error at age 7 by doing the same analysis as in section 4 for two subsamples of the population: the group with most educated and less educated parents. The results show that there is no significant difference between the coefficients in the two regressions, suggesting that this sociological effect is not driving the results. The second analysis provides some simulations of a simple model where the measurement error was larger at age 7 and at age 11 to provide some idea on how large the difference in variance would have to be to replicate the results presented in section 4. The simulations show that for the results to be explained only by a larger variance in the measurement error at age 7, it would have to be at least five times larger at age 7 than at age 11.

3.2 The data

The National Child Development Study (NCDS) is a continuing, multi-disciplinary longitudinal study which takes as its subjects all those living in Great Britain who were born between 3 and 9 March 1958.

It has its origins in the Perinatal Mortality Survey (PMS), sponsored by the National Birthday Trust Fund. This was designed to examine the social and obstetric factors associated with still birth and death early in infancy among the 17.000 children born that one week.

To date there have been six attempts to trace all members of the birth cohort in order to monitor their physical, educational and social development. These were carried out by the National Children's Bureau in 1965 (NCDS1), 1969 (NCDS2), 1974 (NCDS3), and 1981 (NCDS4); by the Social Statistics Research Unit (SSRU) of City University in 1991 (NCDS5); and by the Centre for Longitudinal Studies of the Institute of Education, University of London and the National Centre for Social Research in 1999 (NCDS6). In this paper the following sources of information will be used:

PMS (1958)	NCDS1 (1965)	NCDS2 (1969)	NCDS3 (1974)
Birth	7	11	16
17,4141	15,568	15,503	14,761
Mother	Parents	Parents	Parents
	School	School	School
	Tests	Tests	Tests
Medical	Medical	Medical	Medical

Table 3.1: Sources of information

For the birth survey information was obtained from the mother and from medical records by the midwife. For the purpose of the first NCDS surveys, information was obtained from parents (who were interviewed by health visitors) head teachers and class teachers (who completed questionnaires), the schools health service (who carried out medical examinations) and the subjects themselves (who completed tests of ability and, latterly, questionnaires).

Attrition is often a problem in longitudinal surveys and the NCDS is no exception. Overall responses have remained high, considering the length of the panel. However individuals disappear and reappear. In fact restricting the sample to those who appear in every wave would result in a drastic reduction in sample size. Instead, we will use the available sample for each analysis.

Previous analyses of this data suggest that attrites are more likely to be from disadvantaged backgrounds, although observable differences between the two groups are quite small (Fogelman, 1976, 1983; Robertson and Symons, 1996; Connolly, Micklewright and Nickell, 1992). In this paper this problem will be solved by controlling for background characteristics.

The tests we focus on are standardized tests of reading and mathematics which were administered to the children in their schools by their teachers. We have standardized the results so that they have mean 0 and variance 1.

Our measures of children's background consist on the highest education level attained by any of the two parents, whether one of the parents has a skilled job and whether they want the child to continue studying after age 16.

For measures of school quality we will use different variables depending on their availability, but they will mainly include class size, stability of teachers in the school and whether the school has parent-teacher association.

3.3 Estimated effects of early test scores

3.3.1 Estimated effects of test scores at 7 on test scores at 16

The first table shows the results of running a simple regression of math and reading scores at age 16 on math and reading scores at age 7:

As we can observe the coefficients on age 7 scores are clearly significant and of considerable size, but we are missing variables that we know are going to have some explanatory

	math score 16	reading score 16
cons	0.005	0.0013
(t-stat)	(0.66)	(0.18)
math score 7	0.49	
	(57.43)	
reading score 7		0.63
		(80.65)
R^2	0.24	0.38

Table 3.2: Effects of test scores at 7 on test scores at 16

power. The following sections will control for different sets of variables including schooling, background and age 11 test scores.

3.3.2 Controlling for background and school variables

Including parents' education and number of siblings

Table 3.3 shows the results from running a simple regression of math and reading scores at age 16 on math and reading scores at age 7, the highest education level attained by the parents and the number of siblings. This results replicate a similar analysis done by Currie and Thomas in 1998.

	math score 16	reading score 16
cons	-1.69	-0.84
(t-stat)	(-20.39)	(-11.71)
math score 7	0.42	
	(42.37)	
reading score 7		0.55
		(59.1)
parents education	0.13	0.08
	(26.91)	(18.41)
siblings	-0.08	-0.09
	(-14.32)	(-17.50)
R^2	0.31	0.42

Table 3.3: Effects of age 7 test scores and family background characteristics on math and reading scores at age 16

Including a richer set of background and school variables for age 7 and 16

Table 3.4 shows the results obtained from running a simple regression of age 16 test scores on age 7 test scores, background and school variables. Background variables include parents education, whether some parent works on a skilled job, whether they want the child to continue studying after age 16 and the number of siblings. School variables include whether the school has parent-teacher association at age 7 and 16 (reflecting how involved parents are in the education of their children and also imposing some monitoring for the school's performance), class size at age 7 and age 16,and number of teachers who left the school during the school year at age 16 (reflecting the stability of the school). The choice of variables depended strongly on their availability in the data set and the coherence with the analysis.

All coefficients are significant and have the expected sign except for class size. We would expect a larger class size to result in worse scores, but this is a normal finding in the literature and can be a result of schools assigning students that are having difficulties into smaller classes to allow them to catch up. Also in the regression on reading test whether there is a parent-teacher association is not significant.The R^2 has increased with the inclusion of the richer set of explanatory variables, although not by a large amount.

3.4 Including age 11 scores and school variables

In this section we have now included variables that contain information about the performance of the kids and schools at the middle stage, i.e. at age 11. Table 3.5 reflects the results of running the same regression as in the last section, but now including test scores

	math score 16	reading score 16
cons	-2.54	-1.96
(t-stat)	(-22.54)	(-20.05)
math score 7	0.41	
	(43.26)	
reading score 7		0.50
		(55.86)
parents education	0.12	0.068
	(23.04)	(15.65)
parent has skilled job	0.21	0.19
	(9.27)	(9.95)
par want kid study	0.17	0.24
	(3.44)	(5.51)
siblings	-0.065	-0.07
	(-11.92)	(-15.91)
parent-teacher assoc. at 7	0.11	0.04
	(4.77)	(1.81)
class size at 7	0.002	0.05
	(2.04)	(4.67)
n. teach. left during year at 16	-0.006	-0.003
	(-3.96)	(-2.51)
parent-teacher assoc. at 16	0.062	0.11
	(3.32)	(6.54)
class size at 16	0.023	0.025
	(14.33)	(17.49)
R^2	0.35	0.46

Table 3.4: Effects of age 7 test scores, family background and school characteristics on math and reading scores at age 16

at age 11, percentage of teacher in the school who had less than two years of experience (which in the literature is taken as a bad characteristic of the school teachers) and class size at age 11.

The results in this case are surprising. Some school characteristics loose significance, but the rest of the variables are still significant. But the absolute value of the coefficients are all much smaller than in the previous section. Specially the coefficient on test scores at 7 are reduced from 0.37 to 0.07! And the R^2 has doubled, going from 0.35 to 0.61. These results are consistent with the education process having some relevant role in the middle stages too. We could think that the relevant part of the education process is primary school. There the basics for future success are determined. If kids have been able to perform well up until then they will most probably do well in the future also.

We have to be careful and think about other forces driving these results. One other possible consistent explanation would be that test scores have a bigger measurement error of some underlying ability in early stages than in, say, middle stages. We will test for a version of this in the sixth section of the paper, and we will refute this hypothesis.

3.5 Decomposition of the different correlations between factors affecting the child's performance

3.5.1 Explaining age 7 test scores through background and school variables at age 7

Table 3.6 shows the results of running a regression of age 7 test scores on different background variables. The R^2 is extremely small and the school variables are not significant,

	math score 16	reading score 16
cons	-1.06	-0.8
(t-stat)	(-10.6)	(-9.32)
math score 7	0.08	
	(8.8)	
reading score 7		0.17
		(19.56)
parents education	0.056	0.014
	(13.01)	(3.63)
parent has skilled job	0.06	0.097
	(3.42)	(5.97)
par want kid study	0.003	0.18
	(0.1)	(4.97)
siblings	-0.023	-0.03
	(-5.25)	(-9.85)
parent-teacher assoc. at 7	0.08	0.02
	(3.79)	(1.42)
class size at 7	0.0002	0.002
	(0.16)	(2.93)
math score at 11	0.68	
	(69.29)	
reading score at 11		0.61
		(70.63)
perc. of teacher less 2 exp.	0.07	0.007
	(2.26)	(0.26)
class size at 11	0.0001	0.0014
	(0.17)	(1.46)
n. teach. left during year at 16	-0.002	-0.001
	(-1.81)	(-1.43)
parent-teacher assoc. at 16	0.03	0.065
	(1.95)	(4.83)
class size at 16	0.005	0.01
	(4.18)	(11.14)
R^2	0.61	0.68

Table 3.5: Effects of including age 11 test scores and school characteristics in explaining math and reading scores at age 16

which is consistent with the literature (see Hanushek 1986). But background variables are significant. Better educated or skilled parents and their care about the kids studies will positively affect the performance of the kid in school.

	math score 7	reading score 7
cons	-1.12	-1.91
(t-stat)	(-8.75)	(-15.97)
parents education	0.07	0.08
	(12.85)	(15.87)
parent has skilled job	0.15	0.24
	(5.62)	(9.98)
par want kid study	0.44	0.64
	(8.59)	(13.37)
local authority school	-0.46	-1.19
	(-6.86)	(-3.06)
parent-teacher assoc. at 7	0.034	0.03
	(1.19)	(1.03)
class size at 7	-0.001	0.002
	(-0.77)	(2.05)
R^2	0.054	0.083

Table 3.6: Effects of family background and school characteristics on age 7 test scores

3.5.2 Explaining age 11 test scores through background and school variables

Table 3.7 presents the results of regressing test scores at age 11 on test scores at 7 together with various school and background variables. We see the clear effect that past test scores have. Again, having understood the materials thought at 7 well is highly correlated with understanding the new material at 11. Background variables are also important and school variables are significant, except for class size, which has been behaving differently throughout the analysis presented in this paper and in the literature, resulting from the endogeneity problem.

	math score 11	reading score 11
cons	-2.06	-1.57
(t-stat)	(-17.96)	(-14.13)
math score 7	0.48	
	(45.9)	
reading score 7		0.53
		(47.91)
parents education	0.096	0.09
	(18.89)	(17.88)
parent has skilled job	0.23	0.12
	(9.13)	(5.02)
par want kid study	0.34	0.14
	(6.91)	(3.04)
siblings	-0.07	-0.07
	(-11)	(-12.2)
parent-teacher assoc. at 7	0.08	0.066
	(3.15)	(2.58)
class size at 7	0.001	0.005
	(0.85)	(0.38)
perc. of teacher less 2 exp.	-0.13	0.0013
	(-3.17)	(0.03)
class size at 11	0.01	0.08
	(7.75)	(5.26)
R^2	0.38	0.40

Table 3.7: Effects of test scores at age 7, family background and school characteristics on age 11 test scores

The results in this section should support the idea that background variables affect, together with school variables, the performance in schools through primary school. And in last section we see that the this performance at age 11 is what mainly determines future performance. Therefore, the initial positive effect families and schools have on certain children multiplies out throughout their lives. This suggests that if society were interested in allowing for social mobility it should intervene at the primary school level and compensate disadvantaged children through providing extra help in acquiring habits of study and interest to pursue their studies, and to provide extra libraries and tutoring services, elements which should try to substitute for elements which are missing in their homes and that justifies calling them disadvantaged children.

3.6 Are age 7 tests worse estimators of ability than age 11 test scores?

3.6.1 Do children "learn" to take tests?

In section 4 we saw that including age 11 test scores double the explanatory power of the regression, reduces the coefficients on early in life variables and results in a very large coefficient on age 11 test scores. The literature in sociology and psicology has often studied how the school systems test children in a way that is more advantageous for kids with certain backgrounds or cultures. For example, it is said that white-european families teach their kids by making them repeat what they just said. These families tend to be rather small and devote proportionately more attention to their kids. But latin-american families are big, with old, young and children living in the same house. Kids are expected to learn by looking and listening. They are good at paying attention to many things at a time because that is what they were exposed to in their homes. Therefore sitting down and focusing on simply repeating the material taught in class may be more difficult for a latin kid. If that is true the test will not reflect the true ability of the kid because of the framing of the questions. Well, if that's true than one can think that after four years in school the latin kids will adapt to the new way of learning and presenting what you have learned in school, and at age 11 they will be able to reflect their true ability better.

Now, suppose that the only factor driving test scores is some sort of ability. We will expect kids with less educated parents to have a larger error term in the age 7 tests than in the age 11. This could be a result of educated parents training or educating their children in a compatible way to how they know their kids will be expected to learn in school.

If that was the case, not allowing for different coefficients on the scores at 11 for kids with different backgrounds can derive to the misleading conclusion that there is something else, other than ability driving the education process. In this section then, we try to test whether this sociological phenomenon may be driving the results in section 4. In order to do that we take two sub-samples from the data, one with the best educated parents and one with the worse educated parents. We would expect the coefficient on the test scores at 11 to be insignificant or at least significantly smaller for the well educated parents than for the less educated parents group. The tables bellow show the results of the regression in section 4, but in the two different samples:first for the sample of highly and then for that of the poorly educated parents.The results reject the hypothesis of the coefficient for the well educated parents group to be significantly smaller.

	math score 16	reading score 16
cons	-0.8	-0.28
(t-stat)	(-4.8)	(-2.12)
math score 7	0.08	
	(5.71)	
reading score 7		0.2
		(14.72)
parents education	0.042	-0.002
	(6.32)	(-0.36)
parent has skilled job	0.12	0.07
	(3.52)	(2.64)
par want kid study	0.03	0.14
	(0.45)	(2.38)
siblings	-0.023	-0.04
	(-3.18)	(-6.2)
parent-teacher assoc. at 7	0.06	0.04
	(2.15)	(1.75)
class size at 7	-0.0009	0.001
	(-0.60)	(1.13)
math score at 11	0.71	
	(47.46)	
reading score at 11		0.55
		(46.13)
perc. of teacher less 2 exp.	0.12	-0.004
	(2.4)	(-0.011)
class size at 11	-0.0002	0.001
	(-0.14)	(1.09)
n. teach. left during year at 16	-0.005	-0.001
	(-2.36)	(-0.62)
parent-teacher assoc. at 16	0.02	0.041
	(0.88)	(2.17)
class size at 16	0.005	0.01
	(2.38)	(6.15)
R^2	0.63	0.68

Table 3.8: Effects of including age 11 test scores and school characteristics in explaining math and reading scores at age 16 for the sample of highly educated parents

	math score 16	reading score 16
cons	-0.93	-0.55
(t-stat)	(-1.34)	(-0.89)
math score 7	0.07	
	(4.64)	
reading score 7		0.14
		(9.62)
parents education	0.04	-0.006
	(0.87)	(-0.14)
parent has skilled job	0.0008	0.07
	(0.03)	(2.51)
par want kid study	-0.02	0.15
	(-0.39)	(2.59)
siblings	-0.019	-0.03
	(-2.72)	(-5.15)
parent-teacher assoc. at 7	0.07	0.04
	(1.84)	(1.02)
class size at 7	0.001	0.003
	(0.16)	(1.44)
math score at 11	0.64	
	(37.3)	
reading score at 11		0.66
		(40.16)
perc. of teacher less 2 exp.	0.03	-0.03
	(0.56)	(-0.55)
class size at 11	0.002	0.002
	(0.77)	(1.43)
n. teach. left during year at 16	0.002	-0.003
	(0.81)	(-1.32)
parent-teacher assoc. at 16	0.03	0.07
	(1.26)	(2.8)
class size at 16	0.01	0.01
	(2.71)	(6.2)
R^2	0.68	0.65

Table 3.9: Effects of including age 11 test scores and school characteristics in explaining math and reading scores at age 16 for the sample of poorly educated parents

3.6.2 How bad must tests at 7 compared to tests at 11?

This paper is not able to discard the hypothesis that the results presented are driven by the fact that at age seven tests are a worse estimators of ability than at age 11. In this section a very simple model of the education process is presented. Through running some simulations we characterize how much larger the variance in the noise at age 7 would have to be relatively to the noise at age 11 in order to generate the results presented in the previous sections.

To simplify the analysis, the structural model that we use to do the simulations ignores the variables other than tests scores that we know would determine the education attainment at each stage. We represent the joint effect of these variables on the test scores by a random variable,w, which we will assume to be normally distributed.

At age 7, the test score is determined by three factors: *ability*, representing all those characteristics of the individual that will help him succeed in school, w_7, including all the variables other than ability that could potentially be influencing the individual's achievement, and a random factor u_7, which represents the error of the test when measuring school attainment.

$$t_7 = w_7 + u_7$$

where $u_7 = \sigma_7 \epsilon$, w_7 is distributed $N(ability, 1)$, and ϵ and $ability$ are $N(0, 5)$.

At age 11 and at age 16, test scores are determined not only by the variables mentioned above, but also potentially by what happened in the past, reflecting that there may be learning underlying the process, i.e. what you achieved in the past will determine how much you can achieve today.

$$t_{11} = \alpha \cdot w_7 + w_{11} + u_{11}$$

$$t_{16} = \alpha \cdot (t_{11} - u_{11}) + w_{16} + u_{16} = \alpha^2 \cdot w_7 + \alpha \cdot w_{11} + w_{16} + u_{16}$$

where $u_{11} = \sigma_{11}\epsilon$, $u_{16} = \sigma_{16}\epsilon$, and w_{11} and w_{16} are distributed $N(ability, 1)$.

Without loss of generality we can fix $\sigma_{16} = \sigma_{11} = 1$, since we only care about the relative size of the variance at 7 with respect to the variance at 11, and since σ_{16} affects only the error term in the regression.

Once the data has been simulated we have normalized, just as we did with the actual data. Then we have estimated the following regression:

$$t_{16} = \beta_0 + \beta_7 \cdot t_7 + \beta_{11} \cdot t_{11}$$

The following table shows the relevant results from the simulation and regression:

The results show that, in order to replicate the results presented in section 4, either there is some relevant cumulative effect of education attainment over time, reflecting the relevance of the years in school, or $\sigma_7 \geq 5 \cdot \sigma_{11}$. Looking at the table it seems most likely that the two hypothesis are true, i.e., there is both some learning and some decrease of the variance of the error term over time. What seems less plausible is that there is no learning at all. If one believes the education process does not matter, than it also has to believe that the variance of the test 7 error is at least five times larger than at age 11, i.e. individuals reflect their knowledge significantly better when they are 11 than when they are 7.

α	σ_7	β_7	β_{11}
0	1	0.42	0.4
0	3	0.14	0.58
0	**5**	**0.09**	**0.6**
0	**7**	**0.06**	**0.61**
0.4	1	0.33	0.58
0.4	3	0.11	0.74
0.4	**5**	**0.06**	**0.76**
0.4	**7**	**0.04**	**0.76**
0.7	1	0.27	0.67
0.7	**3**	**0.08**	**0.81**
0.7	**5**	**0.05**	**0.83**
0.7	7	0.03	0.83
1	1	0.22	0.74
1	**3**	**0.06**	**0.86**
1	5	0.03	0.88
1	7	0.02	0.88

Table 3.10: Results after simulating test scores and replicating regression results

3.7 Conclusions

Education has often been thought to be an instrument to allow for social mobility. Empirical evidence has been questioning this by noticing a very large correlation between early test scores and future outcomes and interpreting this results as evidence that no matter what the child does while growing up his future is basically determined. This paper wants to put forward an alternative interpretation of the process driving the observed facts and to provide some evidence for it. We want to show that children's experience between ages of 7 and 11 does matter for future outcomes, and that the strong correlation results from the process between ages 7 and 11 being correlated with what happened to the child before and during that period. Kids with a more advantaged backgrounds will be taught studying habits, may be more be motivated to study and will tend go to better schools. This will be reflected in them getting higher test scores initially. Then, once they get a good understanding of the basic materials they are better prepared to learn the materials they are exposed to later and that will again be reflected in better scores. But it is not necessarily the case that they will do well in the future no matter what they do when growing up. In the same way, what a disadvantaged kid does in the beginning will determine his chances of doing well when growing up and in the future. But if he manages to compensate his disadvantaged background with a "good" education process and, getting good scores at age 11, he will have big chances of changing what was thought to be a predetermined future.

The analysis provides evidence consistent with this by using the NCDS data. First we replicate the standard analysis in the literature, where future outcomes are shown to be correlated to initial background characteristics and tests scores. Then we introduce test scores at age 11 and school variables at age 11 and we see that the test scores at 7 and background variables are still significant, but that the coefficients are much smaller. And on the other hand the coefficient on age 11 test scores are significant and very large. The explanatory power of this new regression is also double as high as for the standard analysis. We take this to point out that there might be something important between the 7 and 11 ages.

Suppose that all that really mattered was ability, which was determined at age 7. If test scores had a higher measurement error at age 7 then at age 11, then we would get these same results too. We haven't figured out a way to show that this can not be the case under any circumstances. We do provide some evidence that it can not be true that the higher variance is a results of the heterogeneous backgrounds of the children.

Given the absence of a methodology that would allow to discard the alternative generating process we provide some simulations that suggest that most likely the two hypothesis are true, that is, there is both some learning and some decrease in the test error over time. What seem less plausible is that there is no learning at all, since that would require the variance in the noise in the test at age 7 to be at least five times larger than that at age 11.

Chapter 4

Rationalizing and Curve-Fitting Demand Data with Quasilinear Utilities

4.1 Introduction

Theoretical models of consumer demand assume consumer's choice is derived from the individual maximizing a concave utility function subject to a budget constraint. Given a finite set of price and consumption choices, if there exist a concave, continuous and monotonic utility function such that these choices are the maxima of this function over the budget set, then we say that the data set is rationalizable.

Afriat (1967) provided the first necessary and sufficient conditions for a finite data set to be rationalizable, that is for it to be a result of the consumer maximizing her utility subject to a budget constraint. Afriat (1981) and Varian (1983) subsequently derived necessary and sufficient conditions for a finite data set to be rationalizable by homothetic or by separable utility functions.

Quasilinear rationalizations are used in a wide range of areas in economics, including theoretical mechanism design, public economics, industrial organization and international trade, one reason being that in this case changes in consumer surplus are equivalent to changes in consumer welfare.

Quah has shown that homothetic utility functions generate monotone demand functions, i.e. $x(p, I)$ is said to be monotone if $x(p_1, I) - x(p_2, I) \cdot (p_1 - p_2) \leq 0$. This is a multivariable version of the law of demand, i.e. an increase in prices results in a decrease in demand. Demand functions generated from quasilinear utility maximization subject to a budget constraint enjoy an additional property not possessed by monotone demand functions: they are cyclically monotone. Such demand functions are said to satisfy the strong law of demand [1]. Cyclically monotone demand functions not only have downwards sloping demand curves, in the sense that they are monotone functions, but also their line integrals are path-independent and therefore measure the change in consumer welfare for a given multidimensional change in prices.

In the first part of this paper we show that a finite data set consisting of pairs of price vectors and consumption vectors can be rationalized by a quasilinear utility maximization subject to a budget constraint if and only if the data set is cyclically monotone. Moreover we derive a linear program, from the associated Afriat inequalities, that is solvable if and only if the data is cyclically monotone.

Suppose the data is not cyclically monotone, then by our theorem there is no quasilinear

[1] see Brown and Calsamiglia (2003) for further discussion

rationalization of the data. In this case we ask what is the "best" quasilinear approximation to the data. We address this question in the second section of the paper in two steps. First we show that our non-parametric characterization of quasilinear rationalization extends to maximization of a random quasilinear utility function of the form $u(x) + \epsilon \cdot x + x_0$, subject to a budget constraint. Random quasilinear utility functions have also been discussed by Brown and Wegkamp (2003). Their specification, $v(x, e) = u(x) + \epsilon \cdot x + x_0$ is a special case of the random utility model suggested originally by Brown and Matzkin (1998). If the shocks are assumed to be bounded, as in Brown and Wegkamp and Brown and Matzkin, then we show that the random model, as in the deterministic case, is testable.

If the shocks are not bounded then any data set can be rationalized by a random quasilinear utility function. This suggests another interpretation of ϵ, as slack variables in fitting a quasilinear utility function to demand data. In this case we choose the quasilinear approximation with least absolute deviation as the "best" quasilinear rationalization of her choices.

4.2 The Strong Law of Demand and Quasilinear Rationalizability

Afriat (1967) provides the first non-parametric test for consumer behavior. He provides a necessary and sufficient condition on finite data for it to be rationalizable by a neoclassical utility function.

Definition 1: Let (p_r, x_r), $r = 1, ..., N$ be given. A utility function u rationalizes the data if for all $r = 1, ..., N$ x_r solves:

$$\max_{x \epsilon \boldsymbol{R}^n_{++}} \quad u(x)$$
$$\text{s.t.} \quad p_r x \leq I = p_r x_r$$

Theorem 4.1 (Afriat 1967). *The following conditions are equivalent:*

1. *There exists a concave, monotonic, continuous, non-satiated utility function that rationalizes the data.*

2. *The data (p_r, x_r), $r = 1, ..., N$ satisfies Afriat inequalities, that is, there exists $U_r > 0$ and $\lambda_r > 0$ for $r = 1, \ldots, N$ such that*

$$U_r \leq U_l + \lambda_l p_l \cdot (x_r - x_l) \ \ \forall r, l = 1, \ldots, N$$

3. *The data (p_r, x_r), $r = 1, ..., N$ satisfies "cyclical consistency", that is,*

$$p_r x_r \geq p_r x_s, \ p_s x_s \geq p_s x_t, \cdots, p_q x_q \geq p_q x_r$$

   implies

$$p_r x_r = p_r x_s, \ p_s x_s = p_s x_t, \cdots, p_q x_q = p_q x_r$$

Definition 2 Let (p_r, x_r), $r = 1, ..., N$ be given. The data is *quasilinear rationalizable* if for some $y_r > 0$ and $I > 0$, $\forall r$ x_r solves

$$\max_{x \in \boldsymbol{R}^n_{++}} \quad U(x) + y_r$$
$$\text{s.t.} \quad p_r x + y_r = I$$

where U is a concave function.

Definition 3 A demand function $x(p, I) : I\!R^{n+1}_{++} \to I\!R^n_{++}$ is *cyclically monotone* if for any given I and finite set $\{p_1, ..., p_m\}$ (m arbitrary):

$$x_1 \cdot (p_2 - p_1) + x_2 \cdot (p_3 - p_2) + \cdots + x_m \cdot (p_1 - p_m) \geq 0$$

This is equivalent to $p(x, I)$ being cyclically monotone, that is that for any given I and finite set $\{x_1, ..., x_m\}$ (m arbitrary)

$$p_1 \cdot (x_2 - x_1) + p_2 \cdot (x_3 - x_2) + \cdots + p_m \cdot (x_1 - x_m) \geq 0$$

Definition 4 A demand function $x(p, I) : I\!R^{n+1}_{++} \to I\!R^n_{++}$ satisfies the *strong law of demand* if it is cyclically monotone.

Definition 5 If u is concave on $I\!R^n$, then $\beta \in I\!R^n$ is a subgradient of u at x if for all $y \in I\!R^n : u(y) \leq u(x) + \beta \cdot (y - x)$.

Definition 6 If u is a concave function on R^n, then $\partial u(x)$ is the set of subgradients of u at x.

Theorem 4.2 (Rockafellar 1970). *Let ρ be a multivalued mapping from $I\!R^n$ to $I\!R^n$. In order that there exists a closed proper concave function f on $I\!R^n$ such that $\rho(x) \subset Df(x)$ for every x, it is necessary and sufficient that ρ be cyclically monotone.*

In his proof he constructs the following function f on $I\!R^n$:

$$f(x) = \inf\{x_m^* \cdot (x - x_m) + \ldots + x_0^* \cdot (x_1 - x_0)\}$$

where the infimum is taken over all finite sets of pairs $(x_r^*, x_r), r = 1, \ldots, m$ (m arbitrary) in the graph of $\rho(x)$. Note that if the graph of ρ has only a finite number of elements then the domain of f is all of R^n. Also note that for this f, x_i^* is the subgradient at x_i.

Theorem 4.3. *The following are equivalent:*

1. *The data (p_r, x_r), $r = 1, ..., N$ is quasilinear rationalizable by a continuous, concave, strictly monotonic utility function U.*

2. *The data (p_r, x_r), $r = 1, ..., N$ satisfies Afriat's inequalities with constant marginal utilities of income, that is, there exists $G_r > 0$ and $\lambda > 0$ for $r = 1, \ldots, N$ such that*

$$G_r \leq G_l + \lambda p_l \cdot (x_r - x_l) \;\; \forall r, l = 1, \ldots, N$$

or equivalently there exist $U_r > 0$ for $r = 1, \ldots, N$

$$G_r \leq G_l + p_l \cdot (x_r - x_l) \;\; \forall r, l = 1, \ldots, N$$

where $G_r = \frac{U_r}{\lambda}$

3. *The data* (p_r, x_r), $r = 1, ..., N$ *satisfies the strong law of demand.*

PROOF:

1. (1) ⇒ (2): From the FOC of the quasilinear utility maximization problem we know:

$$\exists \beta_r \,\epsilon\, \partial U(x), \text{ s.t. } \beta_r = \lambda_r p_r \text{ where } \lambda_r = 1$$

   Also, U being concave implies that $U(x_r) \leq U(x_l) + \beta_l(x_r - x_l)$ for $r, l = 1, 2, ..., N$. Since $\beta_l = p_l \; \forall l = 1, \ldots, N$ we get $U(x_r) \leq U(x_l) + p_l(x_r - x_l) \; \forall r, l = 1, ..., N$

2. (2) ⇒ (3): For any set of pairs $\{(x_s, p_s)\}$, $s = 1, ..., m$ we need that: $p_0 \cdot (x_1 - x_0) + p_1(x_2 - x_1) + ... + p_m(x_0 - x_m) \geq 0$. From the Afriat inequalities with constant marginal utilities of income we know:

$$\begin{aligned} U_1 - U_0 &\leq p_0 \cdot (x_1 - x_0) \\ U_2 - U_1 &\leq p_1 \cdot (x_2 - x_1) \\ &\cdots \\ U_0 - U_m &\leq p_m(x_0 - x_m) \end{aligned}$$

   Adding up these inequalities we get that the left hand sides cancel and we get the condition that defines cyclical monotonicity.

3. (3) ⇒ (1): By Theorem 4.2 we know there exists a utility function U such that p_r is

the subgradient of U at x_r, that is:

$$p_r = \partial U(x_r)$$

which together with $\lambda_r = 1$ constitutes a solution to the first order conditions of the quasilinear maximization problem.

If we require strict inequality in (2) of Theorem 4.3, then it follows from Lemma 2 in Chiappori and Rochet (1987) that the rationalization can be chosen to be a C^∞ function. It then follows from Roy's identity that $x(p) = -\frac{\partial V(p)}{\partial p}$. Hence for any line integral we see that $\int_{p_1}^{p_2} x(p)dp = -\int_{p_1}^{p_2} = \frac{\partial V(p)}{\partial p}dp = h(p_1) - h(p_2)$. That is, consumer welfare is well-defined and the change in consumer surplus induced by a change in market prices is the change in consumer's welfare.

4.3 Random Quasilinear Rationalizations

If consumers have random quasilinear utility functions can anything happen? No, not if each individual's distribution of utility shocks to her marginal utilities are bounded, where agents have random utility functions of the form $V(x, e) = U(x) + \epsilon \cdot x + x_0$. Assuming $U(x)$ is strictly concave, smooth and monotonic, each realization of ϵ gives rise to a quasilinear utility function having all of the properties previously derived, i.e. for a fixed ϵ, the random demand function $x(p, \epsilon)$ is cyclically monotone. Of course a finite family of observations of such demand functions need not be cyclically monotone, since each observation can in principle be drawn from a "different" cyclically monotone demand function. It is therefore surprising that the hypothesis of random quasilinear rationalization of a data set is refutable, if the shocks are bounded and the bounds are known a priori. As a consequence, this hypothesis is testable in the sense of Brown and Matzkin (1995), i.e. there exists a finite family of polynomial inequalities involving only observations on market data that are solvable if and only if the data can be rationalized with a random quasilinear utility function. These inequalities will involve the known bounds on the shocks.

Definition 7 $U(x) + \epsilon \cdot x + x_0$ is a random quasilinear rationalization of the data (p^r, x^r), $r = 1, ..., N$ if $\exists \epsilon_1, \ldots, \epsilon_r$, x_{0r} and $I > 0$ such that x_r is the solution to

$$\begin{array}{ll} \max & U(x) + \epsilon_r \cdot x + x_{0r} \\ \text{s.t.} & p_r \cdot x + x_0 = I \end{array}$$

Definition 8 The data set (p_r, x_r), $r = 1, ..., N$, satisfies the Random Afriat Inequalities, RAI, if it satisfies the Afriat inequalities for a random quasilinear utility function of the form $V(x, \epsilon) = U(x) + \epsilon \cdot x + x_0$. That is, if they satisfy the following:

$$\begin{array}{rcl} U_r & \leq & U_l + (p_l - \epsilon_l) \cdot (x_r - x_l) \text{ for } r, l = 1, \ldots, N \\ p_r - \epsilon_r & \gg & 0 \quad \text{for } r = 1, \ldots, N \\ U_r & > & 0 \quad \text{for } r = 1, \ldots, N \\ \epsilon_r & \gg & 0 \quad \forall r = 1, \ldots, N \end{array}$$

These are linear inequalities in the unknown U_r and ϵ_r. Hence they can be solved in polynomial time using interior point linear programming algorithms.

The utility shock ϵ has compact support if there exists ϵ_{min} and ϵ_{max} such that $\epsilon_{min} \leq \epsilon \leq \epsilon_{max}$, where the quasilinear model is a special case of the random quasilinear model if $\epsilon_{max} = 0$.

Figure 4.1 shows that for two observations, all possible pairs of budget lines defined by the gradients of $U(x)$ at x_1 and x_2, given consumption x_1 and x_2, violate WARP. Hence rationalizations with random quasilinear utilities of the form $U(x) + \epsilon \cdot x + x_0$, where ϵ has compact support is refutable.

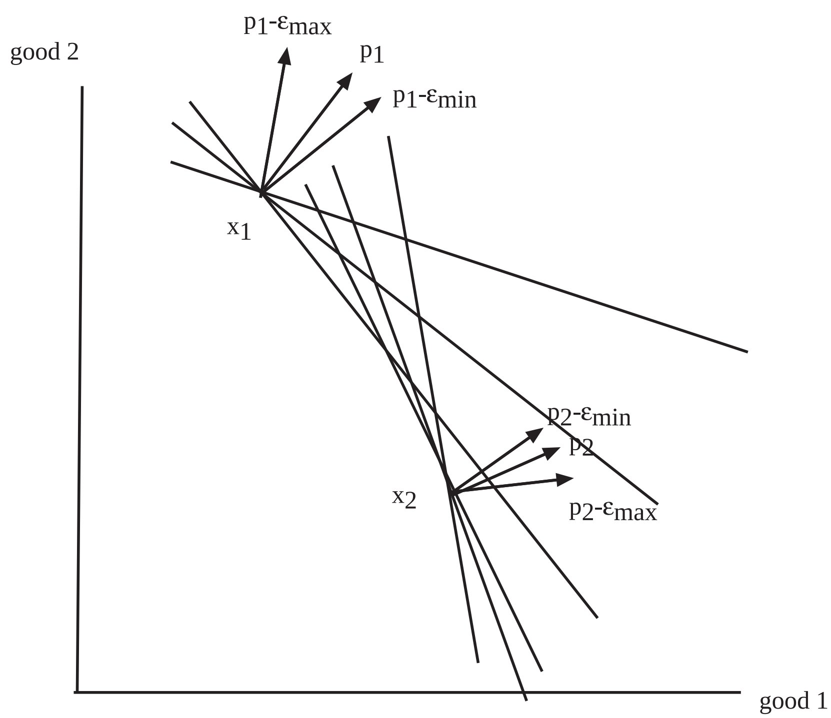

Figure 4.1: Two goods two observation example with bounded shocks

Figure 4.2 shows that for two observations, without bounds on the shocks, the model is not refutable. With unbounded shocks we can generate the interior of the positive orthant of the price vector and therefore WARP will never be violated.

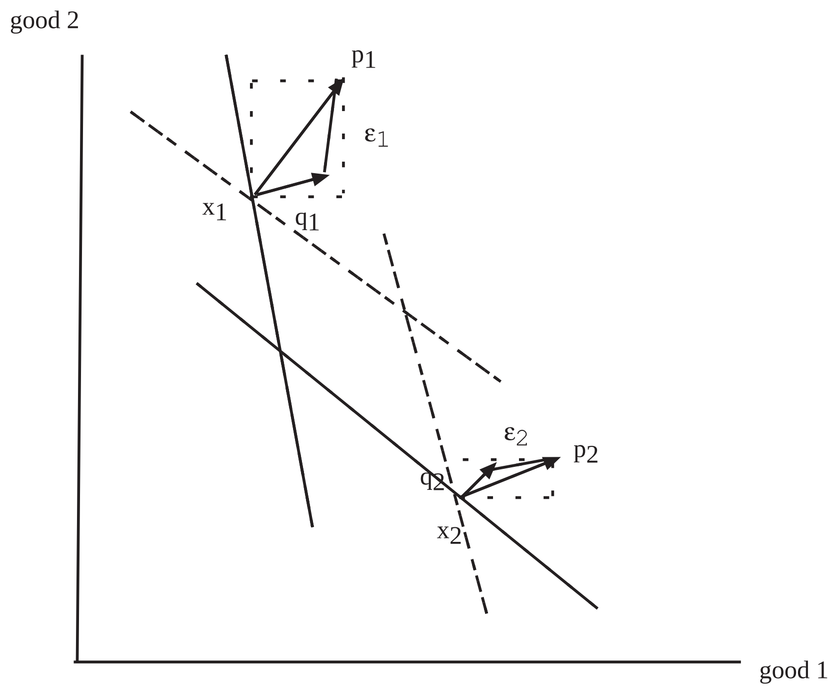

Figure 4.2: Example with unbounded shocks, $\epsilon_r \leq p_r$ and a pair of budget sets where consumption choices do not violate WARP.

Theorem 4.4. *Let the data set* (p_r, x_r), $r = 1, ..., N$ *be given. The data can be rationalized by random quasilinear preferences of the form* $U(x) + \epsilon \cdot x + x_0$ *if and only if it satisfies RAI. Moreover, the model is refutable if and only if the shocks are bounded and the bounds are known a priori.*

This result can be viewed as a relaxed version of Theorem 4.3. If one empirically tests for quasilinear utilities it would be extremely surprising to find that the Afriat inequalities for $\lambda_r = 1\ \forall r$ were satisfied, i.e. that a quasilinear function rationalizes the data. As in any empirical test we would like to allow for some error in the individual's choice.[2] One way of doing this is to say that individuals make small mistakes in evaluating their marginal

[2] Afriat (1972) and Varian (1990) provide the Critical Cost Efficiency Index as a measure of goodness-of-fit that is measured by the minimal distortion in wealths required to rationalize the data. This is the measure used by Andreoni and Miller (2002) to rationalize altruism with experimental data.

utilities when making a consumption choice. This suggests the following family of Afriat inequalities with q's and p's:

$$U_r \leq U_l + q_l \cdot (x_r - x_l) \qquad \forall r, l = 1, \ldots, N$$

$$q_r \leq p_r \qquad \forall r = 1, \ldots, N$$

Note that if we define $\epsilon_r = p_r - q_r$, this reduces to the family of inequalities for random quasilinear utilities. We now compute the solutions to the set of inequalities with unknowns U_r, q_r for $r = 1, \ldots, N$ and pick the solution with least absolute deviation (LAD). This solution is our "best" approximation.

Our notion of curve-fitting demand data with quasilinear utilities extends naturally to the wide class of economic models formulated as solutions to a finite family of polynomial inequalities consisting of the Afriat inequalities for consumers or producers and possibly other polynomial inequalities such as budget constraints or feasibility constraints, e.g. market clearing. We simply relax the Afriat inequalities for agents by introducing slack variables for the gradients of the utility and production or cost functions. In particular, if we apply our relaxation method to the Brown-Matzkin inequalities characterizing a Walrasian equilibrium, we obtain the "best" Walrasian approximation to the observed market data with respect to minimizing least absolute deviations. It is in this sense that economists may use competitive markets as approximations to actual market structures.

Bibliography

Afriat, S.N., 1972. "Efficiency Estimates of Production Functions", *International Economic Review*, 13, 568-598.

———, 1981. "On the Constructability of Consistent Price Indices Between Several Periods Simultaneously", in Deaton (ed.), *Essays in Applied Demand Analysis*. Cambridge, Cambridge University Press.

Andreoni, J. and Miller, J., 2002. "Giving According to GARP: An Experimental Tests of the Consistency of Preferences for Altruism", *Econometrica*, **70**, 737-753.

Arneson, R., 1989."Equality and equal opportunity for welfare", *Philosophical Studies* 56, 77-93

Arrow, K., Bowles, S. and Durlauf, S., 2000, *Meritocracy and Economic Inequality*, Princeton University Press. Princeton, New Jersey.

Benabou, R. and Ok, E. 2001. "Mobility as Progressivity: Ranking Income Prodesses According to Equality of Opportunity", NBER Working Paper No. W8431.

Betts, J.,1995. "Does school quality matter? Evidence from the Nationals Longitudinal Survey of Youth". *The Review of Economics and Statistics*, 77.

Betts, J. and Roemer, J., 2002. "Equalizing opportunities through educational finance reform" (working paper)

Bowles, S. and Gintis, H., 1998. "Is equality passé?", *Boston Review*.

Brown, D.J. and Calsamiglia, C.,2003."The Strong Law of Demand", Cowles Foundation Discussion Paper No.1399.

Brown, D.J. and Matskin, R.,1996. "Testable Restrictions in the Equilibrium Manifold", *Econometrica*, **64**, 1249-1262.

Brown, D.J. and Matskin, R. (1998), "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand", Cowles Foundation Discussion Paper, No.1175, Yale University.

Brown, D.J. and Wegkamp, M.H. (2003), "Test of Independence in Separable Econometric Models " Cowles Foundation Discussion Paper No. 1395.

Cavanagh, M., 2002. *Against Equality of Opportunity*, Oxford University Press.

Chiappori, P.A. and Rochet, J.C., 1987.“Revealed Preferences and Differentiable Demand”, *Econometrica*, 55, 687-691.

Cohen, G.A., 1989. “On the currency of egalitarian justice”, *Ethics* 99, 906-944.

Connolly, S., J.Micklewright and S.Nickell, 1992. ”The occupational success of young men who left school at sixteen”, *Oxford Economic Papers*, 44.

Dworkin, R., 1981a. “What is equality? Part 1: Equality of welfare”, *Philosophy and Public Affairs* 10, 185-246.

——, 1981b. “What is equality? Part 2: Equality of resources”, *Philosophy and Public Affairs* 10, 283-345.

Elster, J., 1992. *Local Justice*, Russell Sage Foundation.

Fogelman, K.,1983. *Growing up in Great Britain: Papers from the National Child Development Study*, London: MacMillan.

Fogelman, K. 1976.*Britain's Sixteen-Year-Olds*, London: National Children's Bureau.

Harmon, C. and I.Walker, 1998. ”Selective schooling and the returns to the quantity and quality of education”, Dublin: Department of Economics at University College Dublin.

Hutchinson, D., H.Prosser and P.Wedge, 1979. ”The prediction of educational failure”, *Educational Studies*,5.

Konow, J. 2003. “Which is the fairest one of all? A positive analysis of justice theory”, *Journal of Economic Literature*, Vol. XLI, pp. 1188-1239.

Llavador, H. and Roemer, J., 2001. ”An equal-opportunity approach to international aid,” *Journal of Development Economics* 64, 147-171.

Quah, J., 2000. “The Monotonicity of Individual and Market Demand”, *Econometrica*, 68, 911-930.

Robertson, D. and J.Symons.”Do peer groups matter? Peer group versus schooling effects on academic attainment”. London: Center for Economic Performance, London School of Economics, Working Paper 311, 1996.

Rockafellar, R.T.,1970. *Convex Analysis*, Princeton, NJ: Princeton University Press.

Roemer, J., 1998. *Equality of Opportunity*, Harvard University Press

——, 2002. ”Equality of opportunity: A progress report,” *Social Choice and Welfare*, Vol. 19 (2) pp. 455-471.

——, Aaberge, Colombino, Fritzell, Jenkins, Marx, Page, Pommer, Ruiz-Castillo, San Segundo, Tranaes, Wagner, and Zubiri, 2003. ”To what extent do fiscal systems equalize opportunities for income among citizens?” *Journal of Public Economics* 87, 539-565

Sen, A. 1992. *Inequality Reexamined*, Cambridge, Mass.: Harvard University Press.

Varian, H.,1983. "Non-parametric Test of Consumer Behavior", *Review of Economic Studies*, 99-210.

Varian, H.,1990. "Goodness-of-fit-Optimizing Models" , *Journal of Econometrics*. Number 13.

Young, H.P., *Equity: In Theory and Practice*, 1994. Princeton University Press, Princeton, New Jersey