Società italiana di economia pubblica
XV CONFERENZA

DIRITTI, REGOLE, MERCATO
Economia pubblica ed analisi economica del diritto

Pavia, Università, 3 - 4 ottobre 2003

A STATISTICAL APPROACH TO FREEDOM

M. D'AGOSTINO, V. DARDANONI, V. PERAGINE

DIRITTI, REGOLE, MERCATO
Economia pubblica ed analisi economica del diritto

XV Conferenza SIEP - Pavia, Università, 3 - 4 ottobre 2003

pubblicazione internet realizzata con contributo della COMPAGNIA di San Paolo

società italiana di economia pubblica

dipartimento di economia pubblica e territoriale – università di Pavia

A Statistical Approach to Freedom

M. D'Agostino (University of Ferrara)
V. Dardanoni (University of Palermo)
V. Peragine (University of Bari)

March 26, 2003

Abstract

This paper is concerned with the appraisals of specific freedoms. We adopt an approach which is empirical, in the sense that appraisals of freedom are tested against observations; and statistical, in that such tests concern large populations rather than single individuals. We argue that this methodology is based on sound theoretical grounds, and we show by means of an empirical application that it is actually useful for operational purposes.

PRELIMINARY DRAFT

1 Introduction

In this paper we propose a model of how hypothetical judgements about freedom can be formed, tested and revised. Our approach is empirical, in the sense that appraisals of freedom are tested against observations; and statistical, in that such tests involve populations and samples rather than single individuals. We argue that this methodology is based on sound theoretical grounds, and we show by means of an empirical application that it is actually useful for operational purposes.

We motivate our approach within the general conceptual framework proposed by MacCallum (1967). Following MacCallum, "every discourse about liberty takes the form of a triadic relationship: an agent (individual or collective) is free from a given set of constraints to choose among a given set of options". Appraisals of freedom are therefore obtained by specifying the set of options, the constraints and the agent they are referred to: z is the agent, y is the option, and x is the constraint.

In this paper we concentrate on *specific* freedoms: the freedom in a society, with respect to a certain class of choices (e.g. freedom of religion, of press, of speech, of movement; freedom to vote, to choose an occupation, to open a shop, to join a trade union etc.). Thus, we assume that y is a variable which indicates the options which are open to the agent z.

Suppose now that the choice variable y takes n mutually exclusive and exhaustive possibilities. To fix the ideas, suppose that y is actually a binary variable that takes on two values, say A and B. For example, suppose that an agent z can choose whether to enter a door which leads him to room A, or enter another door which leads him to room B, but since options are mutually exclusive and exhaustive, he ultimately has to choose one and only one room. Now, z will be unfree to choose if there is a constraint x (a "closed door") which prevents him from entering one of the rooms.

Appraisals of freedom involve discovering the existence and the nature of these "closed doors". Suppose that we, as external observers, cannot directly check whether one of the two doors is actually closed for x, nor can we ask directly z which is his favourite option between A and B. These are situations that we typically face in most interesting cases. We can, however, observe which room is actually entered by z. How can we infer whether z is free or not? Suppose we find z in room A. Clearly there are three possibilities: either (i) z really wanted to be there and, facing no "closed door", he was able to get inside, or (ii) z actually preferred to be in room B and, facing a "closed door", he was forced to get into room A, or (iii) z preferred to be in room A and managed to get in, but the door of room B was closed. While in case (i) we can legitimately say that z was "free" to choose any of the two options, in cases (ii) and (iii) we should conclude that z was not free (since there was at least a closed door).

Now, suppose that the nature of the constraint x is actually related to characteristics of agents: if John Doe faces a closed door, this is so not because he *is John Doe*, but rather because, say, he is tall, or black, or Jewish, or has a low IQ, or has an uncle in the military etc. Suppose now that we find that in room A we have only people with a given characteristic (say having blue eyes), and in room B people without that characteristic. There are then two possibilities: if all people with blue eyes really wanted to enter room A and all people without blue eyes really wanted to enter room B, then we cannot infer the existence of a constraint x, activated on the base of the color of the eyes, in the choice of y; on the other hand, if the color of the eyes was irrelevant for the choice of option y, then we can legitimately infer the existence of a constraint x, activated on the base of the color of the eyes.

Suppose also that we can list a set of individual characteristics that are sufficient to predict the independence between a certain individual characteristic v and the distribution of preferences regarding option y. Suppose, finally, that we find a lack of independence, given the set of all the other relevant characteristics, between the given characteristic v, and the actual choices with respect to option y. Then we can legitimately conclude that there exists a constraint x activated on the base of that characteristic.

Hence, appraisals of freedom involve discovering which characteristics, that agent z has or lacks, are related to constraints to obtaining option y. In general, the link between the possession of a given characteristic and (the constraint to obtain) a given options is practically — and perhaps theoretically — impossible to discover by observing the behavior of a *single* individual.

On the other hand, each feature of an individual implicitely defines a category of individuals — all individuals sharing that feature; then a statistical approach might be implemented to detect the presence of systematic links, in a given society, between characteristics and constraints. In other words, while we are interested in detecting the presence of features that are related to constraints on the access to a given option at the individual level, we can detect them empirically by looking at the statistical (in)dependence between observed choices and observable features of population of individuals.

One of the main points of this paper is that focusing on characteristics that may be related to the existence of constraints on freedom is theoretically and empirically sound. The set of options which are "really" open to an agent may be dependent on a plurality of characteristics: natural factors, social conditions, past choices and actions, etc. Which characteristics, which may be related to the existence of constraints to the availability of y, are "relevant" to the assessment of freedom? Which are not? Where to draw the line?

This is a crucial issue which has been vastly and deeply discussed in the philosophical literature on freedom: see, among others, Berlin 1958, Bobbio 1956, Miller 1983, Oppenheim 1985, Steiner 1974 and Taylor 1979.[1] The question is that of identifying the constraints which are relevant for the definition of freedom: should we consider merely "external" constraints (Berlin, 1958) or also "internal" constraints, as those deriving from the specific desires and preferences of the agent under consideration (Taylor 1979)? Among the external constraints, should we consider only the constraints imposed by other agents (Oppenheim 1985) or also those imposed by nature?

[1] See Carter (2000) for an excellent discussion of most of these topics.

Should threats count as restrictions of freedom (Steiner, 1974)? Should we count all the constraints imposed by other agents or only those for which the other individuals can be held morally responsible[2] (Miller,1983, see also Pettit, 2001)?

Thus, theories of freedom tend to agree on identifying freedom with unconstrained action, but disagree on what constitutes a constraint. According to some theories of freedom, only the existence of other agents that make it physically impossible for z to access some option y would count as a constraint on z's freedom to access y. Other theories will consider a constraint also the existence of other agents that make it highly undesirable for z to access options y (for instance by threatening z). Moving along the "positive" direction, one could consider as a constraint also the fact that z can't afford y; in general, the more "positive" the theory of freedom the larger the set of characteristics that will be regarded as constraints.

In this paper we do not take a specific position about theories of freedom, but we show how they can be embedded into predictive models that have an *empirical content* and can therefore be empirically tested.

2 The Background Problem

Let us start by recalling MacCallum's (1967) triadic relation, "z is free from x to choose y", where z is the agent, y is the option, and x is the constraint.

We can think of an agent z as described by a set of real valued characteristics $(z_1, \ldots, z_n)$, so that z can be construed as an element of R^n. In general, some of these characteristics will be observed, and some will not; we can therefore partion z into (o, u), so that $z = o \cup u$ and $o \cap u = \emptyset$.

On the other hand, the constraint x is activated by the characteristics of the agent, that is, we imagine there is a function $c_y : R^n \mapsto \{0, 1\}$, where if

[2] To use one of the examples above, does the poverty of a tramp who is unable to dine at the Ritz restrict her freedom? One plausible answer is that of arguing that impersonal economic forces cannot reasonably be interpreted as constraints on freedom. That is to say that the "x factor" in MacCallum's formula gets narrowed down to only those obstacles that are imposed deliberately, or that are at least foreseeable, by other individuals. An alternative answer would be that she is unfree only if the fact that she is poor - so the constraints that prevent him to dine at the Ritz - cannot be attributed to her choice, nor to fortune, but to an unjust distribution of resources in the society, and that we can attribute to the society the will that causally determines the situation of poverty. Here, we only notice the link between the theory of freedom and the theory of justice; in particular, we notice the link with the recent literature, in normative economics, on equality of opportunity (Fleurbaey 1995, Roemer 1993, 1998 and Peragine 1999, 2002).

$c_y(z) = 1$, the agent is constrained with respect to option y, while if $c(z) = 0$ the agent is free from constraints.

We follow the mainstream in defining freedom as absence of (some sort of) constraints on the agent's choice. However, as in the room example discussed in the introduction, most often than not we do not have, nor can we possibly have, a perfect and *complete* knowledge of the constraints which are imposed on each individual. We could obtain indirect evidence of the presence of constraints on the action of an agent z from comparing z's actual choice behaviour with z's known preferences, on the basis of the assumption that a free individual usually does not act against his own preferences. But, again, obtaining perfect and complete information about the preferences of each individual in a large population appears to be unfeasible. As in any scientific investigation, we have to replace our lack of knowledge with *hypotheses.* Moreover, if our model has to be an *empirical* one, such hypotheses must be testable against observative or experimental evidence.

So our first point is:

> *In most realistic situations, we do not have, nor can we possibly have, perfect and complete information about the individual preferences and/or the constraints which may act against each individual; hence we have to resort to hypotheses and test them on the basis of observative evidence.*

We can assume that the option y which is actually chosen by z is completely determined by a perfect knowledge of the preferences of z together with a perfect knowledge of all the constraints that may prevent him from obtaining a certain preferred option. So, on the basis of suitable hpotheses about the agent's preferences and the absence of constraints with respect to the preferred option, we can make predictions about the actual option that the agent will take. For this purpose all we need is the following basic inference scheme:

(a) z has a (strong) preference for option y
(b) there are no constraints on z's obtaining option y
―――――――――――――――――――――――――
(c) z will take option y.

This implies, by *modus tollens*, that if z does not take option y, then *either* hypothesis (a) *or* hypothesis (b), or both must be incorrect.

Notice that this is by no means different from what happens in scientific theories, e.g. physical theories, where we *need* auxiliary hypotheses and initial conditions in order to make empirically testable predictions from a theory. Observe also that hypothesis (b), namely the hypothesis that the given agent is "free" to obtain option y, takes the logical form of a theoretical assertion, i.e. a strictly universal one. The assertion "there are no constraints on z's choice of y" is logically on a par with "all ravens are black" or "all bodies obey Newton's gravitation law". All these assertions *forbid* the existence of something (a constraint on z's choice of y, a non-black raven, etc.) in a domain of which we do not have perfect knowledge.

Our second point then is:

> *Judgements about freedom are theoretical, strictly universal, assertions and therefore they (a) can never be verified, but only falsified, and (b) require auxiliary hypotheses in order to make empirically testable predictions.*

However, at a closer look, the very simple prediction model outlined above turns out to be inadequate for (at least) two different reasons. *First*, how are we going to make and test hypotheses about individual preferences? All we can possibly do is to base such hypotheses on a suitable (necessarily finite) description of individuals in terms of their observed characteristics. But this will lead to hypotheses about the preferences of a *class* of individuals, i.e. those sharing the same observed characteristics, and such hypotheses can never lead to deterministc predictions about the preferences of single individuals. All we can achieve is a probabilistic prediction. So, in our model we should replace (a) with (a$'$) "z prefers option y with probability p", and conclusion (c) with (c$'$) "z will take option y with probability p". Clearly, in this case, the observation that z, as a matter of fact, does not choose y has no implication on either (a$'$) or (b).

Second, at the individual level it may well be the case that z is free to take his preferred option but unfree to take some other option that he is not interested in. In this case, we should say that z is not free to choose among the options, although this does not affect his actual choice. However, in such a situation, the individual prediction model fails to reveal a lack of freedom, even in the case in which we have perfect information about z's preferences.

These difficulties appear to be the main motivation of the "statistical concept of freedom" outlined in Gabor and Gabor (1954), as can be inferred from the following passage:

> All acts of individuals which we associate with the adjective "free" can be considered as choices from a number of alternatives. Before the act, an outside observer can enumerate the possibilities from his necessarily imperfect knowledge of restraints, and he can assess to them probabilities, from his even less perfect knowledge of the psychology

> of the individual. Actually, only one of these possible acts will materialise, and the *a priori* latitude which the observer has granted to the individual appears only as a measure of his own ignorance. Thus the objective approach is useless when applied to the single individual, while a "subjective" approach, based on his own estimate of his wishes, fancies and frustrations can hardly lead to an agreed numerical measure, at any rate, not at the present stage of experimental psychology.

Rather optimistically, the Gabors' do not hesitate to make the sweeping claim that "[their] objective statistical approach, which takes account only of acts, *post facto*, and which considers the individual only as a member of a large population, overcomes these obstacles at one stroke." Their methodological tenet to ban any hypothesis about preferences, as incurably "subjective", leads them to a philosophically naive model which has been occasionally criticisized explicitly, for instance by Carter (1993), but largely ignored by the mainstream literature.

In this paper we attempt to revive the statistical approach, taking its motivation seriously, but embedding it in what we hope is a more sophisticated model, taking into account some of the main issues concerning the testing and revision of scientific hypotheses which are vastly and deeply discussed in the philosophy of science literature.

3 The Model

In order to do this we need to introduce some statistical notation; useful discussions of the topics involved are contained in Dawid (1979, with discussion), Holland (1986, with discussion), Stone (1993), Greenland, Robins and Pearl (1999) and Dawid (2001, with discussion).

Given two random variables U and V, $f(u,v)$, $f(u)$ and $f(u|v)$ will denote respectively the joint, marginal and conditional densities. We let $U \perp V$ denote that U is independent on V, that is, $f(u|v) = f(v)$; in words, U is independent on V if knowing the value of V does not add any information about U[3]. Similarly, given three random variables U, V and W, we let $U \perp V | W$ denote that U is independent on V conditionally on W, that is, $f(u|v,w) = f(u|w)$; in words, U is conditionally independent on V given W if knowing the value of V does not any add new information on U after

[3]This formulation of independence is actually equivalent to the usual condition that $f(u,v) = f(u)f(v)$.

knowing W, that is, V is superfluous once W is given.[4]

Now, let X be an observed characteristic, and Z a set of observed characteristics describing the reference population. What is the effect of a change in X on the choice of Y?

First a change in X may affect the agent's *preferences*. The assumption that this is not the case, i.e. that the value of X has no effect on the agent's preferences can be expressed as follows:

(PI) $\quad W \perp X | Z$

where W is a random variable taking the same values as Y and describing the "preferred" options, as opposed to the options actually taken by the individual agents. We call this assumption (PI) for "Preference Independence" assumption.

Second, a change in X may also affect the *constraints* that may be activated on the agent's choice of an option $y \in Y$. Let C_y be a variable taking as value 1 if the agent is constrained in his choice of the option $y \in Y$, and 0 otherwise. The no-effect assumption, in this case, takes the following form:

(CI) $\quad \forall y,\ C_y \perp X | Z$

We call this assumption (CI) for "Constraint Independence", since it says that the value of X has no effect on the distribution of constraints (with respect to options Y) in the population Z.

Observe that (CI) is a "relative freedom" hypothesis, saying that the agents' freedom to obtain any of the options in Y is independent of the characteristic X.

The main "Freedom Hypothesis" (FH) is then expressed by the following statement:

(FH) $\quad \forall X \notin Z, \forall y,\ C_y \perp X | Z.$

Clearly (FH) implies (CI) for every variable X. (See below for a brief discussion of degenerate cases.)

Finally, that the *actual* behaviour of the reference population with respect to the options Y is independent of characteristic X, is expressed by the following statement:

(OI) $\quad Y \perp X | Z.$

[4]An example which shows the difficulties involved in analyzing statistical independence is contained in Appendix A.

We call this statement (OI) for "Observed Independence".

Now, assuming that Y depends only on W and C, it can be shown that:

Theorem 1 *Under (CI), (PI) if and only if (OI)*

Now, (OI) is an empirically testable prediction about observed behaviour which is related, via (PI), to the "relative freedom" hypothesis (CI) and hence to the main freedom hypothesis (FH), in the sense that if (OI) does not hold, then *either* (PI) does not hold *or* (CI), and therefore (FH), does not hold.

The importance of this theorem, then, is that it connects the hypothesis (FH), via the auxiliary hypothesis (PI), to an empirical prediction (OI) that can be tested by means of standard statistical methods (for an example of how to test conditional independence and a flavour of the techniques involved see Appendix B).

Let us now discuss briefly two degenerate cases which may arise in connection with hypothesis (FH). First, observe that (FH) is still true in the case in which one or more constraints C_y apply *uniformly* to all members of the population described by Z. Although this may appear paradoxical, it must be remarked that such a situation would amount, in fact, to a restriction of the option set: an option which is unavailable to every member of the population is not, in some sense, a real option. Of course, under certain conditions, such a restriction of the option set can also be interpreted as a restriction in freedom. From this point of view, freedom can be analysed along two distinct dimensions: (i) the richness and diversity of the options, and (ii) the existence of constraints on their availability for some members of the population. Our model addresses dimension (ii) of the problem and is not intended as a contribution to the analysis of dimension (i), for which we refer the reader to the discussion in van Hees (2001). In any case, observe that the distinction between dimensions (i) and (ii) tends to vanish as the population under investigation grows and, in the limiting case, coincides with the "whole world". In this limiting case, an "option" which is unavailable to every inhabitant of the world can, perhaps, be regarded as not being a real option.

A second degenerate case concerns a situation in which constraints are randomly distributed within the population described by Z. In such a case, our model is inapplicable since we are not able to establish dependence of constraints on any characteristic X such that the subpopulation identified by any specific value of X is large enough for a meaningful statistical analysis[5], and therefore the freedom hypothesis (FH) cannot be refuted. This

[5]Indeed, if the constraints are *really* randomly distributed, then dependence can be

limitation simply reflects the fact that the statistical concept of freedom outlined here, does not apply to single individuals but only to (sufficiently large) populations.

4 Dynamics aspects of the model: the revision process

What if the prediction (OI) is rejected? Since (OI), like any interesting scientific prediction, does not depend on an isolated hypothesis but on a *prediction model* that includes a *system* of hypotheses — in our simplified model it depends on the two hypotheses (PI) and (CI) — we are left with the problem of deciding which of these hypotheses should be rejected (observe that this is what normally happens in testing any genuine scientific hypothesis). Should we reject (CI), and therefore our main freedom hypothesis (FH), and conclude that in the given population there is no free access to options Y? Or should we reject the preference indipendence hypothesis (PI)? Suppose we want to save (CI) from refutation, then we have to show how to revise our prediction model by replacing (CI) with a better version (CI′). How can we envisage such a new hypothesis (CI′)?

A well-known problem of inferences based on statistical independence is related to the so-called Simpson Paradox (for which see Appendix A). In the context of conditional independence, the paradox consists in the fact that adding a variable to the conditions may reverse the truth-value of independence statements, i.e. if $Y \perp X|Z$ is false (true), it may well be that adding a new variable V to the conditions, we find that $Y \perp X|Z, V$ is true (false).

As far as our model is concerned, this implies that a "refutation" of the freedom hypothesis (FH) is never conclusive: in the presence of contrary statistical evidence (to the effect that $Y \not\perp X|Z$) we can always save (FH) by making the "rescue hypothesis" that there exists a yet unobserved variable V which is "confounding", in the terminology of Stone (1993), that is V may affect the distribution of actual choices Y, by affecting the distribution of preferences W, and show dependence on X. This means that we may have neglected a "relevant" condition in our preference independence hypothesis (PI), and want to replace it with a revised version (PI′) $= W \perp X|Z, V$. If we can then show that $Y \perp X|Z, V$, this kind of revision would restore con-

"observed" only on variables X that are characteristic functions of single individuals, since that the constraints act at random simply means (by the Kolmogorov-Chaitin theory of information) that no description of the set of elements to which they apply can be shorter than the enumeration of the elements themselves.

sistency with the observed data. If not, we can alway make the hypothesis that some other "hidden variable" is in action. A conclusive refutation of (CI) would be possible only we assumed that there is a "natural endpoint" to the process of taking new yet unobserved variables into consideration, i.e. that we have considered *all* the possible "relevant" variables that may significantly influence the results of our test. Such an assumption takes different forms in the statistical literature: the "covariate sufficiency" hypothesis of Dawid (1979), or the "no confounding" hypothesis of Stone (1993). But there is no such *natural* endpoint and whether accepting these assumptions or not is a matter of conventional decisions. Indeed, the revision process of a predictive model, such as the one presented here, is alway the result of methodological decisions and the "heuristic value" of such decisions is a question to which all the considerations made in the philosophy of science literature can be applied. Some decisions may be regarded as *ad hoc* manouvres, *regressive* revisions that do not lead to any new testable prediction and so reduce the empirical content of (FH); others may appear to be *progressive* revisions in that they lead to *new* testable (and possibly successful) predictions. The main point we want to stress here is that the dynamics of making hypotheses about freedom, embedding them into a predictive model, testing and revising this model in the light of contrary evidence, need not, and perhaps *should not*, be different from what we learn from the scientific practice of mature science.

5 Concluding Remarks

In this paper, we have shown how different theories of freedom can be embedded into models with predictive power, i.e. models that have an empirical content and can therefore be tested.

We have proposed to detect the possible presence of constraints on the access of an individual to a given option by looking at the statistical (in)dependence between observed choices and observable features of population of individuals. By looking at statistical dependencies, we can infer the presence of potential constraints that are unkown or even not directly observable. As a matter of fact, given the complexity of the interaction among individuals, we do not expect individual constraints to be easily observable. In fact, to appreciate the presence of constraints on single individuals, one should have detailed information about the process of choice. For, to make some options, which are only formally available, genuine opportunities open to individuals it is necessary that the act of choosing is not restricted by external pressures or influences. But it is extremely difficult to have quantitative data about the presence of social pressures and influences and the process leading to the final choices; on the contrary, excellent data can be obtained on group dependencies, and constraints which are not observed or observable can be *inferred* from these data.

Moreover, we have attempted to illustrate the dynamical aspects of our model. In testing freedom hypotheses, exactly as with any other scientific hypothesis, refutations are never conclusive, but what we learn in the process of trials and errors may considerably help us improving and revising our believes.

We stress that our approach intends to be a general framework for testing freedom hypotheses on the basis of their empirical content. Not only does is it not committed to any particular conception of freedom (although it is based on the largely agreed concept of freedom as absence of constraints), but as a matter of fact it does not even need to refer to *human* agents, choices and constraints: any physical system with a non-deterministic behaviour where we can suitably interpret the option variable Y, the preference variable W and the constraint variables C_y would be an acceptable domain of application.

If we focus on social and political applications, different "political programmes" are characterized by the methological decisions concerning the characteristics that should be regarded as "relevant" or "irrelevant", intended as conditions for free access to a certain option. Moreover, by specifying the set of characteristics which should be irrelevant to the access to a certain option, a theory of freedom may constitute the basis of a public policy which could help reducing the presence of constraints related to such characteristics.

From this point of view, the kind of analysis outlined in this paper may be interestingly related to the studies on discrimination. In this area, public intervention could follow two different principles (Roemer, 1998): the first, called the "nondiscrimination principle" (Roemer, 1998), states that "irrelevant" characteristics should not count as admissible criteria for the access to options and positions in a society. The second says that society should "level the playing field" among individuals, so that all those with the same "relevant" characteristics will eventually have the same access to the options. For instance, if race, as a characteristic, is deemed irrelevant as an attribute for accessing a certain class of jobs and, neverthless, we observe that race is "empirically" relevant for accessing to such a job, then a public intervention is justified. The nondiscrimination principle in this case states that race should not count for or against a person's eligibility for that job. According to the "level playing field" principle, the public policy should take the form of an affirmative action: for instance, by spending more educational resources, per capita, on black children, in order to compensate for the constraints they will face, later on, in the competition for the job. Notice that, in the case of the "level playing field" principle, an important empirical issue is involved. For, to justify race-based affirmative action policies it is not enough to state that, on a normative basis, race should be irrelevant with respect to the access to the job; such a policies are based on the convinction that race, as a characteristic, is strongly related to access to jobs and positions. In fact, this policies can be criticized on the empirical ground: a complaint could be made, for instance, by saying that class background is a more relevant characteristics than race with respect to the existence of constraints on the access to the relevant jobs. For instance, according to Roemer (1998), this empirical argument justifies the current white working-class backlash against U.S. affirmative action.

However, it is clear that, while the statistical methodology we propose can help to suggest the possible presence of constraints, the actual identification of constraints that are "ethically relevant" for the assesment of freedom, and the role of public policies, if any, which could help reducing the presence of such constraints, is beyond the scope of this paper.

Appendix A. Simpson's Paradox

Suppose Y is a binary random variable which takes on two values: Vote (V) and Not Vote (NV). Suppose that voting behavior is determined by the following two binary characteristics:

race = While (W), Black (B);

area or residence = Village (V), City (C).

Suppose there are 300 individuals in this population, and the distribution of Y and race is given by the following table:

	Vote	*Not Vote*
W	*120*	*90*
B	*40*	*50*

By inspecting the table, it emerges that Y and race are not independent. In particular, the odds of voting are greater for whites than for blacks.

However, we have assumed that voting behavior is actually determined by both race *and* area of residence. Now, suppose that the complete description of this population is given by the following table:

Villages

	V	NV
W	20	40
B	20	40

City

	V	NV
W	100	50
B	20	10

Inspection of the table then shows that, conditionally on area of residence, voting behavior is actually independent on race! This may be deduced by looking at the conditional distributions of voting behavior for blacks and whites: in the village, blacks and whites have identical distribution of voters and non voters, and in the city both blacks and whites are twice as likely to vote rather than not vote. On the other hand, by merging the two tables into a single one, as done in the table 1 above, we get the impression that blacks tend to be less inclined to vote than whites simply because of the disproportion of blacks living in the village. This example is an instance of the so called Simpson paradox, which, in its simplest terms, can be interpreted as saying that a given set of data can exhibit at the same time positive (negative) marginal dependence and non-positive (non-negative) conditional dependence between two variables.

Appendix B. Testing observed independence: an example

In this appendix we illustrate how statistical technniques can be used for testing conditional independence by means of an example.

Our example concerns the existence of freedom-constraints on the choice of occupation. We have data on 36.000 families from the General Social Survey (USA) in the years 1970-1998. The variable Y takes on three values: 0 indicates that the individual has an occupation with low social status, as measured by the Siegel social prestige scale; 1 indicates that the individual has an occupation with middle social status; and 2 indicates that the individual has an occupation with high social status.[6]

The set of observed characteristics is composed of the following variables:

[6] Such an analysis is closely realted to the studies on discrimination and segregation in the labor economics literature, mainly devoted to detect the presence of race and gender differentials in the labor market. These studies try to detect the presence of persistent wage gaps (in the case of "economic discrimination") or employment differential (in the case of segregation) due to factors like gender and race (for recent surveys see Altonji and Blank, 2002 and Cain, 2002). The metodhologies they adopt to dectect the presence of causal links between variable is, however, different from ours.

- fathers' occupation, with values 0 when the individual's father had an occupation with low social status; 1 indicates that the individual's father had an occupation with middle social status; and 2 indicates that the individual's father had an occupation with high social status;

  education, with values 0 to 6 depending on the number of years of education of the individual;

  sex, with values 0 (male) or 1 (female) depending on the sex of the individual;

  age, with values 0 (18 to 35); 1 (35 to 50); 2 (50 to 65) depending on the age of the individual;

  race, with values 0 (white) or 1 (non white) depending on the race of the individual;

  immigration status, with values 0 (born in USA) or 1 (not born in USA) depending on the immigration origin of the individual;

  family size, with values 0 (0 to 2); 1 (2 to 5); 2 (more than 5) depending on the number of siblings in the family of the individual.

Let Y be occupational social status and $(Z_1, \cdots, Z_4$ be race, sex, father's status and immigration. Suppose we want to test the following independence statements:

> (OI$_1$): Y is independent on race conditionally on fathers' occupation, education, sex, age, immigration status and family size.
>
> (OI$_2$): Y is independent on sex conditionally on fathers' occupation, education, race, age, immigration status and family size.
>
> (OI$_3$): Y is independent on father's occupation conditionally on race, education, sex, age, immigration status and family size.
>
> (OI$_4$): Y is independent on immigration conditionally on fathers' occupation, education, sex, age, race and family size.

To test condition OI_i, in order to avoid to make unnecessary functional form and distributional assumptions typical of regression analysis, we can use a simple chi-square test of independence within each 3×2 table which arrays Y-X after conditioning. For example, given that fathers' occupation, education, sex, age, immigration status and family size take on $3 \times 6 \times 2 \times 3 \times 2 \times 3 = 648$ values, a nonparametric test of (OI$_1$) is simply a (joint) test of independence of Y and race within each of the 648 tables obtained after conditioning.

Dardanoni and Forcina (1998) and Bartolucci, Forcina and Dardanoni (2001) propose tests of conditional independence against the alternative of various relevant notions of positive dependence. For our purposes, the interesting notion of dependence is that of *positive quadrant dependence*, which states that, in comparison with the case of independence, two variables are positively quadrant dependent if there is a greater probability that higher values of one variable are coupled with higher values of the other. This notion has been applied to social mobility measurement in Dardanoni (1993). The advantage of this testing procedure is that it tests the null hypothesis of independence against a *specific* positive dependence hypothesis, thus significantly increasing the power of the test in comparison to the standard chi-square procedure, and also indicating the direction of the departure from independence if the null is rejected.

Following Bartolucci, Forcina and Dardanoni (2001), we estimate the model by maximum likelihood techniques, and carry a likelihood ratio test of the null hypothesis of conditional independence (OI) against the alternative of positive dependence. Bartolucci, Forcina and Dardanoni (2001) show that the appropriate test statistic is asymptotically distributed as a chi-bar-squared variate, and critical values can be obtained by normal approximation.

The results of our tests indicate that the null hypotheses OI_1, OI_2 and OI_3 are rejected in favour of conditional positive dependence, while OI_4 is actually not rejected. In words, we find that, conditionally on all the other variables, Y is positively dependent on race, sex and father's status but is independent on immigration[7].

References

[1] Altonji J. and Blank R. (1999) Race and gender in the labor market. In Ashenfelter O. and Card D. (Eds) Handbook of Labor Economics, Vol 3, Elsevier Science, pp. 3143-3259.

[2] Bartolucci F., Forcina A. and Dardanoni V. (2001), ''Positive quadrant dependence and marginal modelling in two-way tables with ordered margins", Journal of the American Statistical Association, vol. 96, 1497-1505.

[7] In this paper we are focusing on the role of this tests with respect to the acceptance or rejection freedom hypotheses. However, the same framework can be used to compare the causal effect of different characteristics on the same choice, or to compare two different societies in terms of freedoms (either wiht a partial or a complete ranking).

[3] Berlin I. (1958) Two Concepts of Liberty, reprinted in in Four Essays on Liberty, Oxford, Oxford University Press, pp. 118-72

[4] Bobbio N. (1956) 'Della libertà dei moderni comparata a quella dei posteri', in Politica e cultura, Torino, Einaudi, pp. 160-94

[5] Cain G. (1999) The economic analysis of labor market discrimination: a survey. In Ashenfelter O. and Layard R. (Eds) Handbook of Labor Economics, Vol 1, Elsevier Science, pp. 693-785.

[6] Carter, I. (1993) "The Measurement of Freedom", PhD thesis, European University Institute, Florence 1993.

[7] Carter, I. (2001) "A Measure of Freedom", Oxford: Oxford University Press.

[8] Dardanoni V. (1993) ''Measuring Social Mobility", Journal of Economic Theory, vol. 61, pp.372-394.

[9] Dardanoni V. and A. Forcina (1998) ''A unified approach to likelihood inference in stochastic orderings in a nonparametric context", Journal of the American Statistical Association, 93, 1112-1123.

[10] Dawid A. P. (1979) Conditional independence in statistical theory, in Journal of the Royal Statistical Society, 41, 1-31.

[11] Dawid, A. P. (2001) "Causal Inference Without Counterfactuals, with discussion", Journal of the American Statistical Association, 95, pp. 407-448.

[12] Fleurbaey M. (1995) Equal opportunity or equal social outcome? Economics and Philosophy 11, 25-55.

[13] Gabor D. and A. Gabor (1954) ''An essay on the mathematical theory of freedom", Journal of the Royal Statistical Society, Series A (General), vol. 117, part 1. Reprinted in International Journal of Social Economics (1979) 6, pp.330-390.

[14] Greenland, S., Robins, J.M. and J. Pearl (1999) "Confounding and Collapsibility in Causal Inference", Statistical Science, vol.14, pp.29-46.

[15] Holland P. W. (1986) Statistics and causal inference (with discussion), in Journal of the American Statistical Association, 81, 945-960.

[16] MacCallum G. C. Jr. (1967) Negative and Positive Freedom, in Philosophical Review 76, 312-334.

[17] Miller D. (1983) Constraints on Freedom, in "Ethics", xciv, pp.66-86 .

[18] Oppenheim F. (1985) Constraints on Freedom as a Descriptive Concept, in Ethics, vol. xcv, pp. 305-9.

[19] Peragine V. (1999) The Distribution and Redistribution of Opportunity, in Journal of Economic Surveys, 13, pp. 37-69.

[20] Peragine V. (2002) "Opportunity egalitarianism and income inequality", in Mathematical Social Sciences (forthcoming).

[21] Pettit P. (2001) A theory of freedom, London: Polity.

[22] Roemer J.E. (1993) A pragmatic theory of responsibility for the egalitarian, in Philosophy and Public Affairs

[23] Steiner, H. (1974) "Individual Liberty", in Proceedings of the Aristotelean Society", vol. mxxv, pp. 33-50.

[24] Stone R. (1993) "The assumptions on which causal inferences rest", in Journal of the Royal Statistical Society, 55, 455-466.

[25] Taylor C. (1979) "What's wrong with negative liberty", in Ryn A (ed.) The idea of freedom, Oxford University Press, Oxford.

[26] van Hees M. (2001) "Freedom of choice and diversity of options. Some difficulties", manuscript.