UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Microeconomics and Statistical Decision Making

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Economics

by

Augusto Nieto Barthaburu

Committee in charge:

Professor Ross M. Starr, Chair
Professor Patrick J. Fitzsimmons
Professor Clive W. J. Granger
Professor Joel Sobel
Professor Christopher Woodruff

2006

The dissertation of Augusto Nieto Barthaburu is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2006

DEDICATION

A mi familia

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

When I finished my undergraduate studies in Economics and I decided to pursue a Ph.D. I could have never imagined that I would have the opportunity to work with some of the leading researchers in the world. The decision to go to San Diego turned out to be the right one, and I am very happy with the quality of the education I received at UCSD. Ross Starr was as good an advisor as I could have asked for. His constant encouragement and advise helped me immensely to reach this point in my career. I owe to him a great deal of the Economics I know. And more importantly, his constant patience and care have helped me not only academically but also personally.

The other members of my committee where also very important in my formation. Joel Sobel shaped me with his rigorous and deep thinking. Clive Granger was a great inspiration for my last essay (Chapter I in this dissertation), and I hope to have the opportunity to interact more with him in the future in a line of research that I am very enthusiastic with. While I am mostly a theoretically oriented researcher, working as an RA for Chris Woodruff made me appreciate empirical work. Last but not least, Patrick Fitzsimmons made Probability Theory accessible for me (and for many other of my fellow economists at UCSD), and with that helped me acquire an incredibly useful tool set, that I have used every single time I think about a problem ever since.

I also want to mention a few of my undergraduate teachers. Cristina Mirabella may be responsible for me being where I am now. She taught me my first course on Microeconomics, and left me with the thirst for understanding economics that I still feel today. Victor Elias, Osvaldo Meloni and Juan Carlos Abril where as exceptional as teachers as supportive when I decided to follow my

studies in the US.

Financial support from various institutions during my studies was extremely helpful. FOMEC Scholarship from Argentina during the academic year 1999-2000 allowed me to join UCSD. I am particularly grateful to Eduardo Juarez for making that support possible. During my stay in San Diego I received constant support from the UCSD Economics department, without which I could have never finished my studies. The Cal-$(IT)^2$ Institute also provided generous financial support during the academic year 2003-2004.

Chapter I in this dissertation was coauthored with Robert P. Lieli. The dissertation author was the principal investigator on that paper.

My family has been a constant support during my stay in San Diego. Without them I could have never done what I did and for that this dissertation is dedicated to them: To my Mom Maria and my brothers Santiago and Maria de la Paz. My godfather and cousin Popi has also been with me at every moment and for that a part of what I have achieved is also his.

My life in San Diego changed substantially three years ago. It was then when I met Faby, and her love and care has made me a much richer and happier person. She is constantly the measure against which I judge myself, and she makes me want to be a better person. She and her family (specially her Mom Edda) are an important part of my life.

Jorge, Isabel, Veronica, Jorgito and Emilia where my second family during my stay in California, and I am very grateful to them for worring and caring about me. Silvia, Daniel, Pipa and their family where also always there for me. Pablo, Dolores, Joaquin and Paloma always opened their home to me, and I will never forget the moments we spent together.

Robert is an exceptional friend. I shared with him wonderful moments and I am sure that we will be friends forever. Furthermore, I have learned more Economics and Philosophy hanging out with him than I could have learned in any class. I definitely feel that our conversations and discussions have sharpened my point of view enormously, and I feel very enthusiastic about our recently started collaboration, that I expect to be very fruitful in the future.

I am also fortunate to have been part of a humanly and intellectually rich group at UCSD. Daniel, Susana, Zhigang, Nada, Francisco, Andrew, Liangjun, Yoichi, Joseline and Takayuki where a challenging and fun class. Carlos, Giuseppe, Kevin and Marius are friends and former soccer teammates. Julie was my office-mate for a long time, and I have to apologize to her for taking a substantial part of the office with my stuff :). Thai, Anne and Baldomero where exceptional roommates. I also want to mention my Argentinian friends that visited me in San Diego, Charles and Sebastian, as well as those who couldn't make it, Diego, Federico, Flaco, Maximo, Sebastian and Trifon. Eddie is a great friend too.

It's been a great and fun journey, and I am sure I will be linked to San Diego for the rest of my life.

VITA

1974	Born, Tucumán, Argentina.
1999	B.A., Economics, Universidad Nacional de Tucumán, Argentina.
2006	Doctor of Philosophy, Economics, University of California, San Diego.

FIELDS OF STUDY

Major Field: Economics

Studies in Microeconomics and Econometrics.
Professors Ross M. Starr, Joel Sobel and Clive W. J. Granger.

ABSTRACT OF THE DISSERTATION

Essays on Microeconomics and Statistical Decision Making

by

Augusto Nieto Barthaburu
Doctor of Philosophy in Economics
University of California, San Diego, 2006
Professor Ross M. Starr, Chair

Chapter I of this dissertation addresses the problem of optimally forecasting a binary variable based on a vector of covariates in the context of two different decision making environments. First we consider a *single* decision maker with given preferences, who has to choose between two actions on the basis of an unobserved binary outcome. Previous research has shown that traditional prediction methods, such as a logit regression estimated by maximum likelihood and combined with a cutoff, may produce suboptimal decisions in this context. We point out, however, that often a prediction is made to assist in the decisions of a whole *population* of individuals with heterogeneous preferences who face various (binary) decision problems which are only tied together by a common unobserved binary outcome. A typical example is a public weather forecaster who needs to predict whether it will rain tomorrow. We show that a logit or probit regression estimated by maximum likelihood may well be "socially" optimal in this context in that it will lead certain populations of decision makers to social welfare maximizing decisions even when the underlying conditional probability model is grossly misspecified.

Chapter II analyzes the situation of a loan officer that makes sequential decisions on whether to grant loans or not to applicants. Before making each decision, the loan officer can observe some characteristic of the applicant, and in the case that a loan is granted, he can observe if whether it is paid back or not. On the other hand, when a loan is denied the officer cannot observe the applicant's behavior. This selection problem will have an effect on the lender's ability to learn about the population of borrowers. We find that in some cases the lender will be able to make correct decisions in the long run, but in other cases not, i.e. he can make mistakes forever. The crucial factor that determines whether full learning will occur is the nature of the information that the lender observes from the applicant before making each loan.

A Network Externality arises when the satisfaction that a consumer gets from the consumption of a given good depends (usually positively) on the number of consumers that consume the same good. A common feature of markets where NE are present is the phenomenon called "tipping", which is the tendency of one of the competing goods or protocols to win a substantial share of the market. In Chapter III it is argued that for tipping to occur there must be some underlying indivisibility in the consumption space. In addition to the problems posed by externalities for existence of equilibrium, indivisibilities create discontinuous demand behavior (not merely nonconvexity). In the paper we provide sufficient conditions for existence of competitive equilibrium with NE *and* indivisibilities. The key conditions are a large number of consumers and dispersion in their income distribution.

Chapter I

Optimal Binary Prediction for Group Decision Making

Abstract[1]

In this paper we address the problem of optimally forecasting a binary variable based on a vector of covariates in the context of two different decision making environments. First we consider a *single* decision maker with given preferences, who has to choose between two actions on the basis of an unobserved binary outcome. Previous research has shown that traditional prediction methods, such as a logit regression estimated by maximum likelihood and combined with a cutoff, may produce suboptimal decisions in this context. We point out, however, that often a prediction is made to assist in the decisions of a whole *population* of individuals with heterogeneous preferences who face various (binary) decision problems which are only tied together by a common unobserved binary outcome.

[1]Joint work with Robert P. Lieli. I am deeply indebted to Ross Starr for guidance and encouragement throughout my dissertation, and for his invaluable advice with this essay. We have further benefited from comments by David Kaplan, Tridib Sharma and the participants of the 2005 UT Austin-ITAM conference. All errors are our responsibility.

A typical example is a public weather forecaster who needs to predict whether it will rain tomorrow. We show that a logit or probit regression estimated by maximum likelihood may well be "socially" optimal in this context in that it will lead certain populations of decision makers to social welfare maximizing decisions even when the underlying conditional probability model is grossly misspecified.

I.A Introduction

Typically, the reason why one attempts to forecast the future value of a random variable Y is because knowledge of this variable would enable a decision maker, or a group of decision makers, to choose the best action from a set of alternatives. The value of a forecast is the "improvement" in decisions under uncertainty afforded by the new information conveyed by the forecast. Although most of traditional forecasting theory is based on expected loss minimization, the loss function is rarely derived explicitly form an underlying decision problem and the decision-supporting aspect of forecasting tends to be neglected. A typical way to handle the problem of forecasting Y based on a vector X of covariates is the following.

The first step is to specify a "convenient" loss function and minimize expected loss due to the error in forecasting Y. It is well known that in the case of quadratic loss this procedure leads to the conditional expectation $E(Y|X)$ as the best predictor of Y. Second, a parametric model is typically proposed for the conditional expectation and it is estimated by maximum likelihood (ML). Finally, the estimated model is used to make inferences about Y, and the results are presumably taken into account in subsequent decision making. The nature of this decision problem is usually not discussed—the underlying assumption seems to be that forecasts based on quadratic loss and ML estimation are appropriate in a wide range of decision making circumstances. It also remains unclear how exactly the forecast is going to be used in making a decision.

In recent years this traditional approach has come under criticism. In particular, as Granger and Pesaran (2000a,b) and Elliott and Lieli (2005) argue, when forecasts are constructed to aid in a decision making process, it is important

to take into account the nature of the decision problem and the preferences of the decision maker (henceforth DM) at all stages of the forecasting procedure outlined above.[2] This includes the specification of a relevant loss function and estimation.

The paper concentrates on two-action, two-state decision/forecasting problems. In this framework, DMs have a utility function

$$U(a, Y) = u_{a,Y}$$

where $a \in \{-1, 1\}$ denotes the action to be taken and $Y \in \{-1, 1\}$ denotes the state of the world (to be observed only after the decision has been made). Information about the future outcome Y is summarized by a vector X of covariates, observed prior to making a decision.[3] The utility function can be conveniently represented by the two-by-two matrix given in Table I.1. We assume

Assumption I.1 : $u_{1,1} > u_{1,-1}$ *and* $u_{-1,-1} > u_{-1,1}$.

That is, if the decision maker knew the state of the world, then it would be optimal to choose action $a = 1$ when state $Y = 1$ occurs and action $a = -1$ when state $Y = -1$ happens. This assumption rules out the uninteresting case in which there is a dominant action regardless of the state of the world.

The following examples describe situations which fit the preceding setup:

[2]The literature on decision based forecasting is fairly recent, and most of the previous work in the area has focused on forecast evaluation. Some examples of decision based forecast evaluation are McCulloch and Rossi (1990), West, Edison and Cho (1993), Diebold, Gunther and Tay (1997), Pesaran and Skouras (2001) and Bond and Satchell (2004). Treatments in which decision based forecasting methods are proposed are found in Lieli (2004) and Crone, Lessmann and Stahlbock (2005).

[3]In general, the DM's utility function could also depend on X. In our case, however, this would introduce additional complications that are beyond the focus of the paper. Therefore, we assume that utility for a given pair (a, Y) is invariant to the realization of the vector X.

Table I.1: The DM's preferences

	$Y=1$	$Y=-1$
$a=1$	$u_{1,1}$	$u_{1,-1}$
$a=-1$	$u_{-1,1}$	$u_{-1,-1}$

Example I.1 *A school official observes prospective students' test scores, high school GPA, personal information, etc* (X)*. The two possible states of the world are academic success or failure* (Y)*. The two possible decisions are accept or reject the applicant* (a)*.*

Example I.2 *A loan officer observes certain characteristics of a loan applicant, such as credit history, income, household size, etc* (X)*. The possible outcomes are default or no default* (Y)*. The possible actions are grant the loan or not* (a)*.*

Example I.3 *A weather forecaster observes today's weather conditions* (X) *in a certain geographic region. The possible states of the world are rain tomorrow or no rain tomorrow* (Y)*. Farmers in the region have to decide whether to irrigate or not* (a)*; commuters have to decide whether to take along an umbrella or not* (a)*, etc.*

Example I.4 *A city official follows the news about terrorists activities, reads government reports, etc* (X)*. The possible states of the world are terrorist attack or no terrorist attack* (Y)*. Commuters in the city have to decide wether to take the subway or a bike to work* (a)*; police have to decide whether to put additional patrols on the streets or not* (a)*, etc.*

Example I.5 *A financial advisor observes past and present macroeconomic conditions, past and present stock prices, etc* (X)*. The possible outcomes are that an*

option will expire in the money or out of the money (Y). *The advisor issues a bulletin for a set of clients, who have to decide whether to buy the option or not* (a).

In this paper we make a distinction between the forecast user (i.e. the DM) and the forecaster. We further distinguish between two kinds of forecasting activities, which we will refer to as "consulting" and "public/social service", respectively. In the consultant's problem, illustrated by Examples I.1 and I.2, there is a single DM with well-defined preferences facing a concrete decision problem. [4] We assume that the DM truthfully reveals these preferences to the forecaster and that there is a large sample of past observations available on X and Y. Given the data and the DM's utility function, the forecaster's job is to predict Y in a way that will lead the DM to the best (i.e. expected utility maximizing) course of action, at least asymptotically. In fact, the forecaster can just report a recommended *decision rule* to the DM, which directly specifies an action for any observed value of the covariates.[5]

On the other hand, the public service forecasting problem (Examples I.5 to I.4) is characterized by the presence of a heterogeneous population of decision makers facing potentially different (binary) decision problems which are only tied together by a common future outcome Y. The forecaster provides a "social service" by observing the history of X and reporting a single "public" forecast of Y; this information is then used as an input in many separate decision making processes.[6]

[4] In the everyday usage of the word, a "consultant" could also serve a group of DMs, as in Example I.5; however, in this paper we use the term "consultant's problem" to refer to the forecasting exercise where the objective is to help a single DM.

[5] There is actually no need to distinguish between action and point forecast in the consultant's problem; the forecast of Y can simply be *defined* as 1 if the recommended action is 1 and -1 if the recommended action is -1.

[6] As argued in more detail in Section I.D, the term "social service" is used in a wide meaning. It can

By the principles of decision based forecasting, when constructing the forecast and choosing what sort of statistic to report, the forecaster must account for the preferences of a whole population of decision makers involved in a potentially wide range of decisions.

As will be discussed in detail in Section I.B, traditional binary forecasting methods, such as logit and probit estimated by ML, are not in general optimal for individual decision making (the consultant's problem). Intuitively, the reason for this is that ML methods provide "global approximations" to the true conditional probability function $p(x) = P(Y = 1|X = x)$. As will be argued below, a global approximation may miss important features of a particular decision problem. In particular, the DM may be more interested in accurately approximating $p(x)$ in some regions of the support of the conditioning variables X than in others. ML, on the other hand, penalizes deviations from the true function $p(x)$ equally over the whole support of X.

Nevertheless, ML methods will be shown to be appropriate in situations with multiple DMs (the public service problem). Intuitively, when there are many DMs with dispersed preferences it is increasingly difficult to satisfy the individual needs of all of them. Therefore, the forecaster will need to provide an approximation to $p(x)$ that can be used by all DMs, i.e. a global approximation will be more appropriate in this case. This is exactly what ML provides.

The contributions of this paper are twofold. First, the existing literature on decision based forecasting only deals with the single DM case, i.e. setups similar to the consultant's problem. Thus, the paper extends the literature by introducing the public service forecasting problem and formalizing it using the idea of social welfare maximization, where social welfare maximization will be interpreted as the

represent situations in which the forecaster is completely self interested, as in Example I.5.

maximization of the forecaster's objectives. Second, we will provide sufficient conditions under which traditional binary prediction methods such as logit or probit regressions estimated by ML can be interpreted, asymptotically, as a "socially optimal" way to estimate the conditional probability of an outcome even if the model suffers from misspecification.

The paper will be organized as follows: In Section I.B we provide an intuitive comparison of the consultant's and the public service forecasting problems. In Section I.C we formalize the consultant's problem and subsequently, in Section I.D, the public forecaster's problem. Section I.E rewrites the likelihood function of a conditional probability model as a social welfare objective. In Section I.F we provide sufficient conditions on individual preferences and the social welfare objective under which maximum likelihood estimation can be interpreted as socially optimal. Section I.G summarizes and concludes. Some additional discussion and proofs are collected in the Appendix.

I.B Intuitive comparison of the consultant and the public service problems

In forecasting binary variables, the focus is usually on modeling and estimating the conditional probability

$$p(x) = P(Y = 1|X = x).$$

Traditionally, interest in $p(x)$ is justified by the use of a quadratic loss function; this conditional probability is then closely related to the mean-square optimal predictor of Y.[7] In decision based forecasting, interest in $p(x)$ arises since the function $p(x)$

[7] If $Y \in \{0, 1\}$, the conditional expectation of Y given X is $p(X)$ itself. Since we define $Y \in \{-1, 1\}$, the conditional expectation of Y given X is $2p(X) - 1$. Thus, $2p(X) - 1$ is the mean-square optimal point

describes the whole conditional distribution of Y given X, and consequently it has the broader interpretation of summarizing all statistical information contained in X about the outcome Y. Therefore, $p(x)$ is relevant for any expected utility maximization (or expected loss minimization) problem where uncertainty is due to Y and the information available is summarized by X.

As will be shown in Section I.C, the expected utility maximizing decision rule of an individual DM in the consultant's problem is of the form $a^*(x) = sign[p(x) - c]$, where the optimal cutoff c is determined by the utilities in Table I.1.[8] Let $m(x, \theta)$ denote a potentially misspecified parametric model of the conditional probability $p(x)$ (e.g. logit or probit). In discussing the consultant's problem, Elliott and Lieli (2005) propose estimating the unknown parameter vector θ by solving the sample analog of a suitable transformation of the DM's expected utility maximization problem. The resulting empirical decision rule $\hat{a}(x) = sign[m(x, \hat{\theta}) - c]$ can be shown to be asymptotically optimal even if the model $m(x, \theta)$ is largely misspecified. Traditional estimators of $m(x, \theta)$ such as maximum likelihood do not have this property.

Intuitively, this result arises because the loss function implicit in the DM's expected utility maximization problem contradicts the loss function implicit in ML estimation, which becomes an important issue when the postulated conditional probability model is misspecified. On one hand, expected utility maximization dictates that the model should be fit accurately at those points in the support of X where the cutoff intersects $p(x)$—the magnitude of errors is inconsequential away from these locations as long as the fitted model is on the "correct" side of

forecast of Y, even though the only values that Y can take on are -1 and 1, and the function $2p(X) - 1$ can take values in the entire $[-1, 1]$ interval. Allowing point forecasts to fall outside the support of the variable is typical in forecasting practice, even in cases in which Y is not binary.

[8]The function $sign(x)$ is defined as 1 if $x > 0$ and -1 if $x \leq 0$. In less cryptic terms, the optimal decision rule states "take action $a = 1$ if $p(x) > c$ and take action $a = -1$ if $p(x) \leq c$".

the cutoff. ML, on the other hand, penalizes deviations from the true conditional probability function over the entire support of X and as a result the fitted model will give a globally good approximation to $p(x)$. If $m(x, \theta)$ is misspecified for $p(x)$, then trying for a globally good fit may easily result in missing the critically important intersection points between $p(x)$ and the cutoff, even when a large sample of observations on X and Y is available.

Figure I.1 illustrates this point. In the figure the solid line represents the true conditional probability function $p(x)$, the light dashed line represents a locally good fit, and the dark dashed line represents a globally good fit (e.g. a misspecified logit estimated by ML). We see that the global fit is "closer" to $p(x)$ than the local fit for most of the support of the variable X. However, a locally good fit leads to optimal decisions for all values of X, since it is always on the same side of the optimal cutoff as $p(x)$. On the other hand, the global fit misses the important point of intersection between $p(x)$ and the cutoff line, and therefore it leads to suboptimal decisions in a sizable region of the support of X. Hence, although closer to the truth on average, the global fit is inferior to the local fit for decision making purposes. To make an optimal decision, the DM does not need to know the exact value of $p(x)$ for each value x, but only whether $p(x)$ is higher or lower than the cutoff for that particular x. Consequently, a local fit does not need to have any meaning beyond the decision rule it implies.

In the social service framework (multiple DMs) the forecaster's problem is more complex. In this setting, *each* individual DM j has an optimal decision rule of the form $a^*(x) = sign[p(x) - c_j]$, where the optimal cutoff c_j may vary with j, i.e. from individual to individual. Hence, the public forecaster cannot just report a single empirical decision rule; he must instead report an estimate $m(x, \hat{\theta})$ of $p(x)$

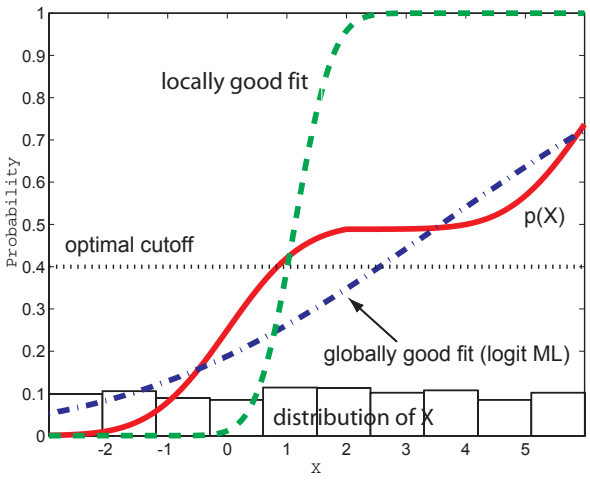

Figure I.1: Globally good fit vs. expected utility maximizing (locally good) fit.

from which DMs derive their own empirical decision rules $\hat{a}(x) = sign[m(x, \hat{\theta}) - c_j]$. The question is: Given a potentially misspecified parametric model $m(x, \theta)$ for $p(x)$ how should θ be estimated?

If the forecaster cares about the satisfaction of a group of decision makers, θ can no longer be estimated based on the expected utility maximization problem of a single individual, as the resulting fit may lead to very poor decisions for individuals whose cutoffs are different (again, see Figure I.1 and consider a cutoff equal to, say, 0.7). Therefore, we propose choosing $\hat{\theta}$ so as to maximize a *social welfare function* (SWF) defined over the utilities of individual decision makers, who then use the estimated decision rule $\hat{a}(x) = sign[m(x, \hat{\theta}) - c_j]$ to make their own decisions. This implies that the forecaster will make value judgments about the utilities of different members of the population, and those value judgments will be reflected in the value of $\hat{\theta}$ chosen.

Figure I.2 provides an illustration. In the figure we depict a relatively

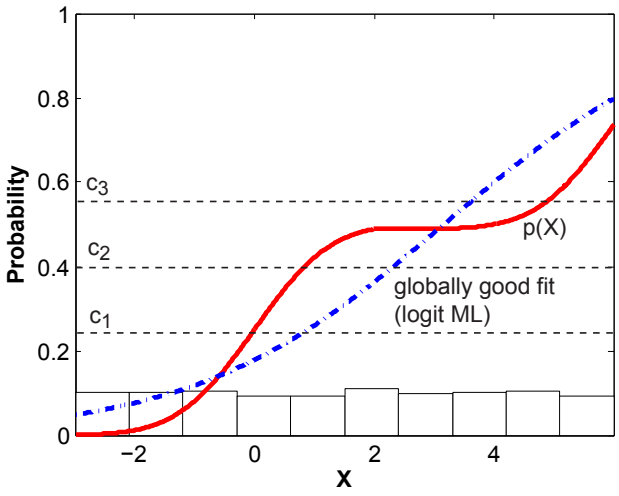

Figure I.2: The social service forecasting problem.

simple multiple DM problem with only three DMs who possess different optimal cutoffs. We observe, however, that a logit specification based on a linear index cannot possibly run through all the intersection points between $p(x)$ and the individual cutoffs. Therefore, given the specification, it is not possible to ensure correct decisions for all DMs, even asymptotically. The choice of the optimal fit will then necessarily require the forecaster to make interpersonal comparisons among individuals. In choosing θ optimally, the forecaster will need to clearly specify the relative importance of the decisions of different DMs.

As can be seen, maximum likelihood estimation of a conditional probability model may not be consistent with maximizing the expected utility of a *given* DM (the consultant's problem). However, it may well be consistent with maximizing the social welfare of a group of decision makers (the public forecaster's problem). If there are multiple DMs with a range of optimal cutoffs, and the forecaster cares about the utility of each DM, then the model will have to be fitted

as accurately as possible at a number of intersection points along $p(x)$ to ensure that each DM has a decision rule which is at least approximately optimal. Thus, it may well be the case that the a public forecasting problem (implicitly) calls for the model to give a *globally* good asymptotic approximation to $p(x)$, which is what ML generally provides. The natural question to ask is under what conditions estimating (potentially misspecified) parametric models by ML leads to empirical decision rules that maximize the social satisfaction of the population (asymptotically). The goal of the rest of the paper will be to formalize both forecasting problems and to provide an answer to this question.

It is also possible to think of the problem posed as an inquiry into the properties of quasi maximum likelihood estimation, i.e. maximization of a misspecified likelihood function. It is a fairly well known result (see, e.g., White 1996) that in this case maximum likelihood estimation can be thought of as asymptotically minimizing the Kullback-Liebler distance between the true and estimated models. Our goal is to offer an alternative, economically relevant, interpretation in the context of binary prediction models; specifically, our results will show that maximum likelihood estimation of misspecified logit or probit models may still possesses optimality properties for a specific *group* of decision makers who rely on these models in making their decisions.

I.C A formal statement of the consultant's forecasting problem

As discussed in the introduction, in the consultant's problem there is a single DM, with a given utility function $u(a, Y) = u_{a,Y}$, where $Y \in \{-1, 1\}$ denotes the (future) state of the world, and $a \in \{-1, 1\}$ denotes the action to be taken

(see Table I.1). Given $X = x$, the DM's expected utility, a function of the action a, can be written as

$$E[u(a, Y) \mid X = x] = \begin{cases} u_{1,1}p(x) + u_{1,-1}[1 - p(x)] \text{ if } a = 1 \\ u_{-1,1}p(x) + u_{-1,-1}[1 - p(x)] \text{ if } a = -1, \end{cases}$$

where $p(x) = P(Y = 1|X = x)$.

The DM will choose action $a = 1$ if and only if the expected utility associated with this choice is greater than that associated with $a = -1$.[9] It is easily shown that this leads to the optimal decision rule

$$\text{"take action } a = 1 \text{ if and only if } p(x) > c\text{"}, \tag{I.1}$$

where the optimal cutoff c is given by

$$c = \frac{u_{-1,-1} - u_{1,-1}}{(u_{1,1} - u_{-1,1}) + (u_{-1,-1} - u_{1,-1})}.$$

Under Assumption I.1 the optimal cutoff c lies strictly between 0 and 1. The numerator is the net benefit from taking correct action (i.e. $a = -1$) when $Y = -1$. The denominator is the total net benefit from taking correct action considering both states. Hence, c can be interpreted as the relative net benefit of taking action -1 versus action 1. The higher this value, the higher the probability that $Y = 1$ has to be to make it worthwhile for the DM to take action 1.

In the consultant's problem the forecaster's primary goal is to recover the optimal decision rule $a^*(x) = sign[p(x) - c]$ of a *given* DM. Suppose the forecaster

[9]The formulation of the DM's expected utility maximization problem implies that we assume no feedback from the action taken to the (conditional) probability of the outcome Y. For example, consider a central bank interested in forecasting whether a currency crisis will occur and deciding whether to take preventive action. If in the bank's estimation the probability of a crisis is high enough, then it will launch measures designed to *reduce* the probability of a crisis. The existence of such feedback implies that we would have to write $p_a(x)$ in place of $p(x)$ in the DM's expected utility maximization problem. This complication will be assumed away in this paper.

adopts the parametric model $m(x, \theta)$ for $p(x)$, where $\theta \in \Theta$, $\Theta \subset \mathbf{R}^d$. As discussed in Section 1, maximizing a given DM's expected utility requires of the model to have a *locally* good fit at x-values where $p(x)$ intersects the optimal cutoff c. More formally, the goal is to choose $\hat{\theta}$ such that

$$sign[p(x) - c] = sign[m(x, \hat{\theta}) - c] \text{ for all } x \in \text{supp}(X), \tag{I.2}$$

provided that such $\hat{\theta}$ exists. If the model is estimated by maximum likelihood, the result is a globally good fit, which is not necessarily consistent with (I.2), even asymptotically, unless the model happens to be correctly specified for $p(x)$. As shown Lieli (2004) and Elliott and Lieli (2005), a locally good fit can be achieved under general conditions by solving

$$\max_{\theta} S(\theta) \equiv \max_{\theta} E_{Y,X}\big\{[Y + 1 - 2c]sign[m(X, \theta) - c]\big\}.$$

Maximization of $S(\theta)$ will recover the DM's optimal decision rule if condition (I.2) is satisfied for some $\hat{\theta} \in \Theta$. (This is a much weaker requirement than correct specification.) Even if no point in the parameter space satisfies (I.2), maximizing $S(\theta)$ will result in an optimal decision rule *conditional* on the model specification $m(x, \theta)$. As maximization of $S(\theta)$ guarantees a locally good fit only, the value of the estimated model $m(x, \hat{\theta})$ at any particular point x does not, in general, have any meaning beyond the implied decision rule $\hat{a}(x) = sign[m(x, \hat{\theta}) - c]$. However, this is all what is needed in the consultant's problem.

The objective function $S(\theta)$ can be derived directly from the DM's expected utility maximization problem. It is easy to show that the function

$$s(a, y) = [y + 1 - 2c]a$$

is an affine transform of $u(a, y)$ and hence is an equally good measure of the DM's utility. Furthermore, maximizing $E_{Y,X}[s(a(X), Y)]$ over the space of all possible de-

cision rules $a: \ supp(X) \to \{-1, 1\}$ is equivalent to solving the "original" expected utility maximization problem posed at the beginning of the section. Restricting the set of possible decision rules to $a(x, \theta) = sign[m(x, \theta) - c]$, $\theta \in \Theta$, yields the objective function $S(\theta)$.

Of course, in practice one cannot directly maximize the objective function $S(\theta)$. Instead, one fits the model $m(x, \theta)$ by solving the sample analog problem

$$\max_{\theta} \mathcal{S}_n(\theta) = \max_{\theta} n^{-1} \sum_{i=1}^{n} (Y_i + 1 - 2c) sign[m(X_i, \theta) - c]$$

for a sample of observations $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$. The solution of this maximization problem is called the "maximum utility" (MU) estimator by Lieli (2004) and Elliott and Lieli (2005). The MU estimator turns out to be a generalization of Manski's (1975, 1985) maximum score estimator.

If the sequence of random variables $(Y_i + 1 - 2c) sign[m(X_i, \theta) - c]$, $i = 1, 2, \ldots, n$, satisfy a (strong) law of large numbers, then $\mathcal{S}_n(\theta)$ converges a.s. to $S(\theta)$ for any fixed $\theta \in \Theta$. This "pointwise" convergence is not enough to guarantee that maximizing $\mathcal{S}_n(\theta)$ is asymptotically equivalent to maximizing $S(\theta)$. The key requirement for this is uniform convergence in θ, which is guaranteed under the following conditions (Elliott and Lieli 2005):

Assumption I.2 *(a)* Θ *is a compact and* $m(x, \theta)$ *is Borel-measurable as a function from* $\text{supp}(X) \times \Theta$ *to* $\mathbf{R}$*;*

(b) (i) the function $\theta \mapsto m(x, \theta)$ *is continuous on* Θ *for all* x *in the support of* X*;*

(ii) $P[m(X, \theta) = c] = 0$ *for each* $\theta \in \Theta$*;*

(c) $\{(Y_i, X_i')\}_{i=1}^{\infty}$ *is a (strictly) stationary, ergodic sequence of observations on* (Y, X').

I.D A formal statement of the public service forecasting problem

In the public forecaster's problem there is a population of DMs, indexed by $j = 1, 2, \ldots, J$, who face various binary decision problems tied together by the common outcome Y. Each DM has a utility function $u_j(a, y)$ and seeks to maximize their own expected utility. As shown in the previous section, the optimal decision rule for any DM j is then given by $a_j^*(x) = sign[p(x) - c_j]$, where c_j is the optimal cutoff determined by the utility function $u_j(a, y)$.

Suppose the public forecaster specifies the parametric model $m(x, \theta)$ for $p(x)$ and reports the estimated conditional probability $m(x, \hat{\theta})$ to the public. Decision makers will be assumed to approximate their optimal decision rule by simply plugging in this publicly reported estimate of $p(x)$ into (I.1):

$$\hat{a}_j(x) = \hat{a}_j(x, \hat{\theta}) = sign[m(x, \hat{\theta}) - c_j]. \tag{I.3}$$

The fact that a single estimate of $p(x)$ is to be used by many different DMs deserves some justification. It can be argued that the forecaster could in principle make a different (and optimal) forecast for each DM. Although there are situations in which the decision problems of the individual DMs are important enough to justify the construction of separate forecasts, our treatment is relevant for many real world situations where the potential gains and costs of individual decisions do not justify the costs of hiring a private forecaster. For example, a financial advisor may advise a group of small investors. Then, the advisor periodically sends out a report to all her clients with predictions about the market for the next month. For each of the clients it would not be practical to hire a private advisor, since that would be much costlier.

As argued in Section I.C, the utility DM j derives from adopting decision rule (I.3) can be measured by $s(\hat{a}_j(X,\hat{\theta}),Y) = (Y+1-2c_j)sign[m(X,\hat{\theta})-c_j]$. The forecaster aggregates these utilities through a "social welfare function"

$$w\big(s(\hat{a}_1(X,\hat{\theta}),Y), s(\hat{a}_2(X,\hat{\theta}),Y),\ldots,s(\hat{a}_J(X,\hat{\theta}),Y)\big).$$

The forecaster can affect social welfare through the choice of $\hat{\theta}$. In particular, we will assume that the forecaster chooses $\hat{\theta}$ so as to solve

$$\max_{\theta} W(\theta) \equiv \max_{\theta} E_{Y,X}\Big\{w\big(s(\hat{a}_1(X,\hat{\theta}),Y), s(\hat{a}_2(X,\hat{\theta}),Y),\ldots,s(\hat{a}_J(X,\hat{\theta}),Y)\big)\Big\}.$$

We point out that we do not necessarily think of the forecaster as a "social planner". Even though we call w a social welfare function, it may really stand for any objective the forecaster wishes to maximize, e.g. a profit function or any measure of the forecaster's performance (see Example I.5). Thus, the forecaster may well be following his own self-interest by trying to maximize what we call "social welfare". Similarly, u or s may have a more general interpretation than a utility function. Therefore, our treatment applies to many more situations than that of a social planner interested in maximizing social welfare.

The public forecaster's optimization problem has been derived under the presumption that individual decision makers will act in accordance with (I.3), i.e. simply substitute the reported conditional probability $m(x,\hat{\theta})$ for $p(x)$ in their optimal decision rules. This assumption abstracts from the possibility of sophisticated strategic behavior on the part of individual agents: If DMs know that the forecaster's objective is to maximize a given social welfare objective $W(\theta)$, in principle they could try to make adjustments to the forecast provided so that it fits their preferences more closely.[10] Considering this potential "feedback" in the

[10]We thank Tridib Sharma for pointing this out.

forecaster's objective would introduce a complicated fixed point problem. Instead we simply assume that for one reason or another agents act as if they fully believe the reported forecast. For example, they might not have any other or better source of information; they might think that $m(x, \hat{\theta})$ coincides with the truth or at least provides an unbiased or consistent estimate of it, etc.

In the rest of the paper we will specialize to the case where the social welfare objective $w(\cdot)$ is given by a weighted sum of individual utilities, yielding

$$\begin{aligned} W(\theta) &= E_{Y,X}\left\{\sum_{j=1}^{J}\gamma_j(Y+1-2c_j)sign[m(X,\theta)-c_j]\right\} \\ &= \sum_{j=1}^{J}\gamma_j E_{Y,X}\left\{(Y+1-2c_j)sign[m(X,\theta)-c_j]\right\}, \end{aligned}$$

where $\gamma_j \geq 0$, $\sum_j \gamma_j = 1$.

It is apparent that the optimal cutoffs c_j are the only channel through which individual preferences enter the public forecaster's optimization problem. It will be convenient to consolidate decision makers sharing the same optimal cutoff value into groups which we will also refer to as "types". Suppose individual cutoffs take on $T-1$ different values for some positive integer T. Without loss of generality, one can assume that DMs $t_0 = 1$ through t_1 have cutoff c_1, DMs $t_1 + 1$ through t_2 have cutoff c_2, etc, and DMs $t_{T-2} + 1$ through $t_{T-1} = J$ have cutoff c_{T-1} for some integers $t_0 = 1 < t_2 < t_3 < \ldots < t_{T-1} = J$. The public forecaster's problem can then be rewritten as

$$\max_{\theta} W(\theta) = \max_{\theta} \sum_{t=1}^{T-1} \lambda_t E_{Y,X}\Big\{(Y-2c_t+1)sign[m(X,\theta)-c_t]\Big\}, \tag{I.4}$$

where

$$\lambda_s = \sum_{j=t_{s-1}}^{t_s} \gamma_j, \qquad s = 1, \ldots, T-1.$$

The coefficients λ_t, $t = 1, \ldots, T-1$, represent the total weight given by the social welfare function to DMs of type t (i.e. to DMs who share the same optimal cutoff value c_t). It follows immediately from the properties of the individual welfare weights γ_j that $\lambda_t \geq 0$ and $\sum_t \lambda_t = 1$. The relative magnitudes of the type welfare weights reflect two different factors. First, given an equal number of DMs of each type, the λ_t's can differ because (some) individuals in one group are weighted more heavily than individuals in another group. Second, even if each DM of each type has the same individual welfare weight γ_j, one of the types may have more individuals than the other. Hence, any given collection of cutoffs and type welfare weights is consistent with many different distributions of individuals across those types and many different configurations of individual welfare weights.

Of course, in practice one cannot directly solve (I.4), as the expectation is taken with respect to an unknown joint distribution. Given a sample of observations $\{(X_i, Y_i)\}$, $i = 1, 2, \ldots, n$, one must form an empirical objective function $\mathcal{W}_n(\theta)$ by replacing the expectation operator in (I.4) by a sample average:

$$\mathcal{W}_n(\theta) = n^{-1} \sum_{t=1}^{T-1} \sum_{i=1}^{n} \lambda_t \Big\{ (Y_i - 2c_t + 1) sign[m(X_i, \theta) - c_t] \Big\}.$$

The maximizer of $\mathcal{W}_n$ will be called the "maximum (social) welfare" (MW) estimator of $m(x, \theta)$. The conditions given in Assumption I.2 are sufficient to ensure that maximizing the empirical social welfare objective is asymptotically equivalent to maximizing $W(\theta)$ as the sample size goes to infinity. In the Appendix we present some simple examples which illustrate how the MW estimator fits a simple logit model to the data for different cutoff and welfare weight configurations.

The basic question we now seek to answer is the following. Are there conditions under which maximization of the social welfare objective (I.4) is equivalent to maximizing the asymptotic log-likelihood function associated with the model

$m(x,\theta)$?

If the model $m(x,\theta)$ happens to be correctly specified for $p(x)$, then there is a parameter value θ° such that $m(x,\theta^\circ) = p(x)$ for all x in the support of X and maximum likelihood will be consistent for θ° under general conditions. But then θ° must also solve (I.4) for *any* collection of types and welfare weights, since we know that the decision rule $sign[p(x) - c_j]$ is individually optimal for each decision maker j.

Hence, the question posed above becomes interesting only if $m(x,\theta)$ is misspecified for $p(x)$, which is actually the more relevant case. Is there a distribution of cutoffs c_t, $t = 1,\ldots,T-1$, over the $(0,1)$ interval and a corresponding set of welfare weights λ_t, $t = 1,\ldots,T-1$, such that solving (I.4) is equivalent to maximizing the asymptotic log-likelihood of some misspecified conditional probability model $m(x,\theta)$? Is there a distribution of cutoffs and weights that "works" for *any* model in a wide class regardless of the extent of misspecification? The answer is yes; the rest of the paper is devoted to the development of this result.

To this end, we will now decompose $W(\theta)$ in a way that will allow us to compare it with the asymptotic log-likelihood of a conditional probability models (to be introduced in Section I.E below). The decomposition is "conditional" on a set of individually optimal cutoffs $c_1 < c_2 < \ldots < c_{T-1}$. To emphasize this, we will write $W_{C_T}(\theta)$ for the decomposed objective function, where $C_T \equiv \{c_0, c_1, \ldots, c_{T-1}, c_T\} \subset [0,1]$ with $c_0 = 0$ and $c_T = 1$.

The basic idea is to put[11]

$$sign[m(X,\theta) - c_t] = 1[m(X,\theta) > c_t] - 1[m(X,\theta) \leq c_t]$$

[11]The function 1[condition] will denote the indicator function. The indicator function equals 1 if the condition it refers to is true and 0 otherwise.

and to decompose the events $[m(X,\theta) > c_t]$ and $[m(X,\theta) \leq c_t]$ as

$$1[m(X,\theta) > c_t] = \sum_{g=t}^{T-1} 1[c_g < m(X,\theta) \leq c_{g+1}] \text{ and}$$

$$1[m(X,\theta) \leq c_t] = \sum_{g=1}^{t} 1[c_{g-1} < m(X,\theta) \leq c_g].$$

After substituting these expressions into (I.4), a fair amount of rearranging and collecting terms, we obtain the following expression for W_{C_T}:

$$\begin{aligned}
&W_{C_T}(\theta) = \\
&E\left\{\left[-\sum_{g=1}^{T-1}\lambda_g(Y - 2c_g + 1)\right] 1[0 < m(X,\theta) \leq c_1]\right\} \\
&+E\left\{\left[\lambda_1(Y - 2c_1 + 1) - \sum_{g=2}^{T-1}\lambda_g(Y - 2c_g + 1)\right] 1[c_1 < m(X,\theta) \leq c_2]\right\} \\
&+\ldots+ \\
&+E\left\{\left[\sum_{g=1}^{t-1}\lambda_g(Y - 2c_g + 1) - \sum_{g=t}^{T-1}\lambda_g(Y - 2c_g + 1)\right] 1[c_{t-1} < m(X,\theta) \leq c_t]\right\} \\
&+\ldots+ \\
&+E\left\{\left[\sum_{g=1}^{T-1}\lambda_g(Y - 2c_g + 1)\right] 1[c_{T-1} < m(X,\theta) \leq 1]\right\}.
\end{aligned}$$

Bringing the expectations "inside" and rearranging yields

$$\begin{aligned}
W_{C_T}(\theta) &= \sum_{t=1}^{T} \alpha_t^{MW} E[Y \mid c_{t-1} < m(X,\theta) \leq c_t] P[c_{t-1} < m(X,\theta) \leq c_t] \\
&+ \sum_{t=1}^{T} \beta_t^{MW} P[c_{t-1} < m(X,\theta) \leq c_t],
\end{aligned}$$

where

$$\alpha_t^{MW} = \sum_{g=1}^{t-1}\lambda_g - \sum_{g=t}^{T-1}\lambda_g, \quad \beta_t^{MW} = \sum_{g=1}^{t-1}\lambda_g(1 - 2c_g) - \sum_{g=t}^{T-1}\lambda_g(1 - 2c_g), \qquad \text{(I.5)}$$

$t = 1, \ldots, T$. Here we use the notational convention that the "empty" sum is zero.

We will now turn to the problem of studying the (asymptotic) log-likelihood function of conditional probability models. The goal is rewrite this function in a way analogous to $W_{C_T}(\theta)$ so that we can establish sufficient conditions under which the two functions (or at least their maximizers) coincide.

I.E The asymptotic log-likelihood function

Let $m(x,\theta)$, $\theta \in \Theta \subset \mathbf{R}^d$, be a potentially misspecified parametric model for $p(x)$. Given a random sample of observations $\{(Y_i, X_i')\}_{i=1}^n$, the normalized conditional log-likelihood can be written as 1/2 times

$$\mathcal{L}_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} \Big\{ (1+Y_i) \log[m(X_i,\theta)] + (1-Y_i) \log[1-m(X_i,\theta)] \Big\}.$$

The maximum likelihood estimator of the model $m(x,\theta)$ is of course obtained by maximizing $\mathcal{L}_n(\theta)$ w.r.t. θ.[12] We will focus on the behavior of this estimator as the sample size goes to infinity. If $E|\log[m(X,\theta)]| < \infty$ and $E|\log[1-m(X,\theta)]| < \infty$ for all θ, then by the strong law of large numbers $\mathcal{L}_n(\theta)$ converges almost surely to

$$L(\theta) \equiv E\Big\{ (1+Y) \log[m(X,\theta)] + (1-Y) \log[1-m(X,\theta)] \Big\}$$

for any fixed value of θ. The function $L(\theta)$ will be referred to as the asymptotic log-likelihood function.

There is a fairly extensive statistical and econometric literature concerned with the conditions under which the maximum likelihood estimator, a maximizer of $\mathcal{L}_n(\theta)$, converges to a maximizer of $L(\theta)$; see, for example, Amemiya (1985), White (1996). As mentioned in Section I.C, the key requirement is (a.s.) *uniform*

[12] If $\{(Y_i, X_i')\}_{i=1}^n$ is a stationary, ergodic sequence rather than i.i.d., maximization of $\mathcal{L}_n(\theta)$ still leads to a consistent estimator of $p(x)$ if $m(x,\theta)$ is correctly specified. This follows, for example, from the arguments in Gallant (1997, Ch. 5.2).

convergence of $\mathcal{L}_n(\theta)$ to $L(\theta)$. For a strictly stationary and ergodic sequence of observations, the following is a set of sufficient conditions for a uniform law of large numbers to hold (c.f. White 1996, Thm. A.2.2):

Assumption I.3 *(a) Θ is a compact subset of $\mathbf{R}^d$ and $0 \leq m(x,\theta) \leq 1$ for all $\theta \in \Theta$ and x in supp(X);*

(b) $x \mapsto m(x,\theta)$ is Borel-measurable for all $\theta \in \Theta$ and $\theta \mapsto m(x,\theta)$ is continuous on Θ for all x in supp(X);

(c) $E \sup_{\theta\in\Theta} |\log[m(X,\theta)]| < \infty$ and $E \sup_{\theta\in\Theta} |\log[1-m(X,\theta)]| < \infty$.

We will henceforth think of the maximum likelihood estimator as asymptotically maximizing $L(\theta)$ and rewrite this function in a form that makes it possible to compare it with the decomposed social welfare objective $W_{C_T}(\theta)$. Let us define a partition C_T of the [0,1] interval as a set of points $\{c_t,\, t = 0, 1, \ldots, T\}$ such that $0 = c_0 < c_1 < c_2 < \ldots < c_{T-1} < c_T = 1$. Equal spacing is not required, but think of T as large and $\max_t(c_t - c_{t-1})$ as small. Given a partition C_T, we can write

$$L(\theta) = \sum_{t=1}^{T} E\Big\{(1+Y)\log[m(X,\theta)]1[c_{t-1} < m(X,\theta) \leq c_t]\Big\}$$
$$+ \sum_{t=1}^{T} E\Big\{(1-Y)\log[1-m(X,\theta)]1[c_{t-1} < m(X,\theta) \leq c_t]\Big\}.$$

If T is large and $\max_t(c_t - c_{t-1})$ is indeed small, $L(\theta)$ should be well approximated by

$$L_{C_T}(\theta) \equiv \sum_{t=1}^{T} E\Big\{(1+Y)\log(c_t)1[c_{t-1} < m(X,\theta) \leq c_t]\Big\}$$
$$+ \sum_{t=1}^{T} E\Big\{(1-Y)\log(1-c_{t-1})1[c_{t-1} < m(X,\theta) \leq c_t]\Big\}.$$

Rearranging terms on the RHS yields

$$L_{C_T}(\theta) = \sum_{t=1}^{T} \alpha_t^{ML} E[Y \mid c_{t-1} < m(X,\theta) \leq c_t] P[c_{t-1} < m(X,\theta) \leq c_t]$$
$$+ \sum_{t=1}^{T} \beta_t^{ML} P[c_{t-1} < m(X,\theta) \leq c_t],$$

where

$$\alpha_t^{ML} = \log(c_t) - \log(1 - c_{t-1}), \quad \beta_t^{ML} = \log(c_t) + \log(1 - c_{t-1}). \tag{I.6}$$

The function $L_{C_T}(\theta)$ is directly comparable with $W_{C_T}(\theta)$. However, we must first argue that $L_{C_T}(\theta)$ approximates the asymptotic log-likelihood $L(\theta)$ well enough so that maximizing L_{C_T} becomes equivalent to maximizing $L(\theta)$ as the partition C_T becomes fine.

Proposition I.1 *Suppose Assumption I.3(c) is satisfied. Then for any $\epsilon > 0$ there exists $\delta > 0$ such that if a partition C_T satisfies $c_t - c_{t-1} < \delta$ for $t = 1, 2, \ldots, T$, then $|L_{C_T}(\theta) - L(\theta)| < \epsilon$ for all $\theta \in \Theta$.*

The proof of Proposition I.1 makes use of the fact that if a r.v. Z has finite expectations, then $E(Z1[|Z| > z])$ has to be small when z is large and that the log function is uniformly continuous over closed intervals. The details of the proof can be found in the Appendix.

Proposition I.1 immediately implies that if a sequence of partitions $\{C_T\}_{T=1}^{\infty}$ satisfies $\max\{c_t - c_{t-1}, t = 1, 2, \ldots, T\} \to 0$ as $T \to \infty$ (e.g. $C_T = \{t/T, t = 0, 1, \ldots, T\}$), then $\sup_\theta |L_{C_T}(\theta) - L(\theta)|$ converges to zero. Of course, this is just another way of stating that $L_{C_T}(\theta)$ converges to $L(\theta)$ uniformly in θ. As discussed above, this means that maximizing $L_{C_T}(\theta)$ is practically equivalent to maximizing $L(\theta)$ when C_T is sufficiently fine.

I.F Comparing $L(\theta)$ and $W(\theta)$

The approximate asymptotic log-likelihood $L_{C_T}(\theta)$ and the social welfare objective $W_{C_T}(\theta)$ have now been expressed as weighted sums of the same form, but with different sets of weights, given by (I.6) and (I.5), respectively. The goal is to choose a sequence of partitions (cutoffs) $\{c_{t,T}, t = 1, \ldots, T\}_{T=1}^{\infty}$ and a corresponding triangular array of welfare coefficients $\{\lambda_{t,T}, t = 1, \ldots, T\}_{T=1}^{\infty}$, $\lambda_{t,T} \geq 0$, $\sum_t \lambda_{t,T} = 1$, such that

(i) $L_{C_T}(\theta)$ converges to $L(\theta)$ uniformly in θ;

(ii) the resulting set of weights

$$\{(\alpha_t^{ML}, \beta_t^{ML}), t = 1, \ldots, T\}_{T=1}^{\infty} \text{ and } \{(\alpha_t^{MW}, \beta_t^{MW}), t = 1, \ldots, T\}_{T=1}^{\infty}$$

make maximizing $L_{C_T}(\theta)$ equivalent to maximizing $W_{C_T}(\theta)$.

The basic question is that of existence: Are there sequences of partitions and weighting schemes satisfying these conditions? If yes, are there many? For any given sequence of partitions satisfying (i), is it possible to find a weighting sequence such that requirement (ii) is satisfied? We will not give a complete answer to all these questions in this version of the paper. We will however construct a specific sequence of partitions and a specific weighting scheme consistent with conditions (i) and (ii) for large T. Thus, we will have shown that maximum likelihood estimation of (misspecified) conditional probability models *can* be interpreted as (implicitly) maximizing the social welfare of a group of DMs under certain circumstances, but we do not claim to have uncovered *all* possible circumstances when this is true.

As the discussion following Proposition I.1 shows, requirement (i) is easily satisfied by choosing $\{C_T\}_{T=1}^{\infty}$ such that $\max\{c_{t,T} - c_{t-1,T}, t = 1, \ldots T\} \to 0$ as

$T \to \infty$. Requirement (ii) will hold for each value of T if W_{C_T} is an increasing affine transformation of L_{C_T}, i.e. there exist constants $a_T > 0$, u_T, v_T such that

$$\alpha_{t,T}^{MW} = a_T \alpha_{t,T}^{ML} + u_T \tag{I.7}$$

$$\beta_{t,T}^{MW} = a_T \beta_{t,T}^{ML} + v_T, \qquad t = 1, \ldots, T. \tag{I.8}$$

In general, equations (I.7) and (I.8) define a complex set of relationships between the welfare weights λ_t and the cutoffs c_t. (For notational convenience, we will suppress the dependence of these quantities on T.) These equations are considerably easier to evaluate if we restrict ourselves to symmetric partitions of the [0,1] interval and symmetric welfare weights, i.e. those which satisfy $c_t = 1 - c_{T-t}$ and $\lambda_t = \lambda_{T-t}$. These restrictions force $u_T = 0$, $v_T = 1$ and $a_T = 1/\log(T)$. Without loss of generality we can assume that T is even. Writing out (I.7) and (I.8) in detail yields

$$a[\log(c_t) - \log(1 - c_{t-1})] = -1 + 2\sum_{g=1}^{t-1} \lambda_g \tag{I.9}$$

$$a[\log(c_t) + \log(1 - c_{t-1})] = -1 + 2\sum_{g=1}^{t-1} (1 - 2c_g)\lambda_g, \tag{I.10}$$

$$t = 1, \ldots, T/2.$$

Using the recursive structure of these equations, it is fairly straightforward to eliminate the welfare weights from the equations, which yields the following nonlinear second-order difference equation in the cutoffs c_t:

$$\log(c_{t+1}) = \log(c_t) + (1 - 1/c_t)[\log(1 - c_t) - \log(1 - c_{t-1})] \qquad t = 1, \ldots, T/2 - 1.$$

One needs solve this difference equation subject to the terminal condition $c_{T/2} = 1/2$ (from symmetry) and requirement (i). It is possible to obtain solutions by trying different initial values for $c_{1,T}$ and $c_{2,T}$ and using numerical simulation.

For simplicity, we will only derive here an approximate "large T" solution to this difference equation.

Suppose T is large and $\max_t(c_t - c_{t-1})$ is small. Then

$$\log(c_{t+1}) - \log(c_t) \approx (c_{t+1} - c_t)/c_t \quad \text{and}$$
$$\log(1 - c_t) - \log(1 - c_{t-1}) \approx (c_{t-1} - c_t)/(1 - c_t).$$

Substituting these approximations yields

$$c_{t+1} - 2c_t + c_{t-1} = 0.$$

It is trivial to verify that the sequence $c_t^* = t/T$ satisfies this equation. The corresponding welfare weights λ_t^* can be obtained recursively from (I.9) or (I.10). The resulting weights show a characteristic U-shape and are depicted in Figure I.3. We summarize the derivations in this section by the following proposition.

Proposition I.2 *For T large, the cutoffs $c_t^* = t/T$ and the corresponding welfare weights λ_t^* implied by equation (I.9) for $t = 1, \ldots, T-1$ define populations of decision makers and a social welfare objective*

$$W^*(\theta) = \sum_{t=1}^{T-1} \lambda_t^* E_{Y,X}\Big\{(Y - 2c_t^* + 1) sign[m(X,\theta) - c_t^*]\Big\},$$

such that maximizing the asymptotic log-likelihood $L(\theta)$ is equivalent to maximizing $W^*(\theta)$ as $T \to \infty$ for any *joint distribution of (Y, X) and model specification $m(x, \theta)$ for which $W^*(\theta)$ and $L(\theta)$ exist.*

Thus, if Assumptions I.2 and I.3 are also satisfied, then maximizing the log likelihood $\mathcal{L}_n(\theta)$ is asymptotically equivalent to maximizing the empirical social welfare objective $\mathcal{W}_n^*(\theta)$ as n and T go to infinity. In short, for the special

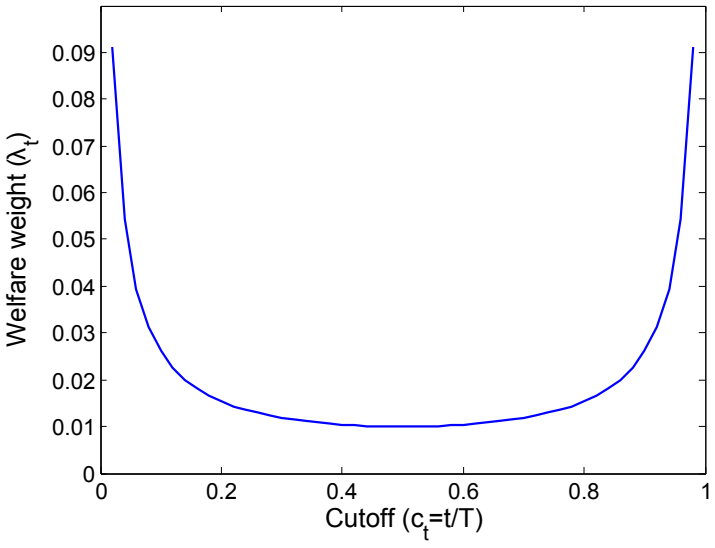

Figure I.3: Welfare weights implied by maximum likelihood for the partition $c_t = t/T$

populations of decision makers and social welfare function defined in Proposition I.2, the ML and MW estimators of $m(x, \theta)$ asymptotically coincide.

Figure I.4 offers an illustration. This graph depicts two (misspecified) logit models—one of them estimated by ML the other by MW using a large sample of observations on X and Y. The MW estimator was constructed using the cutoffs $t/50$, $t = 1, \ldots, 49$, with the corresponding type weights given by λ_t^*. As the graphs shows, the resulting fit is barely distinguishable from the ML estimate.

We see from Figure I.3 that when cutoffs are "uniformly" dispersed, the set of group weights that make the MW and ML estimators asymptotically identical are U-shaped. That is, low and high cutoffs (types) receive heavy consideration in the social welfare function, while cutoffs in the middle of the $(0, 1)$ interval are less important for the forecaster. (As noted in Section I.D, this distribution of

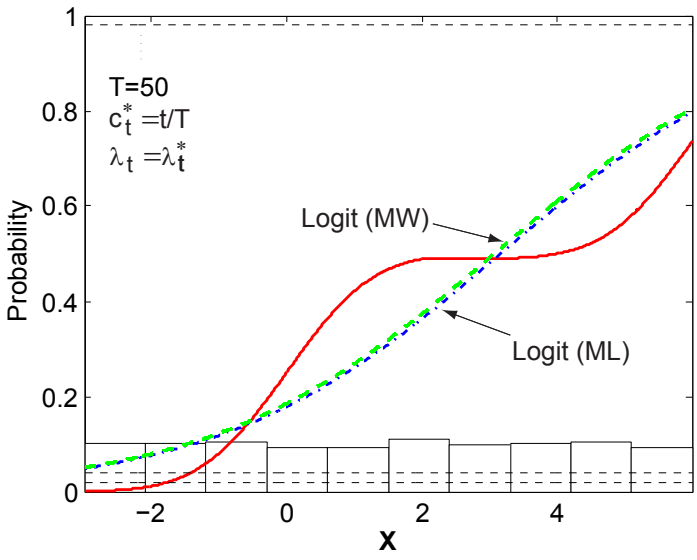

Figure I.4: The equivalence of ML and MW under c_t^* and λ_t^*

weights can come from two different sources: individuals in the extremes could be highly regarded, or there could be many more individuals in the "tails" than in the middle.) Thus, ML estimation is roughly equivalent to trying to reproduce as closely as possible the optimal decision rules for individuals with high or low cutoffs. DMs with low cutoffs are those for whom the gains of making correct decisions when $Y = 1$ are high relative to the gains of making correct decisions when $Y = 0$, and the other way around for DMs with high cutoffs. Thus, we can interpret ML as being optimal when the important (or more numerous) DMs are those who value highly correct decisions in one specific state of the world, as opposed to individuals for whom correct decisions have roughly the same value for either future state of the world.

A remarkable feature of the result presented above is that it hardly places any restrictions on the shape of the conditional probability function $p(x)$ and the model specification $m(x, \theta)$. Another way of saying this is that the cutoff/weight combination described in Proposition I.2 works uniformly for a large class of model specifications and joint distributions of (Y, X), where "works" means "makes ML and MW equivalent". Thus, the model $m(x, \theta)$ can be grossly misspecified for $p(x)$ and ML will still be socially optimal as long as the population of decision makers and the social welfare function satisfy the conditions of Proposition I.2.

I.G Conclusion

In this paper we defined and formalized the notion of public forecasting. A public forecaster needs to construct a single forecast for a population of heterogeneous individuals who will then use the information provided in a range of decisions. Thus, a public forecaster faces an optimization problem which is in many ways more complex than that faced by "consultant", i.e. a forecaster who has a single DM as a client. We focused on the case in which the outcome to be predicted as well as the decisions to be made are binary.

We postulated that a public forecaster's goal is to fit a conditional probability model to the data by maximizing a social welfare function defined over the utilities of individual DMs who then use the estimated model to make their own decisions. Two important features of this optimization problem are the distribution of optimal cutoffs (which are determined by the preferences of the individual DMs in the population) and the welfare weights that the forecaster assigns to the individual DMs who share the same cutoff.

We provided sufficient conditions under which the "maximum social wel-

fare" (MW) estimator asymptotically coincides with the maximum likelihood (ML) estimator even when the conditional probability model is grossly misspecified. In particular, these conditions call for the cutoffs to be uniformly and "densely" distributed over the $(0, 1)$ interval with the forecaster placing a lot of weight on the group of DMs with very low or very high cutoffs. Cutoffs in the middle portion of the (0,1) receive lower weights. Thus, the social welfare weights which make ML and MW asymptotically equivalent exhibit a characteristic U-shape. This shape could arise, for example, if there were many DMs with very low or very high cutoffs, not many in the "middle", and the forecaster cared more or less equally about each individual. An important feature of this U-shaped scheme is that it "works" uniformly over a very large class of probability models irrespective of the issue of misspecification. In other words, maximum likelihood is "socially optimal" under the cutoff/weight combination given above, regardless of whether the estimated model is correctly specified for the true conditional probability function or not.

We do not claim that this U-shaped cutoff/weight combination is the only one that makes maximum likelihood socially optimal. The derivation of the result leaves open the possibility for other sets of cutoffs (i.e. populations of decision makers) and weights (i.e. social welfare functions) to have the same property. In other words, the conditions provided are merely sufficient—a more general characterization of these conditions remains an interesting topic for further research. Another direction this research can follow is towards understanding how much a given cutoff/weight combination causes MW to differ from ML.

This chapter was coauthored with Robert P. Lieli. The dissertation author was the principal investigator on this paper.

Appendix

Examples of Maximum Welfare (MW) estimation Figure I.5 on page 34 illustrates how the MW estimator operates for various combinations of types c_t and weights λ_t and how it compares to the ML estimator. The graphs are based on a random sample $\{(Y_i, X_i)\}$ of 1000 observations; the covariates are uniformly distributed over the (-3,6) interval, while the outcomes Y_i were generated according to the conditional probability function $p(x)$ shown in the graphs by the solid line. In each case the $m(x, \theta)$ is a simple logit model based on a linear index, which is misspecified for $p(x)$. The maximum likelihood estimate of the model is the same in each panel, as it does not depend on the distribution of the cutoffs or the welfare weights.

In panel (a) the welfare weights are equal for all three types and the cutoffs positioned so that it is possible for the logit model to reproduce the individually optimal decision rule for each decision maker, which is of course socially optimal as well. Hence, the MW estimator "finds" this fit, while ML gives a global approximation to $p(x)$, which is suboptimal for DMs 2 and 3. In panel (b) we have a more dispersed set of cutoffs, but the social welfare function is virtually ignorant of DM 1. Hence, the MW estimator fits the model so as to reproduce (approximately) the individually optimal decision rules for DM 2 and 3. Finally, in panel (c) the cutoffs and weights are dispersed enough so that the MW estimator produces a global approximation to $p(x)$ very similar to the fit achieved by ML. We want to find general conditions under which this is the case.

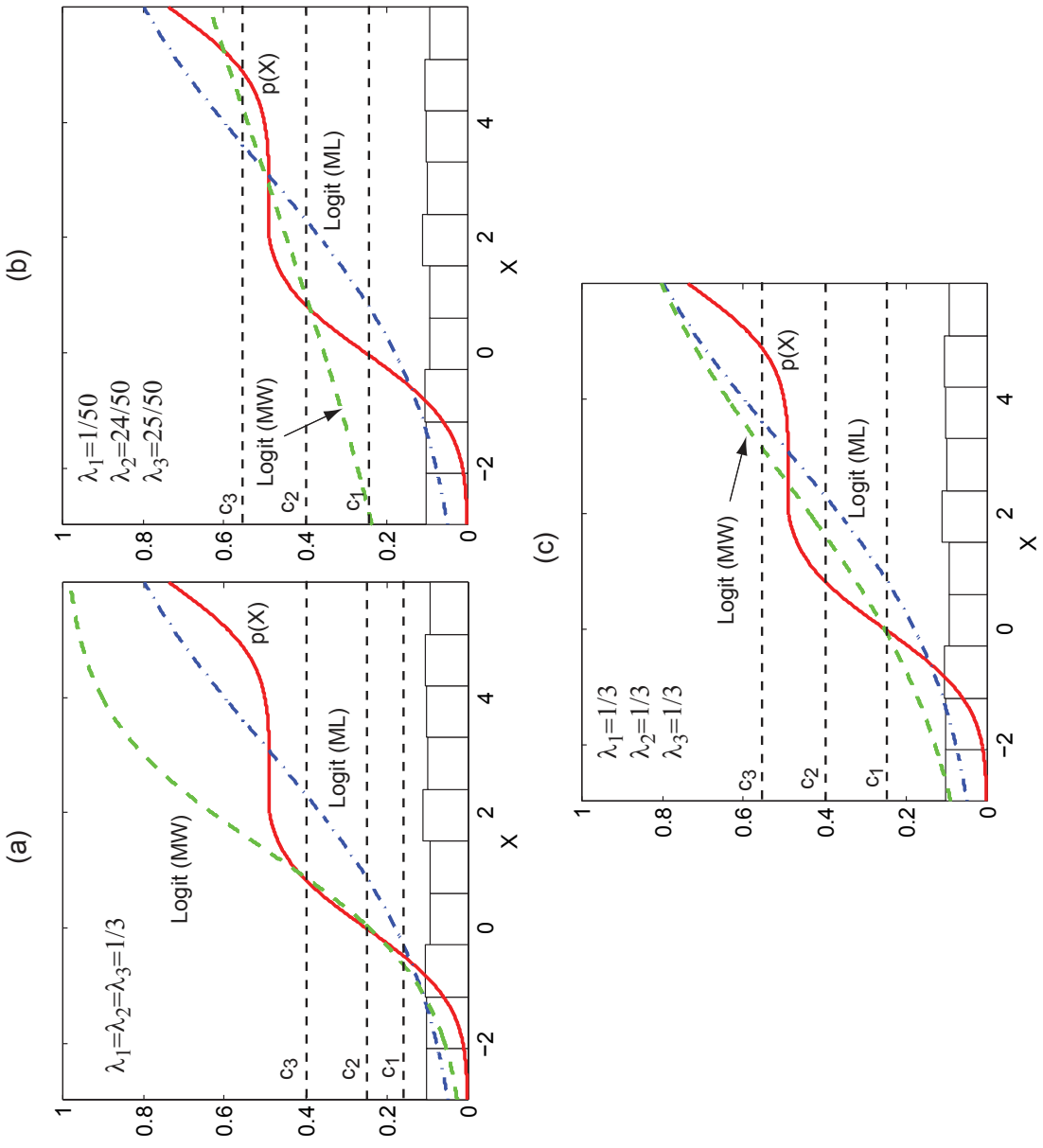

Figure I.5: Examples of maximum welfare (MW) estimation

Proof of Proposition 1 We will first bound the absolute difference between $L_{C_T}(\theta)$ and $L(\theta)$ in several steps.

$$
\begin{aligned}
&|L_{C_T}(\theta) - L(\theta)| \\
&\leq \sum_{t=1}^{T} E\big\{(Y+1)\big|\log(c_t) - \log[m(X,\theta)]\big|1[c_{t-1} < m(X,\theta) \leq c_t]\big\} \qquad \text{(I.11)} \\
&+ \sum_{t=1}^{T} E\big\{(Y+1)\big|\log(1-c_{t-1}) - \log[1-m(X,\theta)]\big|1[c_{t-1} < m(X,\theta) \leq c_t]\big\}.
\end{aligned}
$$

Treating the first and last terms in (I.11) and (I.12) separately, we replace each term in these sums with a further upper bound. This yields

$$
\begin{aligned}
&|L_{C_T}(\theta) - L(\theta)| \\
&\leq 2E\big\{\big|\log[m(X,\theta)]\big|1[0 < m(X,\theta) \leq c_1]\big\} \qquad \text{(I.12)} \\
&+2\sum_{t=2}^{T-1} E\big\{\big|\log(c_t) - \log(c_{t-1})\big|1[c_{t-1} < m(X,\theta) \leq c_t]\big\} \qquad \text{(I.13)} \\
&+2E\big\{\big|\log[m(X,\theta)]\big|1[c_{T-1} < m(X,\theta) \leq 1]\big\} \qquad \text{(I.14)} \\
&+2E\big\{\big|\log[1-m(X,\theta)]\big|1[0 < m(X,\theta) \leq c_1]\big\} \qquad \text{(I.15)} \\
&+2\sum_{t=2}^{T-1} E\big\{\big|\log(1-c_{t-1}) - \log(1-c_t)\big|1[c_{t-1} < m(X,\theta) \leq c_t]\big\} \qquad \text{(I.16)} \\
&+2E\big\{\big|\log[1-m(X,\theta)]\big|1[c_{T-1} < m(X,\theta) \leq 1]\big\}. \qquad \text{(I.17)}
\end{aligned}
$$

Using the upper bound contained in (I.12) to (I.17), it is now possible to show that $L_{C_T}(\theta)$ can approximate $L(\theta)$ to an arbitrary degree of accuracy, uniformly in θ. Fix $\epsilon > 0$. To handle (I.12), observe that for any $\theta \in \Theta$,

$$
\begin{aligned}
&\big\{0 < m(X,\theta) \leq c_1\big\} \subset \big\{\big|\log[m(X,\theta)]\big| \geq |\log(c_1)|\big\} \\
&\subset \big\{\sup_{\theta\in\Theta}\big|\log[m(X,\theta)]\big| \geq |\log(c_1)|\big\}.
\end{aligned}
$$

Therefore,

$$E\big\{\big|\log[m(X,\theta)]\big|1[0 < m(X,\theta) \le c_1]\big\}$$
$$\le E\Big\{\sup_\theta\big|\log[m(X,\theta)]\big|1\big[\sup_\theta|\log[m(X,\theta)]| \ge |\log(c_1)|\big]\Big\}.$$

By Assumption I.3(c), one can choose $c_1 > 0$ small enough so that this last quantity is less than $\epsilon/12$, making (I.12) less than $\epsilon/6$ (c.f. Resnick 1999, p. 183). Without loss of generality we can assume $c_1 < 1/2$, which makes (I.15) less than $\epsilon/6$ as well. By a similar argument, one can choose a point $c_{T-1} \in (c_1, 1)$ close enough to unity such that (I.17) is also less than $\epsilon/6$. (The fact that T is not yet determined is immaterial. The eventual value of T does not affect the position of this point: $T-1$ is only a label that will remain attached to this point as T changes.) Again, w.l.g. we can assume $c_{T-1} > 1/2$, which makes (I.14) less than $\epsilon/6$ as well. Note that the choice of c_1 and c_{T-1} is independent of the value of θ.

To handle (I.13) and (I.16), observe that the function $\log(\cdot)$ is uniformly continuous over the closed interval $[c_1, c_{T-1}]$. Hence, there exists $\tilde{\delta} > 0$ such that $x, y \in [c_1, c_{T-1}]$ and $|y-x| < \tilde{\delta}$ implies $|\log(y) - \log(x)| < \epsilon/12$. Choose the points $c_2 < c_3 < \ldots < c_{T-2}$ in (c_1, c_{T-1}) such that $\max\{c_t - c_{t-1},\, t = 2, \ldots, T-1\} < \tilde{\delta}$. This implies that (I.13) is strictly less than

$$2\frac{\epsilon}{12}\sum_{t=2}^{T-1} P[c_{t-1} < m(X,\theta) \le c_t] \le \frac{\epsilon}{6}.$$

The same argument shows that (I.16) is also bounded by $\epsilon/6$. Again note that the choice of $\tilde{\delta}$ was independent of θ.

Let $\delta = \min\{c_1, 1 - c_{T-1}, \tilde{\delta}\}$. By the arguments above, any partition $C'_{T'}$ such that $\max\{c'_t - c'_{t-1},\, t = 1, 2, \ldots, T'\} < \delta$ satisfies $|L_{C'_{T'}}(\theta) - L(\theta)| < 6(\epsilon/6) = \epsilon$ for all θ. Q.E.D.

References

[1] Amemiya T. (1985): Advanced econometrics. *Oxford: Basil Blackwell.*

[2] Bond S. A. and S. E. Satchell (2004): Asymmetry, Loss Aversion and Forecasting. Department of Economics, University of Cambridge.

[3] Crone F., Lessmann S. and Stahlbock R. (2005): Utility Based Data Mining for Time Series Analysis - Cost-sensitive Learning for Neural Network Predictors. Procedings of the Workshop on Utility-Based Data Mining. Chicago, Illinois, August 2005.

[4] Diebold F. X., Gunther T. A. and Tay A. S. (1997): Evaluating Density Forecasts with Applications to Financial Risk Management. Department of Economics, University of Pennsylvania.

[5] G. Elliott and Lieli, R. P. (2005): Predicting Binary Outcomes. Department of Economics, University of Texas, Austin.

[6] Gallant, A.R (1997): *An Introduction to Econometric Theory.* Princeton University Press. Princeton, NJ.

[7] Granger, C.W.J. and M.H. Pesaran (2000a): A Decision Theoretic Approach to Forecast Evaluation. In: W. Chan, W. Li and H. Tong (eds.) *Statistics and Finance: An Interface, London: Imperial College Press*: 261-278.

[8] Granger, C.W.J. and M.H. Pesaran (2000b): Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting* 19: 537-560.

[9] Lieli, R. P. (2004): Three Essays on the Prediction of Binary Variables. Ph.D. Dissertation. Department of Economics, University of California, San Diego.

[10] McCulloch R. and P. E. Rossi (1990): Posterior, Predictive, and Utility-based Approaches to Testing the Arbitrage Pricing Theory. *Journal of Financial Economics* 28: 7-38.

[11] Manski, C.F. (1975): Maximum Score Estimation of the Stochastic Utility Model of Choice. *Journal of Econometrics* 3: 205-228.

[12] Manski, C.F. (1985): Semiparametric Analysis of Discrete Response. Asymptotic Properties of the Maximum Score Estimator. *Journal of Econometrics* 27: 313-333.

[13] Pesaran, M.H. and S. Skouras (2001): Decision-Based Methods for Forecast Evaluation. In: M.P. Clemens and D.F. Hendry (eds.) *Companion to Economic Forecasting, Basil Blackwell.* 39: 863-883.

[14] Resnick S. (1999): *A probability path.* Boston: Birkhaeuser.

[15] West D. K., Edison H. J. and Cho D. (1993): A Utility-based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics.* 35: 23-45.

[16] White H. (1996): *Estimation, Inference, and Specification Analysis.* New York: Cambridge University Press.

Chapter II

Mistakes in the Loan Market: A One-Armed Bandit Model

Abstract[1]

This paper analyzes the situation of a loan officer that makes sequential decisions on whether to grant loans or not to applicants. Before making each decision, the loan officer can observe some characteristic of the applicant, and in the case that a loan is granted, he can observe if whether it is paid back or not. On the other hand, when a loan is denied the officer cannot observe the applicant's behavior. This selection problem will have an effect on the lender's ability to learn about the population of borrowers. We find that in some cases the lender will be able to make correct decisions in the long run, but in other cases not, i.e. he can make mistakes forever. The crucial factor that determines whether full learning will occur is the nature of the information that the lender observes from

[1]I am deeply indebted to Ross Starr for guidance and encouragement throughout my dissertation and for his invaluable advice with this essay. I thank Joel Sobel for useful comments on an earlier version of the paper, as well as participants in the UCSD Economic Theory reading group.

the applicant before making each loan.

II.A Introduction

The objective of this research paper is to recognize and address a problem that arises naturally in those settings in which decisions have to be made under uncertainty, and the result of those decisions affect the ability of the decision maker of observing the outcome of the variable of interest when it is realized. In particular, consider a loan officer who has to decide whether to grant or deny loans to a group of applicants. Assume that those decisions are made sequentially, and the next decision will be made after the expiration of the current loan. Abstracting from differences in borrowers characteristics (to which we will go back below), the loan officer problem is at any time to decide to grant the loan or not, based in his current beliefs about the probability that the loan will be repaid. Now, granting a loan implies two things: First, the officer takes a lottery, that will have a positive payoff if the loan is repaid and a negative payoff if it is defaulted on. Second, by granting a loan the officer will also be able to observe the behavior of the borrower (i.e. if he pays the loan back or not) and can use that information to update his beliefs about the probability that a loan is paid back. Therefore, there is an extra value in granting a loan beyond that of the lottery that it implies, and that is the value of the new information gained by being able to observe the response of the borrower. On the other hand, when a loan is denied the lottery taken has a certain payoff, and no new information is gained about the distribution of applicants.

The problem exposed above is of the form of a general class of problems known in the statistical literature as the "Two-armed Bandit Problem". Two armed bandit problems were introduced first in statistics, following ideas laid down in Thompson (1933). Early treatments of the two-armed bandit problem in statistics can be found in Bradt, Johnson and Karlin (1954) and Bellman (1956). Berry

(1972) provides a good summary of the literature. In its general form, the two-armed bandit problem consists of an experiment design in which at each stage the experimenter has to decide to sample from either of two binary distributions, and can observe the result of each trial (i.e. a success or a failure) before deciding from which distribution to sample the next time. The name given to the problem comes from gambling, since this experiment design is analogous to the problem of a gambler that can play in either of two slot machines, both of them with two possible outcomes, success or failure, and with unknown probabilities of success. The gambler pulls one of the arms, observes the outcome, and then decides which arm to pull next.

In economics bandit models have been used for example to model decisions in labor markets and pricing under demand uncertainty. Rothschild (1974) uses a two-armed bandit type specification to model the situation of a store owner deciding which price, between two possible prices, to charge to the next customer to enter the store. As different stores face a different sequence of buyers, the model explains how a price distribution can be observed in a given market for the same good, a deviation of the "law of one price".

Our problem, described in the first paragraph, is a special case of the two-armed bandit problem, since one of the arms has a known outcome (i.e. if a loan is denied the outcome is known). This special case has been named the "One-armed Bandit" problem, and was studied by Bradt, Johnson and Karlin (1954). This specification is simpler than the general problem, and has the appeal that many intuitive properties, that do not in general hold for the two-armed bandit case, are true in this special case.

In this paper we concentrate on the application of the loan officer pre-

sented above. It should be noticed though that this problem is identical in form to other interesting problems in economics. In general, the results in this paper carry over to other situations in which the decision making process is such that in the situation that a negative decision has been made, no new information is gained about the underlaying probability distribution of the individual that the decision was made about. Another important application is university admissions. Typically, in university admissions, at least at a first stage, applicants are evaluated using a small number of characteristics, as SAT score and high school GPA for example. The university sets a cutoff in the SAT-GPA space, which is usually interpreted as the cutoff line above which the probability of obtaining a certain minimum GPA or better in college is greater or equal to a given preset target, and rejects all remaining applicants. The problem with this procedure, as stated above, is that then the university can only observe the performance of the accepted individuals, but not of the rejected. This problem is very similar to the one developed here, and most of the insight gained in our problem can be carried over to it.

The problem of correctly identifying "credit worthy" applicants has been previously approached from an econometric point of view in the credit scoring literature. Altman (1968) is usually credited with pioneering the use of statistical methods in credit scoring. Boyes, et. al. (1989), Crook and Banasik (2002), Feelders (2002), Fortowsky and LaCour-Little (2001), Kraft, et. al. (2003) and Ladd (1998) are some examples. A good review of some of those methods can be found in Altman (1981).

In this paper we work in a setting in which the loan officer can observe a random variable X_i for applicant $i = 1, 2, 3, ...$, that distinguishes applicants from one another. In particular, we think of X_i as being some borrowers' characteristic,

as for example race, gender, income, household size, occupation, education, etc., that the loan officer can use in order to decide whether to grant a loan or not. In our setting the lender observes X_i for applicant i, decides whether to grant the loan or not, and then observes if the loan was paid back in the case that it was granted.

The objective of the current paper is to investigate the consequences of this decision making setting, in which no information can be obtained from the rejected individuals, on the quality of the decisions that the loan officer will be expected to make. In particular, we can ask the question: can we expect that credit worthy individuals (in a sense that will be made precise later) will always receive loans in the long run, and on the other hand, individuals not credit worthy will be rejected in the long run?

In fact, we find that the answer to the previous question depends on the nature of the variable X_i. Below we consider two cases, when X_i are qualitative variables, such as race or gender, and when X_i are quantitative variables, as for example income or age. The main result of section II.C states that when X_i are qualitative, the loan officer may make mistakes, even in the long run. Those mistakes are such that it is possible that a certain group of the population of applicants with a given value of X_i, although profitable to be lent to, may be rejected from some moment on, and will not be able obtain loans ever again.

The intuition behind that result is as follows: Assume that at a given stage the state of information of the loan officer dictates that an applicant with certain characteristics has to be denied a loan. Then in the next stage an applicant with the same characteristics should also be rejected, since no new information was gained from the rejected individual (i.e., the state of information at the following

stage will be identical, and consequently will lead to the same decision). Now bear in mind that individuals with those characteristics may, as a group, have a high enough probability of returning the loans (in the sense that it would be profitable to be lent to). Therefore if mistaken decisions are made about a certain group of applicants, those mistakes will be made subsequently.

The problem in this case arises because a "bad realization" in the sequence of applicants may occur for that group, making the loan officer believe that their probability of paying back is lower than the truth (e.g., unless the probability of payback for a group is equal to 1, there is a positive probability that the first N applicants from that group default). Then the loan officer will form beliefs about these applicants, and at a certain point it may become optimal for him to stop lending to that group. From that stage on no new information will ever be collected, and the group will never receive loans again.

In section II.D we consider the case of X_i being quantitative. In this case there will usually be an order in X_i, as for example with income, so applicants can be ranked by their value of X_i. In general when this is the case, it will be natural to expect that the probability of a loan being paid back, conditional on X_i, will change somehow smoothly with X_i. For example, we would not expect that the probability of pay back is substantially different for an applicant with an income of \$25,000 than for an applicant with an income of \$24,999. In view of this observation, in section II.D we will assume some smoothness of the probability that a loan is paid conditional on X_i, as X_i changes. Under this assumption, we will show that the kind of mistakes that the loan officer could make in the case of X_i being qualitative will not be expected to be made in this case.

The intuition for this result goes as follows: As was pointed out above, in

the case of qualitative X_i, some group that has been excluded from getting loans will never be included again. On the other hand, when X_i is quantitative, even when a group has stopped getting loans, the loan officer may, in the future, return to make loans to it. It is true that no new information is being gathered about those applicants, but in this case information being collected about applicants with other values of X_i can be useful to update beliefs about the first group. The following example shows a situation in which this is the case:

Example II.1 *Without loss of generality, assume that the probability of pay back is monotonically increasing in X_i. In this case, typically at any stage the decision rule of the loan officer will have the form "grant the loan to applicants with a value of X_i greater or equal than a certain value". For instance, assume that in reality the minimum income that would be required to grant a loan is \$23,000, but because of previous experience at a certain stage the decision rule dictates "grant the loan to applicants with income of \$25,000 or more". This rule will of course prevent the lender to get information about individuals with \$24,999 of income. Now, if that rule is used for some time, eventually the loan officer will learn (under certain regularity conditions) that the probability of a loan being paid back for incomes of \$25,000 is in excess of the minimum probability required to grant a loan. As the loan officer knows that the conditional probability does not change much with X_i, he can then use that information to revise the decision rule, and include some individuals with less than \$25,000 of income.*

By correcting the decision rule using information gathered for individuals close to the cutoff the loan officer will then be able to include applicants that were incorrectly excluded before, something that was not possible when X_i was qualitative. The main result of section II.D will be that when X_i is quantitative, under

some regularity conditions on the underlying conditional probability function, the loan officer will be expected to make correct decisions in the long run. The sense in that those will be correct decisions is that the cutoff in the variable X_i will be correctly set, and only applicants with a high enough probability of paying the loan back (as to be profitable to be lent to) will be granted loans.

A word has to be said here about the implications of these results. The denial of loans to groups that deserve them can be interpreted as discrimination. The reason is that applicant groups that are credit worthy are being denied loans on the base of belonging to that group. The cause, of course, is that unfortunate previous experience makes the lender do so. So, even when not being moved by any discriminatory motive, a profit maximizing lender may exclude a group of individuals from loans, and since no new information will be gathered about these individuals from then on, the group will be excluded forever. As we pointed out above that can be the case when X_i is qualitative, but not when it is quantitative. Therefore the message of these results is that when lenders are profit maximizers, "discrimination" is much more likely to occur related to a qualitative variable, such as race or gender, than related to a quantitative variable, as for example income or age (quotation marks are used to indicate that this is not discrimination in the strict sense of the word). This suggests that the only cases in which government intervention may be justified is when discrimination is related to variables to the first type. In that case, intervention may take the form of temporarily subsidize loans to a certain group, until enough information can be collected about them to be confident about their probability of returning the loan. After that point, profit maximizing lenders can be expected to make correct decisions, and therefore subsidies can be removed. Bear in mind that here we are not considering the case

in which the government wants to favor some particular group, and instead the only reason for intervention is to solve the problem that can arise due to lack of information about that group.

It should also be mentioned that the general problem of inability to observe the outcome of certain decisions can also be tackled from an econometric point of view. In recent years the there has been a growing interest in the economic forecasting literature towards taking into account the intended use of the forecasts (i.e. decision making) at all stages of the forecasting practice. In particular, the observation that traditional forecasting loss functions do not necessarily produce optimal forecasts for a given use is central in this new area of research, as pointed out by Granger and Pesaran (2000a, 2000b), Lieli and Elliott (2004), Pesaran and Skouras (2002) and Pesaran and Timmermann (2004) among others. As some of these papers argue, the correct measure to compare the ex-post performance of different forecasts has to be based on the preferences of the decision maker, and they propose statistics to perform such comparisons or evaluations. What has been overlooked so far in this literature is the fact that when decisions are made, in some cases, those decisions will affect the ability of the forecaster to observe the realizations of some of the variables of interest, and therefore will affect his ability to perform the desired comparisons. This angle of the problem will not be explored in the current paper, but appears as interesting for future research.

The paper will be organized as follows: In section II.B we present the basic framework that will be used to study the problem. In section II.C we consider the case when X_i is a qualitative variable, and we state and prove the results mentioned above. In section II.D we tackle the case of X_i being quantitative, and show how in that case the problems that could arise in section II.C are not present. Finally,

section II.E contains the conclusions and directions for further research.

II.B Basic Framework

To make our problem concrete, say that there is a population of individuals, and for each of them the loan officer will make a decision that will affect an ex-post payoff. Each individual is characterized by a pair (X_i, Y_i), where X_i is a vector of covariates or characteristics of each individual and Y_i is the outcome from the experiment (i.e. it is a binary random variable that takes the value 1 if the individual will pay the loan and 0 if the loan will be defaulted on). Notice that if the loan is denied, the individual does not have the chance of repaying it or defaulting on it. However, we assume that there exists a random variable Y_i that will determine whether the loan will be repaid if it is granted. If the loan is not granted for individual i, then the loan officer cannot observe the value of the variable Y_i. We assume that (X_i, Y_i) are iid for $i = 1, 2, ...$

Next, we assume that for each individual $(i = 1, 2, ...)$ the decision maker has a utility (payoff) function

$$U(a_i, Y_i)$$

where a_i is an action to be taken, in our case whether to grant a loan or not. Assume that a_i is binary variable, taking the value 1 if the loan is granted and 0 if it is not. At this point it is important to make clear that we are assuming here that the different decisions do no not affect the outcome, just the ability of the officer to observe it. Formally, we say that $p(x) = \Pr[Y = 1 \mid X = x]$ does not depend on a (notice that we have dropped the index i since we are assuming that (X_i, Y_i) are iid). In terms of our example, this assumption of $p(x)$ not depending

Table II.1: The loan officer's utility function

	$a = 0$	$a = 1$
$Y = 0$	0	z
$Y = 1$	0	w

on a will be true for instance in the case of small consumption loans, in which case the fact that the individual gets the loan will not affect either his ability nor willingness to repay.

Given the binary character of the variables a_i and Y_i, we can assign values to the utilities for each combination of a_i and Y_i. Notice, however, that when the loan is denied ($a_i = 0$), utility will be the same whether the applicant would have paid the loan back or not. We can tabulate utility values as in Table II.1, where we normalized the utility of not granting the loan to zero. We assume that $w > 0 > z$.

In our model, applicants arrive sequentially, and the loan officer can observe the results of the last loan before deciding whether to grant the next one or not. We assume that the officer objective is to maximize the discounted sum of the payoffs (utilities) from the loans. The discount factor will be b, such that $0 < b < 1$.

In the basic framework, we will assume that all applicants have identical characteristics, i.e. the value of the vector X_i is the same for all applicants. This will be our benchmark case. The analysis of this case is interesting since it can be used as a benchmark against which more general cases can be compared.

The problem of the loan officer at any stage is to decide whether to grant the loan or not to the applicant in turn. Granting the loan amounts to taking a lottery that will pay w with probability p and z with probability $(1 - p)$. Denying

the loan implies getting 0 for sure. The parameter p is not known to the loan officer. Instead, we assume that he has prior beliefs about it represented by the probability distribution function $F(p)$.

If the problem had only one stage, the officer will want to grant the loan whenever $wp+z(1-p) \geq 0$. In a multi-stage setting, however, there is an additional value to granting a loan, beyond that of the lottery that action implies, and that is the value of the new information gained about the distribution of p. This will be key in our problem, since the loan officer may sometimes give loans whose (utility) expectation is negative, just because the value of information is lager than the expected loss in that stage.

Before we can characterize the officer decision problem, we need to introduce some notation. Define

$V_{m,n}$: Expected return obtained proceeding optimally, after $m+n$ loans have been granted, of which m have been repaid and n have been defaulted on.

$F_{m,n}$: Updated beliefs about p, starting from initial beliefs F and after $m+n$ loans have been granted, of which m have been repaid and n have been defaulted on.

With these definitions, we can write an expression for $V_{m,n}$ for the cases in which a loan is granted and in which a loan is denied. In case in the current stage a loan is granted, then

$$V_{m,n} = \int_0^1 p dF_{m,n}(p)[w + bV_{m+1,n}] + \int_0^1 (1-p) dF_{m,n}(p)[z + bV_{m,n+1}] \quad \text{(II.1)}$$

The interpretation of this equation is the following: The mean probability of the loan being paid back, conditional on having beliefs $F_{m,n}$, is $\int_0^1 p dF_{m,n}(p)$. In that case, then the officer gets a payoff of w in this stage, and in the next stage (when behaving optimally) the expected future payoff will be $V_{m+1,n}$, since at that time

one more loan has been paid back. Similarly the mean probability of a default, conditional on having beliefs $F_{m,n}$, is $\int_0^1 (1-p)dF_{m,n}(p)$. In that case the officer gets a payoff of z in this stage and in the next stage the expected future payoff will be $V_{m,n+1}$.

Similarly to equation (1), we can write an expression for the value of denying the loan at the current stage as follows:

$$V_{m,n} = 0 + bV_{m,n} \tag{II.2}$$

which implies that $V_{m,n} = 0$.

From equations (1) and (2) we conclude that

$$V_{m,n} = \max\{0, \int_0^1 pdF_{m,n}(p)[w+bV_{m+1,n}] + \int_0^1 (1-p)dF_{m,n}(p)[z+bV_{m,n+1}]\} \tag{II.3}$$

We can now state and prove the first result of the paper:

Theorem II.1 *If a loan officer that behaves optimally denies a loan at any stage, then he will deny loans at every stage after that.*

Next, we present the proof of theorem II.1.

Proof: Assume that at a certain stage, after $m+n$ loans have been granted, where m have been repaid and n have been defaulted on, the optimal decision is to deny the current loan, i.e.

$$\int_0^1 pdF_{m,n}(p)[w + bV_{m+1,n}] + \int_0^1 (1-p)dF_{m,n}(p)[z + bV_{m,n+1}] < 0$$

but contrary to the theorem, after r denials optimal behavior allows to grant a loan. Then the expected payoff of this path is

$0+b0+b^2 0+...+b^{r-1}0+b^r\{\int_0^1 pdF_{m,n}(p)[w+bV_{m+1,n}]+\int_0^1(1-p)dF_{m,n}(p)[z+bV_{m,n+1}]\}$

But notice that this expression is negative by hypothesis, since

$$\int_0^1 pdF_{m,n}(p)[w + bV_{m+1,n}] + \int_0^1 (1-p)dF_{m,n}(p)[z + bV_{m,n+1}] < 0.$$

On the other hand, denying the loan for the r^{th} period yields an expected payoff of 0. Then the proposed behavior is not optimal and a loan must be denied in all future stages. Q.E.D.

The intuition behind the previous result goes as follows: We mentioned above that if the loan is conceded the extra information will be used to update previous beliefs about p, whereas if the loan is denied, no extra information will be collected. It follows that if a loan is ever denied, it should be denied forever (since the state of information was such that the loan was to be denied and such information did not change). Then in some cases a loan will be denied forever if the information gathered so far is bad enough. Note, however, that nothing ensures that this is the correct decision. In fact, we will see below that loans could be denied forever even when the probability of repay p is such that it will be profitable to grant loans to all applicants, i.e. when p is such that $wp+z(1-p) \geq 0$.

The previous result has another interesting implication, one that has been discussed in the literature on credit scoring. In the past has been some interest in analyzing if there is discrimination in the loan markets, in particular against some racial group. One can argue that in a profit driven market discrimination will not occur, and profitable loans will be granted. But as the above theorem shows, even a loan officer that is purely driven by profits may in fact discriminate against some group (i.e. loans will be denied to them even when the group as a whole is credit worthy, in the sense of being profitable to be lent to). The key here is that even when those borrowers are credit worthy, bad experience from the past may make the loan officer decide to deny loans at some stage. and when that happens, no

new information will be collected that will make him change his mind in the future.

In section II.C we will extend the basic model to the case in which not all applicants are the same, but instead the loan officer can observe some characteristic of them before making the decision of granting or not the loan.

II.C Different Types of Borrowers: The Qualitative Characteristic Case

In this section we analyze the case in which borrowers are not identical, as assumed above, but rather there is a characteristic, that we call X_i for $i = 1, 2, ...$ that distinguish different borrowers. For example, X_i could be income, race, marital state, number of children, etc. In general loans will be evaluated based on a number of characteristics, and not a single one. Here, however, we will assume that X_i is a single random variable. The reason for this assumption is that we want to investigate the difference that X_i being qualitative as opposed to quantitative makes on the quality of the decisions that the loan officer will make. In fact, we will show below that when X_i is qualitative, the loan officer will be exposed to making the same mistakes that we encountered in section II.B. In this case, the loan officer can use the knowledge about X_i in making decisions, but we will see that this knowledge will only protect him for making mistakes for some groups of borrowers, and not for all of them. On the contrary, when X_i is quantitative, we will show that the officer will be expected to make correct decisions in the long run.

In this section we will study the case in which borrowers are distinguished by a single characteristic X_i, which is a discrete random variable. We assume that before deciding whether to grant or not a loan to applicant i, the loan officer can

observe X_i. We will assume that X_i is binary for all i, taking the values 0 and 1. This assumption is only made for simplicity, and the results obtained will be true when X_i can take a finite number of different values.

We think of X_i as being a qualitative random variable, such as race or sex. In real credit scoring, there are some laws that prohibit lenders from using certain characteristics to make decisions on granting loans. Here we allow the loan officer to use this information. Then our treatment can be viewed in either of two ways: a) X_i is one of the variables that the lender can use to separate borrowers, or b) X_i is one of the variables that cannot be used, in which case we would be thinking of what would be the situation if not such laws would be in place.

From these interpretations the second appears to be the most interesting. This is so in view of controversy about the laws that aim to prevent discrimination in the credit markets. The controversy is about whether the market will work well without regulation (in the sense that individuals who are profitable to be lent to will get loans) or on the contrary if the market solution could involve discrimination, as for example racial discrimination. The argument in favor of the free market is that profit driven lenders will not want to discriminate against any group from which they can make a profit. We will see, however, that this may not be the case. The key here is, as pointed out above, that when bad experience is collected about certain group (that may be "good" borrowers as a group), the lender may, behaving optimally, stop lending to it. And in certain cases this could imply never getting new information about those individuals. We will show below that even in the case in which there are other groups getting loans, the information about these groups may not be enough for the officer to correct his mistakes, the "discriminated" group may not receive loans again.

Although this situation may look as discrimination, it should be pointed out that it arises from pure profit driven behavior. While it is true that "credit worthy" groups of individuals may be rejected, that result does not arise from any group preference (as for example racial preference) from the lender, but from its desire to maximize profits. In this sense this situation does not fit into the definition of discrimination. And in fact it could explain why groups of good borrowers may be excluded from loans in a perfectly profit driven market.

As indicated above, we assume that there is a random variable X_i for each applicant $i = 1, 2, ...$, that can take the values 0 or 1. Then from the perspective of the loan officer, each applicant is viewed as a pair (X_i, Y_i), where X_i is some borrower characteristic and Y_i is a random variable that will be 0 if the applicant will default on the loan (in case of it being granted) and will be equal to 1 if the loan will be paid back. It is important to note that X_i does not enter the lender's utility function, and the only value of observing it for the officer is to help him learn something about the probability of $Y_i = 1$ for the particular applicant. Throughout the paper we will assume that (X_i, Y_i) are iid for $i = 1, 2, ...$

We use the following notation: $p_0 = \Pr[Y = 1 \mid X = 0]$ and $p_1 = \Pr[Y = 1 \mid X = 1]$. The loan officer's beliefs about (p_0, p_1) are represented by the joint distribution function $F(p_0, p_1)$. Next we introduce the following assumptions

(A.3.1) The loan officer's beliefs about (p_0, p_1) are such that p_0 and p_1 are regarded as independent, so we write $F(p_0, p_1) = F_1(p_0)F_2(p_1)$, where F_0 and F_1 are the marginal distributions (beliefs) of p_0 and p_1 respectively.

(A.3.2) Initial beliefs about how X affects the probability of $Y = 1$ (subscripts have been dropped since (X_i, Y_i) are iid) are such that

$$\int_0^1 p_1 dF_1(p_1) > \int_0^1 p_2 dF_2(p_2) \tag{II.4}$$

In words, this assumption amounts to say that the loan officer regards customers with $X = 1$ as more reliable than customers with $X = 0$. This belief may come from some previous experience or prejudice, although we don't need to make this precise here.

Assumption (A.3.2) implies the following initial behavior rule:

"*If at the initial period a loan is to be granted to an applicant with* $X_i = 0$*, then a loan will also be granted to an applicant with* $X_i = 1$." Furthermore, we will assume

(A.3.3) The previous decision rule will be maintained forever.

Therefore, at any stage, if the state of information is such that an applicant with $X_i = 0$ is to be granted a loan, we constrain the loan officer to grant loans to individuals with $X_i = 1$. What this constraint does is to act as a protection for individuals with $X_i = 1$. The reason is that no matter how bad the realization they may get, they will keep receiving loans as long as $X_i = 0$ individuals do. The only way they will stop getting loans is that first $X_i = 0$ individuals are denied, and subsequently there is bad experience for $X_i = 1$ borrowers.

Now our goal is to show that even in this case, in which we are imposing an artificial constraint that will protect $X_i = 1$ applicants, there is a positive probability that those individuals will be denied loans in the long run, even when they are profitable to be lent to. The intuition for this result resembles that of the last section. Since if treated separately $X_i = 0$ individuals will be denied forever with some positive probability, and so do $X_i = 1$ individuals, then there will be a positive probability that both of these events happen. Also, note that given the fact that we are imposing an artificial constraint, and that constraint is acting as a protection to the $X_i = 1$ group, this case will represent a lower bound for

the general case, in the sense that if we remove the constraint, the probability of the group $X_i = 1$ to be discriminated against will be higher. Therefore showing that this probability is positive for the constrained case suffices to prove that it is positive for the unconstrained case.

The result in this section is contained in the following theorem:

Theorem II.2 *Assume that there are two groups of borrowers, indexed by* $X = 0$ *and* $X = 1$, *and that (A.3.1), (A.3.2) and (A.3.3) are true . Then there is a positive probability that the group* $X = 1$ *will be denied loans forever after a certain point.*

Proof: As we are assuming that there is no correlation between the probabilities p_1 and p_2, then each group can be considered separately. In view of this result, the situation reduces to two different cases the case in Theorem II.1. Q.E.D.

What this theorem asserts is that even in the case in which the loan officer can observe applicants characteristics before making decisions, when those characteristics are qualitative this will not be enough to prevent the officer from making mistakes with some positive probability, or in other words, this will not be enough to ensure that the loan officer will make correct decisions in the long run with probability 1. The reason is that when characteristics are qualitative, even when the decision maker may have some information about the ranking of characteristics in terms of the probability of pay back, there is no way of quantifying how big is the difference in that probability. As a consequence, the decision maker has to consider both groups separately. Therefore, the main result of the previous section, that once loans are denied to a group then they will be denied forever, still holds in this case.

We will see in the next section that this will not be the case in the situation in which the borrower's characteristic is a quantitative random variable. The key

difference is that in that case, the loan officer, however he decides to set his cutoff in terms of X_i, will be able to get information about the behavior of individuals that are "very close" to the cutoff (in the sense of having a value of X_i close to the cutoff value). That information will allow the officer to revise his beliefs down if necessary, and can possibly make him to start granting loans again to individuals with some value of X_i that were not getting loans at some point. In the next section we will explore the case in which X_i are quantitative random variables, and we will see that in that case the loan officer will not be expected to make mistakes in the long run. In fact, it will be shown that when X_i are quantitative, under some monotonicity assumption about $p(x) = \Pr[Y = 1 \mid X = x]$, the loan officer will (in the long run) set the right cutoff (in the X dimension) with probability 1.

II.D The quantitative Characteristic Case

In this section we study the case in which borrowers are distinguished by a single characteristic, a random variable X, but as opposed to the last section, that random variable is continuous, as for example household income. We are interested in studying this case since we expect to obtain fundamentally different results than in the previous section. In particular, it will be shown below that when borrowers characteristics are quantitative we will not expect the loan officer to make the same mistakes as in the qualitative characteristics case in the long run.

The intuition goes as follows: In the qualitative case, there is no obvious way of using information about a group of borrowers to make decisions concerning another group. This is so because characteristics are in general qualitative, and in general do not suggest any way in which the probability of pay back should be

related to them. Therefore, when a group of applicants is denied a loan, no new information is collected about them, and as a consequence they are to be denied loans forever.

On the other hand, when the characteristic is a continuous variable, the decision making procedure will typically involve setting a "cutoff" value for the variable X such that applicants with values X_i greater than the cutoff will get the loans and applicants with X_i below the cutoff will be denied. Again no new information is being collected for applicants below the cutoff. However, information about individuals that are granted loans but are "close" to the cutoff, in the sense of having a close value of X_i to the cutoff value, can be used to update the information set of individuals on the other side of the cutoff that are also close to it. Therefore, applicants of a certain value X_i can be considered for loans even after they have been denied in the past. In the long run, it will be shown that this feature will lead to the conclusion that the cutoff is set a the right value, in the sense that only applicants that are profitable to be lent to get the loans or that the applicants that are denied are so because they are not good credit risks.

Again, applicants arrive sequentially, and are indexed by $i = 1, 2, 3, ...$ As before, the loan officer can observe the variable X_i before deciding whether to grant the loan or not to applicant i. In this section $X_i \in \mathbf{R}$ for $i = 1, 2, 3, ...$ We denote by $F(x)$ the (marginal) distribution function of the variable X_i, $i = 1, 2, 3, ...$ Remember that (X_i, Y_i) are iid for $i = 1, 2, 3, ...$

The lender has initial beliefs $q_0(x)$ about the conditional probability $p(x) = \Pr[Y = 1 \mid X = x]$, for $x \in [a, b]$. Although $p(x)$ is unknown, we will assume that it is monotonically increasing in x for all $x \in [a, b]$. We will assume that $p(x)$ is a smooth function of x. In particular, we will assume that it is differ-

entiable for all x in the interior of the support of X:

(A.4.1) The support of X is the interval $[a, b] \subset \mathbf{R}$.

(A.4.2) $p'(x)$ exists for all $x \in (a, b)$.

(A.4.3) $p'(x) > 0$ for all $x \in (a, b)$.

We can now analyze the behavior of the loan officer. Because of our monotonicity assumption, we know (and so does the loan officer) that $p(b) > p(a)$. Therefore it must be the case that $q_0(b) > q_0(a)$. Assume that the lowest probability of payback required for a loan to be granted is k. To make the case interesting, we will assume:

(A.4.4) $q_0(b) > k$

i.e. according to the officer's initial beliefs, there are some values of X for which applicants are good credit risks. This assumption rules out the uninteresting case in which initial beliefs dictate that no applicant, regardless of it's X value, is credit worthy. In that case, no loan will be ever granted. We assume further that:

(A.4.5) $p(b) > k > p(a)$

or in other words, according to the true conditional probability, there are some applicants that are creditworthy and some that are not. Then there is a value $c \in [a, b]$ such that $p(c) = k$. The value c is the value at which the cutoff would be set if the true conditional probability function $p(.)$ would be known.

In addition to the assumptions on the true conditional probability $p(x)$ stated above, we add the following Lipswich condition:

(A.4.6) There exist $0 < g$, such that $0 < p'(x) \leq g$ for all $x \in [a, b]$.

This condition ensures that the probability of pay back does not change much with X. Remember that we are assuming that $p(x)$ is monotonically increasing in x, as it would be expected for example if the random variable X rep-

resents income. In light of this assumption, this is an appealing condition, since we wouldn't expect the conditional probability of payback to take big jumps as we change the borrower's characteristic X by a small amount. For example, we wouldn't expect that applicants with \$25,000 of income have a substantially higher probability of pay back than those with \$24,999 of income.

The following condition will also be useful in the proof of the main result:

(A.4.7) $\exists \varepsilon > 0$ such that loans will be granted to all applicants with $X_i \in [b - \varepsilon, b]$ for all $i = 1, 2, 3, ...$

This condition will ensure that there is a small interval in the support of the characteristic for which applicants with those values of X will always receive loans. This assumption is not very restrictive either. Remember that we were considering the case in which $p(b) > k$. In that situation, due to the continuity of $p(x)$, we know that there exist $\varepsilon > 0$ such that $p(x - \varepsilon) > k$. Furthermore, in view of the condition that $p'(x) \geq h > 0$, an ε with that property can be easily computed. Therefore the condition amounts to say that loans will be always granted to a group of applicants that are credit worthy.

We will denote by $q_i(x)$ the updated beliefs of the loan officer about the probability $p(x)$, after applicant i has been evaluated, and (if the loan was granted) his behavior has been observed. In view of assumption (A.4.3), for each $i = 1, 2, ...$ the function $q_i(.)$ defines a cutoff value, such that loans will be granted for all x such that $q_i(x) \geq k$. We call that cutoff value t_i, i.e. $q_i(t_i) = k$. In other words, t_i is the cutoff in the income dimension set by the lender before applicant $i + 1$ is evaluated. Then the decision rule for the lender will be

"grant a loan to applicant $i + 1$ if his income is larger than t_i, and deny him the loan otherwise".

To this point nothing has been said about the updating rule that the loan officer uses to correct his beliefs in view of new information. The typical assumption in the literature is that beliefs are updated using Bayes rule. Here we refrain from assuming a particular updating rule, since different rules may be preferred by different lenders. Instead, we prefer to impose some minimal condition that any sensible updating rule should have. In particular, what the next condition states is that the updating rule used by the loan officer will be such that his beliefs about the conditional probability of payback for a given interval in the support of X will converge (in the sense of convergence almost surely) to the true conditional probability for that interval when the number of observations in the interval grows to infinity. We introduce the following condition on the updating rule used by the loan officer:

(A.4.8) The updating rule used is such that $q_i(x) \rightarrow_{a.s.} p(x)$ as $i \rightarrow \infty$ for all x such that $t_i < x$ infinitely often.

The intuitive argument behind this condition is of course the strong law of large numbers. We know that the variables X_i are iid for $i = 1, 2, 3, ...$ Then when we say a "sensible updating rule", the meaning of that is that a law of large numbers applies for all values of x for which enough information is collected, in the sense that the cutoff falls to the left of x for infinitely many values of i. When this is the case, that particular x will be in the interval that will get loans for infinitely many periods.

Our objective here is to show that in this case the loan officer will not make the kind of mistakes that he could make in the qualitative characteristics case. The reason is that since he can get a very good estimate of the conditional probability $p(x)$ in the interval $[b-\varepsilon, b]$, he can then use that knowledge and extend

it to other values in the support of X. The key for doing that is the knowledge about the true conditional probability that the officer has, namely that contained in condition 3. We will prove that when the lender has that knowledge, then in the long run he will make correct decisions, in the sense that applicants will be granted loans if and only if $p(x) \geq k$.

The next theorem contains the main result of the section:

Theorem II.3 *Assume (A.4.1)-(A.4.8). Let $t_i \in [a, b]$ such that $q_i(t_i) = k$. Then $t_i \to c$ as $i \to \infty$.*

What the theorem claims is that in the long run the loan officer will use the right cutoff, in the sense that it will grant loans to credit worthy applicants and it will deny loans to bad credit risks.

Proof: We will show that $\limsup t_i = \liminf t_i = c$. We first show that $\limsup t_i = c$. Assume this is not the case. Then either $\limsup t_i > c$ or $\limsup t_i < c$. We want to show that in either of these cases the loan officer would be making suboptimal decisions.

Suppose that $\limsup t_i = t < c$.We know that, by monotonicity of $p(x)$, $p(t) < p(c)$. Now pick $\delta > 0$ such that $t + \delta < c$. Since $\limsup t_i < t + \delta$, almost all elements of the sequence $\{t_i\}$ fall in the interval $[a, t+\delta]$. Then there is N such that $t_i \leq t + \delta$ for all $i > N$. But that means that loans will be granted to applicants in the interval $(t + \delta, c)$ for all $i > N$, and therefore $q_i(x) \to_{a.s.} p(x)$ for $x \in (t + \delta, c)$. Now, as for all such x it is the case that $p(x) < p(c) = k$, this will imply that the loan officer is granting loans in the long run to applicants that he learned not to be credit worthy, in the sense of having a lower probability of returning the loan than the minimum probability required to grant a loan. Therefore the officer must be behaving suboptimally.

On the other hand, suppose that $\limsup t_i = t > c$. Again, by monotonicity of $p(x)$, $p(t) > p(c)$. For any $\delta > 0$ such that $t - \delta > c$, $t_i < t + \delta$ for almost all i. Also we know that $t_i > t - \delta$ infinitely often. Then by condition 5 above, since this is true for any δ, $q_i(x) \rightarrow_{a.s.} p(x)$ for $x \in (t, b]$, and by continuity of $p(x)$, $q_i(x) \rightarrow_{a.s.} p(x)$ for $x \in [t, b]$. Now by condition 3, we can pick δ such that $p(t - \delta) > p(c)$. But then the fact that $t_i > t - \delta$ infinitely often implies that the loan officer will set the cutoff infinitely often above a point for which he will know that the probability of pay back is larger than the minimum required. Therefore he is making suboptimal decisions. Then we have shown that $\limsup t_i = c$.

A symmetric argument shows that $\liminf t_i = c$. Then since $\limsup t_i = \liminf t_i = c$ we conclude that $\lim t_i = c$ and the proof is complete. Q.E.D.

The result contained in theorem II.3 is quite surprising. Its interpretation is the following: When the borrowers' characteristic that the loan officer uses is quantitative, then he may make mistakes in the short run, but in the long run all credit worthy individuals (and only those) receive loans. The key behind the result is the ordering inherent to the variable X, that can be translated to the conditional probability of payback. Then the loan officer is able to learn from individuals that are not getting loans, since they are related by X to some individuals that are being approved. That ability to learn from those who are rejected is the main difference with the qualitative case, and is the main reason behind the sharp difference in the results for both cases.

II.E Conclusion

The paper considered a decision problem in which a loan officer makes sequential decisions on whether to grant loans or not to individuals who apply

for them, based on the observation of some borrowers' characteristics that can be useful to distinguish among different borrowers in terms of their probability of paying the loan back.

It was shown that the kind of information contained in the observable characteristic X_i has an important effect on the quality of decisions that the loan officer will be expected to make in the long run. In particular, we distinguish between two cases, when X_i is a qualitative variable and when X_i is a quantitative variable. In the first case, the nature of X_i does not allow the loan officer from using information from one group of applicants to update his beliefs about another. Given that, if it is the case that a certain point in time the loan officer decides to deny loans to a certain group, that group will be denied loans from that moment on. The reason is because they are denied loans, no new information is being collected about those applicant. Therefore, if the state of information was such that they had to be denied loans at a certain point, it will stay that way from there on, and those individuals will no longer receive loans.

The important point in this case is that this situation of a group being excluded from the market from some moment on may happen even when those individuals are "credit worthy", in the sense of having a high enough probability of paying the loan back as to be profitable to be lent to. The reason why such situation can arise is because the initial realization of the payback sequence for those individuals may be so discouraging that the loan officer may decide to stop "buying" information about those individuals, and therefore will never learn that they are in fact credit worthy.

On the other hand, it was shown that when X_i is a quantitative variable, those problems are not expected to occur. When this is the case, typically we can

expect some "smoothness" in the function $p(x)$, the conditional probability of a loan being paid back when $X_i = x$. By smoothness we mean that we don't expect $p(x)$ to change much do to a small change in x. When this is the case, it was shown that in the long run the loan officer will be expected to make correct decisions, in the sense that it will grant loans to individuals for whom $p(x) > k$, where k is the minimum probability of payback that will make granting a loan profitable.

The reason why the problems encountered in the qualitative characteristic case are not present here is because in this case the loan officer, even when not gathering information for a certain group of individuals, will be gathering information for other individuals that can be related to them by proximity in the x dimension. Therefore the lender will be able to correct mistakes made in the past, and possibly include again among those who receive loans some individuals that where rejected in the past.

The results in the paper have strong policy implications. In fact, the message is that if we assume that lenders are profit maximizers, then we would not expect discrimination to occur related to quantitative variables. Therefore, the argument that low income people are being discriminated would not be supported by our results. On the other hand, discrimination could occur related to a qualitative variable, such as race or gender. In that case, intervention may be justified in the form of subsidies to loans for a certain group. Nevertheless, our findings would support such subsidies only for a period of time, until enough information can be obtained about the payback probability for the group in question. Beyond that, profit maximizing lenders behaving optimally should be able to make correct decisions and no discrimination is expected to happen.

It was also mentioned that the problem dealt with in this paper is one

application of a general type of problems, where the decision maker cannot observe the realization of certain variable of interest for certain decisions. Another important application is university admissions, since rejected applicants cannot be evaluated and therefore their performance cannot be observed.

In section II.C we assumed that there is no a priori correlation between the probabilities of payback of the groups of applicants. An extension to this work would be to relax that assumption. In that case, our intuition tell us that we could expect that the loan officer will be less exposed to make mistakes in the long run. However, it appears that those results will depend on the knowledge that the loan officer has about the correlation between the probabilities for different groups. This constitutes an interesting line for future research.

It was mentioned above that this problem can be approached for an econometric point of view as well. In particular, there has been interest in recent years in developing forecasting methods that take into account the preferences of the users of the forecasts. Also, there has been interest in developing methods to evaluate and compare forecasts from the point of view of the forecast user, taking as a measure of forecast efficacy the quality of decisions it leads to. What has been overlooked so far in this literature is the fact that when decisions are made, in some cases, those decisions will affect the ability of the forecaster to observe the realizations of some of the variables of interest, and therefore will affect his ability to perform the desired evaluations or comparisons. This angle of the problem appears as interesting for future research.

References

[1] Altman, E. I. (1968): Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23: 589-609.

[2] Altman, E. I., R. B. Avery R. A. Eisenbeis and J. F. Sinkey (1981): *Application of Classification Techniques in Business Banking and Finance.* JAI Press: Greenwich, CT.

[3] Bellman, R. (1956): A Problem in the Sequential Design of Experiments. *Sankhya* 16: 221-229.

[4] Berry, D. A. (1972): A Bernoulli Two-Armed Bandit. *Annals of Mathematical Statistics* 43: 871-897.

[5] Boyes W., D. Hoffman and S. Low (1989): An Econometric Analysis of the Bank Credit Scoring Problem. *Journal of Econometrics* 40: 3-14.

[6] Bradt, R. N., S. M. Johnson and S. Karlin (1954): On Sequential Designs for Maximizing the Sum of n Observations. *Annals of Mathematical Statistics* 27: 1060-1070.

[7] Crook, J. and J. Banasik (2002): Does Reject Inference Really Improve the Performance of Application Scoring Models? Working paper. Credit Research Centre, The School of Management, University of Edinburgh.

[8] Feelders, A. J. (2002): An Overview of Model Based Reject Inference for Credit Scoring. Working paper. Institute for Information and Computing Sciences. Utrecht University.

[9] Fortowsky, E. and LaCour-Little, M. (2001): Credit Scoring and Dis-parate Impact. Working paper. Wells Fargo Home Mortgage. Available at fic.wharton.upenn.edu/fic/lacourpaper.pdf

[10] Granger, C. W. J. and M. H. Pesaran (2000a): Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting* 19: 537:560.

[11] Granger, C. W. J. and M. H. Pesaran (2000b): A Decision Theoretic Approach to Forecast Evaluation. Working paper. Department of Economics, University of California, San Diego.

[12] Kraft, H., G. Kroisandt and M. Muller (2003): ssessing the Discriminatory Power of Credit Scores Under Censoring. Working paper. Fraunhofer Institut fur Techno-und Wirtschaftsmatemathik, Kaiserslautern, Germany.

[13] Ladd, H. (1998): Evidence on Discrimination in Mortgage Lending. *Journal of Economic Perspectives* 12: 41-62.

[14] Lieli, R. and G. Elliott (2003): A Flexible Framework for Predicting Binary Variables. Working paper. Department of Economics, University of California, San Diego.

[15] Pesaran, M. H. and S. Skouras (2002): Decision Based Methods for Forecast Evaluation. *In M. P. Clements and D. F. Hendry (Eds.) A Companion to Economic Forecasting.* Oxford: Basil Blackwell, 241-267.

[16] Pesaran, M. H. and A. Timmerman (2004): Real Time Econometrics. CESifo Working Paper No. 1169.

[17] Rothschild, M. (1974): A Two- Armed Bandit Theory of Market Pricing. *Journal of Economic Theory* 9: 185-202.

[18] Thompson, W. R. (1933): On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* 25: 285-294.

Chapter III

Existence of Competitive Equilibrium with Network Externalities

Abstract[1]

A Network Externality arises when the satisfaction that a consumer gets from the consumption of a given good depends (usually positively) on the number of consumers that consume the same good. A common feature of markets where NE are present is the phenomenon called "tipping", which is the tendency of one of the competing goods or protocols to win a substantial share of the market. We argue that for tipping to occur there must be some underlying indivisibility in the consumption space. In addition to the problems posed by externalities for existence of equilibrium, indivisibilities create discontinuous demand behavior (not

[1]I am deeply indebted to Ross Starr for encouragement and direction throughout my dissertation, as well as invaluable advice with this essay, and to Joel Sobel for useful comments on an earlier draft of this paper. Financial support from the Cal-(IT)^2 Institute is gratefully acknowledged. Responsibility for the remaining errors is entirely mine.

merely nonconvexity). In the paper we provide sufficient conditions for existence of competitive equilibrium with NE *and* indivisibilities. The key conditions are a large number of consumers and dispersion in their income distribution.

III.A Introduction

A Network Externality (NE in what follows) arises when the satisfaction that a consumer obtains from the consumption of a given good depends (usually positively) on the number of consumers that consume the same good[2]. Usual examples are the telephone and e-mail services. The fact that a new individual uses the service enhances the usefulness of the service for existing users (the service allows to interact an additional person).

A common feature of markets where NE are present is the phenomenon called "tipping", which is the tendency of one of the competing goods or protocols to win a substantial share of the market. We argue that for tipping to occur there must be some underlying indivisibility in the consumption space, that encourages consumers to coordinate on a given network good. In addition to the problems posed by externalities for existence of equilibrium, indivisibilities create discontinuous demand behavior (not merely nonconvexity). Intuitively, indivisibilities imply "holes" in the opportunity set, that can cause the demand correspondence to fail to be upper-hemicontinuous. Therefore the problem is more complicated that the one in Aumann (1966), that deals only with individual demand not being convex. The interaction of externalities and indivisibilities pose an interesting problem for existence of equilibrium. The goal of the present research is to address that problem.

In the paper we provide sufficient conditions for existence of competitive equilibrium with NE *and* indivisibilities combined. The key conditions are a large number of consumers and dispersion in their income distribution. In fact, we show

[2]The word number is used here in a broad sense here. A consumer may care differently about the addition of two different individuals to the network. However, the assumption that only the number of consumers matter simplifies the analysis greatly and therefore it is adopted often in the literature.

that in an economy suitably defined, with uncountably many consumers, there exists a competitive equilibrium if the income distribution is dispersed, a condition similar to the one used in Yamazaki (1978). Dispersion in the distribution of income (in a sense that will be made precise below) ensures that the set of consumers "jumping" in their demand behavior is negligible for any small change in prices, and therefore aggregate demand is indeed continuous.

NE distinguish from the usual concept of consumption or production externality in the following sense:[3] In a regular consumption externality the satisfaction that an individual (say consumer A) obtains from a given consumption bundle depends (positively or negatively) on the level of consumption of certain good (say good x_i) by another individual (say consumer B). An example would be B polluting the environment (the air A breaths) with his car. Now, the fact that B consumes more of the good in question (car rides) *may or may not* affect $A's$ relative valuations with regard to the goods in his consumption set. In the case of a NE, however, the level of consumption of a given good by B *does* change $A's$ relative valuation of this particular good with respect to other goods in his consumption set.

Also, for consumer A to be benefited (or hurt, if it is a negative NE) from $B's$ consumption he has to consume the good in question. Then if A is not buying the good we wouldn't expect a change in his welfare coming from the addition of B to the set of buyers of the good. We can express these concepts formally in the following way: A positive consumption externality arises when $\partial u^A(\mathbf{x}^A)/\partial x_i^B > 0$, whereas a positive network externality also requires that $\partial u^A(\mathbf{x}^A)/\partial x_i^B = 0$

[3]The distinction exposed here is for consumption externalities. Note however that NE are also possible in production, as for example when the use of a given production process by one firm makes this process more efficient for other firms that use it. The distinction with a regular production externality is analogous to that presented here.

if $x_i^A = 0$, where the superscripts denote to which consumer the consumptions belongs, $\mathbf{x}^A$ is A's consumption vector, x_i and x_j are some of the components of this consumption vector and the externality is in x_i. Note also that as pointed out above $\partial MRS^A_{x_i,x_j}(\mathbf{x}^A)/\partial x_i^B$ can have either sign in a general externality, but we should have $\partial MRS^A_{x_i,x_j}(\mathbf{x}^A)/\partial x_i^B > 0$ in the case of a positive NE. Then we see that a network externality specifies how $B\prime s$ consumption will affect A in a very particular way.

Most of the existing literature on NE belongs to the field of Industrial Organization, and some of it to Public Finance. A common feature of this literature is that in general it addresses the problem from a partial equilibrium point of view. Examples of this work are Dybvig and Spatt (1983), Farrell and Saloner (1985, 1986 a, 1986 b and 1988), Katz and Shapiro (1985, 1986 and 1994) and Rohlfs (1974). Discussions about NE are found in Besen and Farrell (1994) and Liebowitz and Margolis (1994).

Quite surprisingly, there has been almost no theoretical work on NE in a General Equilibrium framework. An exception is the work by Starr (1999), who presents a general equilibrium model of NE in which network goods are produced by firms using a technology characterized by set-up costs. He then shows that an Average Cost Pricing Equilibrium exist in this economy.

In this paper the problem is treated from a different perspective. Whereas in Starr's framework NE are due to cost sharing by consumers by joining the same network, here we concentrate in the case in which there is a direct effect from an individual's consumption of the network goods on people in the same network, the word direct meaning that this effect does not come through a reduction in price from cost sharing in an Average Cost Pricing setting, but rather that the

consumption of network goods by an individual enters other individuals' utility functions. Then we investigate existence of competitive equilibrium when these direct effects are present.

III.A.1 Network externalities and tipping

A common feature of markets with NE is the phenomenon called "tipping". This is the tendency for one of the competing goods or protocols to pull away from its competitors and win a substantial share of the market. we argue here that the occurrence of tipping is due not only to NE, which make advantageous for consumers to buy popular products if they want to interact with other consumers, but also to some indivisibility intrinsic to the good in question. In fact, if no indivisibility were present, there would be no need for a consumer to chose only one of the competing goods or protocols, since it would be possible for him to buy a mix of many of them in the desired quantities. And that would be better for him if variety is something he likes. But it is impossible for a consumer to buy 10 hours of usage of an operating system per month. Also to use many operating systems does not seem practical since it would be annoying to switch among them, and it would imply to duplicate set up costs, as for example the costs of learning how to use them.

The claim here is that the existence of the indivisibility is essential for tipping to occur. In fact we don't observe tipping in goods subject to NE such as restaurants meals, since individuals can go to many different restaurants in a given month. Typically we observe tipping in goods such as operating systems for computers, consumer electronics standards, languages, etc., goods that have indivisibilities and non-negligible set up costs.

Another reason for investigating the situation of NE with indivisibilities is that this case has been the most popular in the previous work. In fact most of the papers referenced above refer to the case where a set of consumers face the decision of buying one among several indivisible goods exhibiting the NE property. The most usual situation is that in which the goods are substitutes, and different network goods are incompatible with each other. Then under this setting consumers will try to buy popular goods, giving rise to tipping.

A second situation that has been analyzed before is that of a market where a given standard is being used, and a new technology appears. The decision of users of these technologies is then whether to switch to the new technology of stick with the old. If the new technology is superior to the old, then it might be efficient that all consumers switch. But if NE are important, it might be difficult to induce the first consumer to switch. In fact one possible outcome is that the market gets stuck with an inferior standard just because it is widely used. Again in most of this literature it has been implicitly assumed that the competing technologies are indivisible, in the sense that it is not possible for an agent to use a mix of them or to use some features of one and some of another. This implicit assumption is relevant in many situations, since in many cases the use of a given technology implies large setup costs and it is impractical to switch back and forth between two protocols. We claim that also in this case the results are influenced by the implicit indivisibility assumption, since if technologies could be used in arbitrary quantities (if it would be possible, without large costs, to mix the use of different technologies), then we expect agents to slowly start incorporating superior technologies to take advantage of their good features, keeping the previous to reap the benefits of NE. And that this process will lead to the new superior technology

to eventually replace the old.

These observations suggest that the case of NE with indivisibilities has indeed been the most studied in the literature. As it was pointed above, not much work has been done in General Equilibrium, though.

In the rest of the paper we model NE and indivisibilities in a General Equilibrium framework, and investigate their consequences on existence of competitive equilibrium. In doing so it will be important to keep in mind the distinction made above, since it suggests that we should model the network effects and the indivisibilities.

The problem of indivisibilities has been addressed previously in the literature. Yamazaki (1978) shows that in large economies a competitive equilibrium exists even when the individual consumption sets are allowed to be non-convex, as it is the case with indivisibilities in consumption. The usual problem in this setting is that individual demand behavior may be discontinuous, since at certain prices a small change in the price vector may cause a large "jump" in the consumption bundle of an agent. The key condition used in Yamazaki (1978) to solve this problem is an assumption on "dispersion of income". That condition says basically that for any price vector, there is no income level such that a large portion of the population has exactly that income. In Yamazaki's work that assumption allows for the "smoothing" effect of large numbers, that ensure that aggregate behavior is continuous even when individual behavior is not. The intuition goes as follows: In large economies, each consumer is small compared to the size of the population (in a sense that will be made clear below). Then, even when small changes in prices may bring large changes in consumption for a set of agents, under suitable conditions that set of agents will be "small" compared with the size

of the population, and therefore that "jump" in individual consumption will be small compared with aggregate consumption. The results in Yamazaki (1978) tell us that in large economies even though individual consumption may be discontinuous when goods are indivisible, aggregate consumption is indeed continuous and therefore a competitive equilibrium exists.

It is worth to mention that our problem in this paper, as well as Yamazaki's, are different from that in Aumann (1966). In Aumann's treatment preferences are allowed to be non-convex, but the consumer's opportunity set is assumed to be $\mathbf{R}_+$ (which is of course convex). In that setting individual demand is not necessarily convex valued, but it is upper-hemicontinuous, though. What Aumann (1966) shows is that in markets with a continuum of consumers even though the individual demand correspondence may not be convex valued, the aggregate demand correspondence will be. In our setting the consumption set is not convex, since we allow for indivisibilities in consumption. The consequence is that the individual demand correspondence may not be upper-hemicontinuous, and that problem will persist even in markets with uncountably many consumers. Therefore we will need to impose an assumption analogous to Yamazaki's dispersion of income to ensure continuous demand behavior in the aggregate.

It is not obvious, though, that results similar to those in Yamazaki (1978) will be true in our case. As it was mentioned above, the results in Yamazaki (1978) rely upon the fact that individual behavior is negligible in comparison to aggregate behavior, and that the set of consumers "jumping" to a very different consumption bundle is small for any price vector and a small change in prices. In the case of economies with NE it is not clear that this will be the case. In the presence of NE, when making decisions consumers are not only looking at prices but also at

what other consumers do. As mentioned above the occurrence of tipping is a well documented fact in these markets. Now the idea of tipping is opposed to the concept of small numbers, since tipping implies a substantial portion of the population buying the same network good (where the word "substantial" means that it is non-negligible with respect of the size of the population). It is possible then that under these conditions we observe jumps for large groups of consumers, and that could pose a problem for existence of competitive equilibrium.

Since it is not clear that in markets with NE a result analogous to Yamazaki's will obtain, the goal of the present research is to inquire about existence of a competitive equilibrium in an economy with NE and indivisibilities. In the remaining sections it will be shown that in such an economy a competitive equilibrium exists, in spite of the tipping behavior that can be expected under such circumstances.

The problem of externalities has been studied in the General Equilibrium literature. Arrow and Hahn (1971) show that in general a competitive equilibrium exists when externalities are present. In their analysis the utility functions and production sets are affected by the allocation of the economy (i.e. by all consumption vectors and production vectors of all agents). In equilibrium, each household maximizes its utility subject to what all other households and firms do, and each firm maximizes its profits subject to what all other firms and households do. The equilibrium allocation is therefore consistent in the sense that at that allocation everybody is maximizing conditional on everybody else's actions.

Another interesting question that arises is the efficiency properties of a competitive equilibrium with NE. This question is not answered in the current article, and remains an interesting topic for further research. However, our intu-

ition tells us that equilibrium in this case will not in general be Pareto efficient. We can think about this by an example. It has been documented that competition between network goods not always ends with the market adopting the best protocol as a standard (see for example David (1985)). There are many reasons why an inferior protocol may become the standard, and one of them is that protocols may be owned or sponsored by firms that will make an effort for imposing them in the market. Usual examples in the Industrial Organization literature are the competition in the home video market between VHS and Betamax and the competition between the Dvorak and QWERTY keyboards. When the market for a given good is dominated by an inferior protocol equilibrium is clearly not Pareto efficient, since it would be advantageous for everyone in the network to switch to the superior protocol (assuming of course that everybody in the network regards that protocol as superior). But each individual has no incentive to switch unilaterally since that way he would loose the interaction benefits.

Another reason for equilibrium with NE not to be PE is the usual that in the presence of positive (negative) externalities individuals will consume less (more) of the externality than the optimal quantity, since when making decisions they don't take into account the effects of their consumption on other individuals. This of course is also true in the case of NE, in which consumers do not consider the effect that their consumption has in the value of the network for other consumers. These arguments suggest that if we can show that an equilibrium exists with NE, we can't expect it to be Pareto efficient.

Before going to the model we discuss briefly the links between the theory of NE and the theory of clubs. Work has been done in existence of equilibrium in club economies (see the references below), and the current research can also be

interpreted as a complement to that work.

III.A.2 Links with the theory of clubs

The description of NE above suggests a very close relation with what has been called "Clubs" in the literature. The theory of clubs, starting with Buchanan's (1965) seminal article, addresses the question of the efficient provision of certain "imperfect" public goods. The word imperfect comes from the fact that in general exclusion is possible for these goods, and also from the fact that consumption is partially rival. This is so because the consumption of the public good by one more individual affects the level of satisfaction that other individuals get from the good. This in general is due to congestion, and the literature on clubs ask about the optimal quantity of the public good and the optimal membership of the club sharing a public good of that size. Other references for this literature are Berglas (1976) and Ng (1973 and 1974). A comprehensive survey is presented in Sandler and Tschirhart (1980).

In our case we can think of a NE as a case of "negative congestion", i.e. one in which existing consumers benefit from the addition of new consumers to the club. The analogy is not complete though, since we can say that a new member of the network will enlarge the size of the public good, i.e. that the quantity of the public good is not being held fixed. However, the club literature can give great insight as to how to think about NE.

There is also a distinctive feature of clubs that will prove to be very useful in analyzing NE, which is the indivisibility of club membership. As pointed out by most of the clubs literature, club membership is indivisible, in the sense that the individual either belongs to a club or not. As we argued above, this indivisibility

is a distinctive feature of most of the goods in which NE are present and it seems to be the key feature of these markets for exhibiting "tipping".

In recent years, there has been some interest in exploring existence of equilibrium in club economies. Some of the papers on this topic are Cole and Prescott (1997), Ellickson, Grodal, Scotchmer and Zame (1999) and Scotchmer (1997). A common feature of these papers is that they consider clubs as predetermined consumption units that are small compared with the population (i.e. all possible club types are a characteristic of the economy, what is endogenous is how many of the given clubs of each type will form in equilibrium). This smallness helps them to solve the "integer problem", which arises when the population is not an integer multiple of club size. Using this they show that in a properly defined economy with clubs a competitive equilibrium exists.

In the case of NE, the assumption of small clubs is not very appealing. This is so because typically in the case of NE, in view of the observations in the literature about tipping, we would expect that a significant portion of the population joins a given "club". Then the existence theorems for club economies with small clubs do not seem applicable to NE. Instead, it would make more sense to look for an existence theorem in which the size of the clubs can be large compared with the population.

Another desirable feature we can ask to our model is that the sizes of the clubs are not predetermined, but rather that in equilibrium any size can obtain for a given club or network. For these reasons we can interpret the current paper as an extension to the results on club economies to the case in which club sizes are endogenous and in equilibrium can be of the same order of magnitude than the size of the population.

In the rest of the paper we will explore the possibility of existence of a competitive equilibrium under NE. In section III.B we present the model and the main result of the paper, namely the existence theorem. In section III.C a preliminary lemma is stated, which will be helpful in the proof of the existence theorem. In section III.D we provide the proof. Section III.E contains some remarks and concludes.

III.B The model and main result

The General Equilibrium literature has dealt with the problem of indivisibilities in consumption. Two notable examples are Mas-Colell (1977) and Yamazaki (1978). In the other hand, existence of equilibrium with externalities has been shown for example in Arrow and Hahn (1971). As we argued above, when we have NE we have to deal with both of this problems, indivisibilities *and* externalities, and that in that case it is not obvious that the previous results on existence of equilibrium will obtain, due to the tipping behavior that is expected in this situation. In this section we present a model that incorporates both of these features, indivisibilities and externalities. The model is more in the spirit of Yamazaki's, from which some basic results have been taken. However, in our case the problem is more complex and therefore the strategy of proof is not exactly the same.

III.B.1 Private Goods

There are N private goods. These goods are perfectly divisible. There are no externalities in these goods. They are not produced, they come only from endowment.

III.B.2 Network Goods

There are M different "activities" that consumers can perform in networks. Access to the networks can be obtained by buying a certain network good (then there are M network goods).Each household satisfaction from joining a given network depends on the *number* of households in the network. Then all households are identical from the point of view of others in their effect on welfare of households in the networks they join.

Jumping a little ahead, there is a measure space of consumers (households) $(A,\mathcal{A},\mu)$, where $A = [0,1]$, $\mathcal{A}$ is the Borel $\sigma-$ algebra on $[0,1]$ and μ is Lebesgue measure on measurable subsets of $[0,1]$. Then we define a network as a pair (B,j), where B is the (measurable) set of consumers joining the network and $j \in \{1,2,...,M\}$ indicates which activity the network performs. The assumption above about consumers only concerned with the number of other consumers in each network translates formally into $\mu(B)$ entering the individual utility functions. We also assume that only one network will form for each activity. Then it is not possible to have a club composed by set B and another club composed by set C performing the same activity. Once a consumer buys a network good, he will interact with everybody who bought the same network good.

Network goods are indivisible. Then each household will chose either to belong to a given network or not. Therefore the consumption set for a consumer is

$$X = \mathbf{R}_+^N \times \{0,1\}^M$$

where $\{0,1\}$ indicates the quantity of each network good purchased.

III.B.3 Consumers

There is a (nonatomic) measure space of consumers $(A, \mathcal{A}, \mu)$, $A = [0, 1]$, $\mu(A) = 1$. (μ is Lebesgue measure and $\mathcal{A}$ is the Borel $\sigma-$ algebra of measurable subsets of $[0, 1]$).

Each consumer is endowed with $e(a) \in \mathbf{R}_+^{N+M}$ (but only quantities of the N private goods, the first N components in $e(a)$, can be positive in the endowment vector). We assume that mean endowment, $\int_A e(a)$, is finite.

We define a *consumption allocation* and a *total consumption allocation* as:

Definition III.1 *A consumption allocation* $\mathbf{c}$ *is an integrable function* $\mathbf{c} : A \to \mathbf{R}^{N+M}$ *such that for almost every* $a \in A$, $\mathbf{c}(a) \in X$. *We say that* λ *is a total consumption allocation corresponding to the consumption allocation* $\mathbf{c}$ *if* $\lambda = (\int_A \mathbf{c}_1, ..., \int_A \mathbf{c}_{N+M})$ *(*$\mathbf{c}_1$ *is the first component of the function* $\mathbf{c}$*, and so on).*

Notice that in the previous definition $\mathbf{c}(a)$ is a (vector valued) function specifying in the k^{th}coordinate the consumption of good k for individual $a \in A$, whereas λ is the ($N + M$ coordinate) vector that results from integrating the consumption allocation $\mathbf{c}$ over the set of consumers A. Notice that the last M coordinates of the vector λ represent the proportion of the population participating in each of the M networks (since the total measure of the population is equal to 1).

Each consumer satisfaction depends on his own consumption of private goods and network goods, plus on the measure of the set of consumers joining the networks he joins. Consumer $a's$ preferences are represented by the continuous

utility function:

$$u(a, x, \lambda) \qquad x \in X$$

In what follows, we will assume strict monotonicity in own consumption of all goods, i.e.

$$u(a, x, \lambda) > u(a, x', \lambda) \qquad if \qquad x \geq x', x \neq x'.$$

In showing existence of equilibrium a distributional assumption on individual incomes will prove very useful. The endowment distribution μ_e is the image measure of μ with respect to the mapping e, i.e. for every Borel set $B \in \mathbf{R}^{N+M}$

$$\mu_e(B) = \mu(\{a \in A : e(a) \in B\})$$

The following definition is taken from Yamazaki (1978):

Definition III.2 *Denote the price vector $p = (p_N, p_M)$, where p_N represents the first N components of p. The endowment distribution is said to be dispersed if the resulting wealth distribution from endowment (total income minus income received from firms profits), the measure*

$$\mu_{e,p}(D) = \mu(\{a \in A : pe(a) \in D\})$$

on $\mathcal{B}(\mathbf{R})$ is absolutely continuous with respect to Lebesgue measure on $\mathbf{R}$ for each $p \in \mathbf{R}_+^{N+M}, p_N \neq 0$.

In Yamazaki (1978), the endowment distribution is assumed to be dispersed. Intuitively, this dispersion assumption says that the measure $\mu_{e,p}$ does not give positive measure to any particular wealth value $k \in \mathbf{R}$. Technically, this is equivalent to say that the measure $\mu_{e,p}(D)$ has a density function. In Yamazaki's

work, which considers an exchange economy, this assumption is helpful in showing existence of equilibrium. The reason is because it prevents large sets of consumers from having the same income (for some prices) and therefore behaving similarly. Since individual behavior may not be continuous when the consumption set is not convex, existence of equilibrium relies heavily on the smoothing of aggregate behavior due to "large numbers". If consumers are "small" compared to the population, then we can have continuous aggregate demand behavior even when individual behavior is not. But if large sets of consumers exhibit "jumps" in demand at the same time, we may find non-continuous aggregate demand. In Yamazaki (1978) this assumption ensures that this is not the case.

In our case, income does not come only from endowment, since there are firms that are owned by individuals. Then, in our case the distributional assumption above will not be sufficient. In section III.B.7 we introduce an appropriate condition for our economy, that will ensure that income from all sources is dispersed.

III.B.4 Firms

In our economy private goods come only from endowment (they are not produced), and are used for consumption and as inputs in the production of the network goods. There is a finite set F of firms, with M firms in it, each producing one network good using private goods as inputs. We assume that these firms behave competitively, despite of the fact that there is only one firm producing each network good.

Firm $f \in F$ is endowed with a production set Y^f. The following assumptions on Y^fwill be maintained throughout the paper:

(1) $0 \in Y^f$.

(2) Y^f is closed.

(3) Y^f is convex.

In Y^f private goods are inputs (nonpositive) and network goods are outputs (nonnegative).

III.B.5 Prices

The price space will be PS, the $N + M$-dimensional unit simplex.

III.B.6 Supply Behavior

Firms maximize profits (py, for $y \in Y^f$) subject to production sets, taking prices as given. The supply set of firm f at prices p is (there are no externalities in production):

$$S(f, p) = \{y \in Y^f : py \geq pz \text{ for all } z \in Y^f\}.$$

III.B.7 Income and Demand Behavior

Household income comes from 2 sources: Endowment and shares on firms profits. Income for household a at prices p is defined as:

$$M(a, p) = pe(a) + \sum_{f \in F} \alpha(a, f)\Pi(f, p)$$

where $\int_B \alpha(a, f)$ for $B \in \mathcal{A}$ is set $B's$ share on firm f profits ($\alpha(., f)$ is a measurable function), and $\int_A \alpha(a, f) = 1$, and $\Pi(f, .)$ is the profit function for firm f (i.e. the function that specifies the maximum profit for firm f for each price vector).

As mentioned above, we will need a distributional assumption on individual incomes. The following example shows why the assumption on the distribution

of endowment is not sufficient to ensure dispersion in the distribution of income in our case:

Example III.1 *Consider an economy with one private good and one network good. Price of the private good is equal to 1.* $A = [0, 1]$ *and* $e(a) = a$ *for* $a \in A$*. Profits for the only firm are equal to* $\frac{1}{2}$ *and shares on firms profits are* $\alpha(a, f) = 2 - 2a$ *for* $a \in A$*. Then the endowment distribution is dispersed, but income is equal to 1 for all* $a \in A$*.*

As seen in the example, we need something in addition to dispersion of endowments to obtain dispersion of incomes. As it is clear from the example, the problem arises because the distribution of ownership shares is correlated with the distribution of endowment.

Throughout the paper we will impose the following assumption on the income distribution:

Condition III.1 *The income distribution for* $p \in PS$ *is the image measure of* μ *with respect to the mapping* M *, i.e.,*

$$\mu_{M,p}(B) = \mu(\{a \in A : M(a, p) \in B\})$$

for every Borel set $B \in \mathbf{R}$*. We assume that the income distribution is dispersed, i.e. the measure* $\mu_{M,p}$ *is absolutely continuous with respect to Lebesgue measure on* $\mathbf{R}$ *for each* $p \in PS$*,* $p \neq 0$*.*

Notice that we may have large portion of the population (possibly all of it) in a small income interval and still satisfy the condition. All that is required is that there is no large group of consumers with *exactly* the same income for any price vector in PS.

Consumers maximize $u(a, x, \lambda)$ for given λ subject to consuming in their budget set. The budget set for consumer $a \in A$ at prices $p \in PS$ is:

$$B(a, p) = \{x \in X : px \leq M(a, p)\}$$

The demand set for agent a at prices p and total consumption allocation λ is:

$$D(a, p, \lambda) = \{x \in B(a, p) : u(a, x, \lambda) \geq u(a, z, \lambda) \quad for \quad z \in B(a, p)\}$$

III.B.8 Equilibrium

We already defined a consumption allocation. A production allocation is a function $s : F \to \mathbf{R}^{N+M}$ such that for all $f \in F$, $s(f) \in Y^f$. An allocation is a pair $(\mathbf{c}, s)$ where $\mathbf{c}$ is a consumption allocation and s is a production allocation.

Next we define competitive equilibrium in our economy:

Definition III.3 *A competitive equilibrium in our economy is a price vector $p^* \in PS$, and an allocation (d^*, s^*) such that*

(1) $d^*(a) \in D(a, p^*, \lambda^*)$ for a.e. $a \in A$.

(2) $s^*(f) \in S(f, p^*)$ for all $f \in F$.

(3) $\int_A d^* \leq \sum_F s^* + \int_A e$

(4) $\int_A d^* = \lambda^*$,i.e. λ^* is consistent.

The last condition for an equilibrium is important. It says that at equilibrium, agents must be optimizing *given* the equilibrium consumption choices of other agents. The main result of the paper is stated in the following theorem, which is proven in section III.D.

Theorem III.1 *In the economy specified above there exists a competitive equilibrium.*

III.C Preliminary results

In this section we present without proof a result that will be helpful in proving theorem III.1. For a proof of this result, the reader is directed to Yamazaki (1978). The proof given there does not rely on any property of the consumption set and therefore applies here as well.

A known problem for existence of equilibrium, known in the literature as the "Arrow corner" arises when for some price vector the income of some individuals or households is equal to zero. The traditional solution is to bound income away from zero, avoiding discontinuous behavior that can happen in that case. When the economy is "large", in the sense that there are uncountably many consumers, our previous assumption on dispersion of income will in turn ensure that the set of households that could exhibit the Arrow corner for any price vector is a set of measure zero, and therefore does not matter for aggregate demand. The following result tells us that this is indeed the case.

We say that a point $x \in X$ has *local cheaper points* at prices $p \neq 0$ if for any neighborhood $U(x)$ there exist $z \in U(x) \cap X$ such that $pz < px$.Then for $p \neq 0$ define

$C^p = \{x \in X : x \text{ does not have local cheaper points}\}$

$A^p = \{a \in A : \text{there is a vector } x \in C^p \text{ such that } px = M(a,p)\}$

Then the set A^p is the set of households whose "budget line" contains points in C^p.

The following lemma, taken from Yamazaki's Corollary 1 (1978, pp. 549), asserts that in proving existence of equilibrium we don't have to worry for the behavior of households in the set A^p.

Lemma III.1 *If the distribution of income is dispersed, then A^p is a set of measure zero.*

Then we see that the set of households A^p will not have a positive effect on aggregate behavior, and therefore can be ignored in proving existence of equilibrium.

III.D Proof of the existence theorem

In this section we will present the proof for the main result of the paper, namely that in the model presented in section III.B there exist a competitive equilibrium in the sense defined in the same section.

Before going to the proof, we will introduce some notation and definitions.

We already defined a consumption allocation as a function $\mathbf{c} : A \to \mathbf{R}^{N+M}$. Call C the set of all possible consumption allocations (then C is the set of all measurable functions $\mathbf{c} : A \to \mathbf{R}^{N+M}$ such that $\mathbf{c}(a) \in X$ for a.e. $a \in A$). The set of *total* consumption allocations will be denoted Ω.Therefore

$$\Omega = \mathbf{R}^N_+ \times [0,1]^M$$

A total consumption allocation $\lambda \in \Omega$ if there is a consumption allocation $\mathbf{c} \in C$ such that $\lambda = (\int_A \mathbf{c}_1, ..., \int_A \mathbf{c}_{N+M})$. Note that in any $\lambda \in \Omega$ the first N coordinates indicate aggregate demand for the N private goods, whereas the last M coordinates indicate the proportion of the population that participates in each of the networks.

For the sum of the sets $A+B$ we denote the set H such that $x \in H$ if there is $y \in A, z \in B$, such that $x = y + z$. We then define the set of total production allocations as the set $Y = \sum_F Y^f$ (i.e. it is the social production possibility set).

One extra assumption on Y will be helpful in proving the main result of the paper:

(A.1) $Y \cap \mathbf{R}_+^{N+M} = \{0\}$ (No free lunch)

Also, since the first N components of $y^f \in Y^f$ are nonpositive and the last M nonnegative, we have the following consequence:

(A.2) $Y \cap -Y = \{0\}$ (Irreversibility)

This is so since a good that is an input for a given firm cannot be produced by any other firm.

Finally define the set Δ as

$$\Delta = \Omega - Y - \{\int_A e\}.$$

The strategy of proof will be the following: First we construct the correspondences:

(1) Total Demand Correspondence: $\Lambda : PS \times \Omega \to \Omega$

$$\Lambda(p, \lambda) = \int_A D(a, p, \lambda)$$

(2) Excess Demand Correspondence: $Z : PS \times \Omega \to \Delta$

$$Z(p, \lambda) = \int_A D(a, p, \lambda) - \int_A e - \sum_F S(f, p) = \Lambda(p, \lambda) - \int_A e - \sum_F S(f, p)$$

(3) Price Adjustment Correspondence: $\rho : \Delta \to PS$

$$\rho(z) = \{p^* \in PS : p^* = \arg\max_{p \in PS} pz\}$$

Then we have the correspondence $\rho \times \Lambda \times Z : PS \times \Omega \times \Delta \to PS \times \Omega \times \Delta$. We then find conditions under which a fixed point theorem can be applied to this correspondence, and after that we show that the fixed point is a competitive equilibrium.

In fact, we don't carry over this procedure directly. Instead, we will find an equilibrium for a sequence of suitable bounded economies indexed by $k = 1, 2, ...$ We then find a limit point for the sequence of competitive equilibria of the bounded economies, and show that this limit point is itself an equilibrium for the unbounded economy.

Before starting to construct the correspondences, we define

$$b = \int_A e_1 + \int_A e_2 + ... + \int_A e_N$$

This quantity will be used in truncating the consumption and production sets later. If the distribution of income is dispersed, then $b > 0$. This is so since if $b = 0$, this implies that $e_i(a) = 0$ for a.e $a \in A$, $i \in \{1, 2, ..., N\}$, and since no production is possible, profits are zero for all firms too. But this contradicts the distribution of income being dispersed, since then income is equal to zero for a.e. $a \in A$.

III.D.1 Supply Correspondence

The supply correspondence for firm $f \in F$ is:

$$S(f, p) = \{y \in Y^f : py \geq pz, z \in Y^f\}$$

This correspondence may not be well defined, since the set Y^f is in general unbounded for $f \in F$. Then we truncate the production set for firm f. For $k = 1, 2, ...,$ define the k^{th} truncated production set for firm f as

$$Y^{f,k} = \{y \in Y^f : -kb \leq y_i \leq 0 \text{ for } i = 1, 2, ..., N,\ 0 \leq y_i \leq k \text{ for } i = N{+}1, ..., N{+}M\}$$

If we call

$$I^k = [-kb, 0]^N \times [0, k]^M \qquad\qquad k = 1, 2, ...$$

then the k^{th} truncated production set is $Y^{f,k} = Y^f \cap I^k$.

It is easily seen that the set $Y^{f,k}$, for $k = 1, 2, ...,\ f \in F$, is: (i) Convex, since Y^f and I^k are convex, (ii) Compact, since Y^f is closed and I^k is compact and (iii) $0 \in Y^{f,k}$.

Next we define the supply correspondence in the k^{th} truncated economy as

$$S^k(f,p) = \{y \in Y^{f,k} : py \geq pz, z \in Y^{f,k}\}$$

the aggregate supply correspondence in the k^{th} truncated economy as

$$\sum_{f \in F} S^k(f,p)$$

and the profit function for firm f in the k^{th} bounded economy as

$$\Pi^k(f,p) = py \qquad\qquad y \in S^k(f,p)$$

Lemma III.2 *The aggregate supply correspondence in the k^{th} truncated economy, $\sum_{f \in F} S^k(f,p)$, is nonempty, upper-hemicontinuous and convex valued. Furthermore, the profit function $\Pi^k(f,p)$ is a continuous function of p, for all $p \in PS$.*

Proof: First, $S^k(f,p)$ is nonempty, since it is the set of maximizers of a continuous function on the compact set $Y^{f,k}$. Next, $S^k(f,p)$ is also convex valued. In fact, suppose that $y_1, y_2 \in Y^{f,k}$, $y_1, y_2 \in S^k(f,p)$. Then we have $py_1 = py_2$. As $Y^{f,k}$ is convex, then $\alpha y_1 + (1-\alpha)y_2 \in Y^{f,k}$, for $\alpha \in [0,1]$. Then $p[\alpha y_1 + (1-\alpha)y_2] = \alpha p y_1 + (1-\alpha)py_2 = py_1$and therefore $\alpha y_1 + (1-\alpha)y_2 \in S^k(f,p)$.

Now we show that $S^k(f,p)$ is upper-hemicontinuous. Take a sequence $p^n \to p$, $y^n \to y$, $y^n \in S^k(f,p^n)$. We want to show that $y \in S^k(f,p)$. Suppose this is not the case. Then there exist $y' \in Y^{f,k}$, such that $py' > py$. We also have

$$p^n y' \to py'$$

$$p^n y^n \to py$$

then for some $N > 0$, $n \geq N$ implies

$$p^n y' > p^n y^n$$

which is a contradiction to the fact that $y^n \in S^k(f, p^n)$. The contradiction proves the upper-hemicontinuity.

It is a known result that the finite sum of upper-hemicontinuous correspondences is itself upper-hemicontinuous. Therefore we have that $\sum_{f \in F} S^k(f, p)$ is upper-hemicontinuous.

Finally we show that $\Pi^k(f, p)$ is continuous in p. Take a sequence $p^n \to p$, $y^n \in S^k(f, p^n)$. Without loss of generality take a convergent subsequence $y^n \to y$ (remember Y^k is bounded). By upper-hemicontinuity of $S^k(f, p)$ we have that $y \in S^k(f, p)$. Then we have

$$p^n y^n \to py$$

which is

$$\Pi^k(f, p^n) \to \Pi^k(f, p)$$

Since this is true for any convergent subsequence y^n continuity is proven. Q.E.D.

Having shown this, we define household income and show it is continuous in prices. Household income in the k^{th} economy is defined as

$$M^k(a, p) = pe(a) + \sum_F \alpha(a, f)\Pi^k(f, p)$$

We require that $\int_A \alpha(a, f) = 1$ for all $f \in F$. We already showed that $\Pi^k(f, p)$ is continuous. It follows trivially then that $M^k(a, p)$ is well defined and continuous in p.

III.D.2 Total Demand Correspondence

Household demand at prices p and total consumption allocation λ was defined above as the set $D(a, p, \lambda)$. This correspondence may not be well defined, since for some prices the set $B(a, p)$ may not be bounded. In order to overcome this problem and obtain an individual demand correspondence that is well defined we start by truncating the consumption set X. The k^{th} truncated consumption set is

$$X^k = \{x \in X : x \leq (kb, ..., kb, 1, ..., 1)\}$$

and the truncated budget set is defined as

$$B^k(a, p) = \{x \in X^k : px \leq M^k(a, p)\}$$

We also define the subset of consumers

$$A^k = \{a \in A : e(a) \leq (kb, ..., kb, 0, ..., 0)\}$$

(i.e. households whose endowments are in the truncated consumption set).

Now we will argue that A^k are of positive measure for $k = 1, 2, ...$ In fact $A^k \subset A^{k+1}, k = 1, 2, ...,$ so we only have to show that A^1 is of positive measure. Suppose not. Then this means that $e(a) > b$ for a.e. $a \in A$. Therefore we have for some $i = 1, 2, ..., N$

$$\int_A e_i > b = \int_A e_1 + \int_A e_2 + ... + \int_A e_N \geq \int_A e_i$$

which is a contradiction. We then define the induced measure spaces $(A^k, \mathcal{A}^k, \mu^k)$, where $\mathcal{A}^k$ is the Borel $\sigma-$ algebra of measurable subsets B^k of A^k such that there is a measurable set $B \subset A$ and $B^k = B \cap A^k$. The measure μ^k is the restriction of μ to sets in $\mathcal{A}^k$.

In section III.B we defined the set of consumption allocations C. The set of truncated consumption allocations C^k, $k = 1, 2, ...$, is the subset of C such that $\mathbf{c}(a) \in X^k$ for a.e. $a \in A^k$. Then the set of truncated total consumption allocations Ω^k is the subset of Ω such that $\lambda \in \Omega^k$ if there is $\mathbf{c} \in C^k$ such that $\lambda = \int_{A^k} \mathbf{c}$.

The set Ω^k so defined will then be the interval

$$\Omega^k = [0, \mu(A^k)kb]^N \times [0, \mu(A^k)]^M$$

This set is trivially compact and convex.

For $\lambda \in \Omega^k, p \in PS$, the individual truncated demand correspondence is

$$D^k(a, p, \lambda) = \{x \in B^k(a, p) : u(a, x, \lambda) \geq u(a, z, \lambda), z \in B^k(a, p)\}$$

and aggregate truncated demand correspondence is

$$\Lambda^k(p, \lambda) = \int_{A^k} D^k(a, p, \lambda)$$

Lemma III.3 *The aggregate demand correspondence Λ^k is: (i) Well defined and convex valued for $(p, \lambda) \in PS \times \Omega^k$ and (ii) Upper-hemicontinuous in (p, λ).*

Proof: (i) For $(p, \lambda) \in PS \times \Omega^k$, we show that truncated individual demand is well defined for a.e. $a \in A^k$. $B^k(a, p)$ is nonempty for a.e. $a \in A^k$. It is also compact, since it is obviously closed and it is a subset of X^k which is bounded. Then truncated individual demand correspondence $D^k(a, p, \lambda)$ is the set of maximizers of the continuous function $u(a, x, \lambda)$ on a compact set and therefore is well defined for a.e. $a \in A^k$.

The argument in Hildenbrand (1970), theorem 3 in page 614, which applies equally here, shows that $D^k(., p, \lambda)$ is measurable for $(p, \lambda) \in PS \times \Omega^k$. Therefore

$$\Lambda^k(p, \lambda) = \int_{A^k} D^k(a, p, \lambda)$$

is well defined and nonempty for $(p, \lambda) \in PS \times \Omega^k$.

It is well known that the integral of a correspondence with respect to an atomless measure is convex valued (see for example Hildenbrand (1970), page 615). Then $\Lambda^k(p, \lambda)$ is convex for $(p, \lambda) \in PS \times \Omega^k$.

(ii) Now we show that $\Lambda^k(p, \lambda)$ is upper-hemicontinuous. Analogously as in section III.B, we define the sets

$C^{p,k} = \{x \in X^k : x \text{ does not have local cheaper points}\}$

$A^{p,k} = \{a \in A^k : \text{there is a vector } x \in C^{p,k} \text{ such that } px = M^k(a, p)\}$

It was stated in section III.C that A^p is a set of measure zero. But $A^{p,k} \subset A^p$. So $A^{p,k}$ is also a set of measure zero.

Take the sequence $(p^n, \lambda^n) \to (p, \lambda), \gamma^n \to \gamma, \gamma^n \in \Lambda^k(p^n, \lambda^n)$. We want to show that $\gamma \in \Lambda^k(p, \lambda)$. For each n there exist an integrable function $d^n : A^k \to \mathbf{R}^{N+M}$ such that $d^n(a) \in D^k(a, p^n, \lambda^n)$ a.e. in A^k and $\int_{A^k} d^n = \gamma^n$.

Let $L(a)$ be the set of cluster points of $d^n(a)$. Notice that $L(a)$ is nonempty since the sequence $d^n(a)$ is bounded. We want to show that a.e. in $A^k/A^{p,k}$ $L(a) \subset D^k(a, p, \lambda)$. This in turn will imply that $\int_{A^k} L(a) \subset \int_{A^k} D^k(a, p, \lambda)$.Then by showing that $\gamma \in \int_{A^k} L(a)$ we'll be done.

To show $L(a) \subset D^k(a, p, \lambda)$ a.e. in $A^k/A^{p,k}$, take $a \in A^k/A^{p,k}$ such that $d^n(a) \in D^k(a, p^n, \lambda^n)$ and let $d^n(a) \to g(a)$ (i.e. $g(a) \in L(a)$). Then $g(a) \leq (kb, ..., kb, 1, ..., 1)$ since $d^n(a) \leq (kb, ..., kb, 1, ..., 1)$ for all n. Now $p^n d^n(a) \leq M^k(a, p^n)$ and $p^n \to p$ implies $pg(a) \leq M^k(a, p)$. Now take $y \in X^k$ such that $py < M^k(a, p)$. Then $p^n y < M^k(a, p^n)$ for n sufficiently large, and therefore

$$u(a, d^n(a), \lambda^n) \geq u(a, y, \lambda^n)$$

for n sufficiently large. Then by continuity of u in both arguments we have

$$u(a, g(a), \lambda) \geq u(a, y, \lambda)$$

Now take $y \in X^k$ such that $py = M^k(a,p)$. Since $a \in A^k/A^{p,k}$, then there exist a sequence $y^i \to y$ such that $py^i < M^k(a,p)$ for all i. Then for each i we have

$$u(a, g(a), \lambda) \geq u(a, y^i, \lambda)$$

and therefore

$$u(a, g(a), \lambda) \geq u(a, y, \lambda)$$

again using continuity of u. Then $g(a) \in D^k(a,p,\lambda)$. Therefore we have that

$$\int_{A^k} L(a) \subset \int_{A^k} D^k(a,p,\lambda)$$

as was to be shown. Finally we show that $\gamma \in \int_{A^k} L(a)$. Since d^n are bounded (i.e. $d^n(a) \in X^k$ for all n), d^n is bounded by the constant function $h(a) = (kb, ..., kb, 1, ...1)$. Then we can use the result in Aumann (1976) that states:

Let d^n be a sequence of measurable functions from A^k into $\mathbf{R}^{N+M}$, which are bounded by the integrable function h. Then each cluster point of $\int_{A^k} d^n \ (= \gamma^n)$ belongs to $\int_{A^k} L$. (i.e. γ, which is a cluster point of the sequence $\int_{A^k} d^n$, is in $\int_{A^k} L$).

Then we have shown that $\gamma \in \Lambda^k(p,\lambda)$. Consequently we have shown that $\Lambda^k(p,\lambda)$ is upper-hemicontinuous and the proof is complete. Q.E.D.

III.D.3 Excess Demand Correspondence

At the beginning of this section we defined the excess demand correspondence as $Z : PS \times \Omega \to \Delta$

$$Z(p,\lambda) = \int_A D(a,p,\lambda) - \int_A e - \sum_F S(f,p) = \Lambda(p,\lambda) - \int_A e - \sum_F S(f,p)$$

The set Δ was defined as

$$\Delta = \Omega - Y - \{\int_A e\}.$$

Similarly, we define the set Δ^k as

$$\Delta^k = \Omega^k - Y^k - \{\int_{A^k} e\}$$

The set Δ^k is certainly bounded, and it is closed since Ω^k and Y^k are compact (proof omitted here). Therefore it is compact. Δ^k is also convex since Ω^k and Y^k are convex.

The truncated excess demand correspondence $Z^k : PS \times \Omega^k \to \Delta^k$ is defined as

$$Z^k(p,\lambda) = \int_{A^k} D^k(a,p,\lambda) - \int_{A^k} e - \sum_F S^k(f,p) = \Lambda^k(p,\lambda) - \int_{A^k} e - \sum_F S^k(f,p)$$

Lemma III.4 *The truncated excess demand correspondence $Z^k(p,\lambda)$ is upper-hemicontinuous and convex valued.*

Proof: It has already been shown that $\Lambda^k(p,\lambda)$ and $\sum_{f\in F} S^k(f,p)$ are upper-hemicontinuous and convex valued for $(p,\lambda) \in PS \times \Omega^k$, and since $\int_{A^k} e$ is a real valued vector, then $Z^k(p,\lambda)$ is a sum of upper-hemicontinuous and convex valued correspondences. Therefore it is upper-hemicontinuous and convex valued. Q.E.D.

III.D.4 Price Adjustment Correspondence

At the beginning of this section the price adjustment correspondence $\rho : \Delta \to PS$ was defined as

$$\rho(z) = \{p^* \in PS : p^* = \arg\max_{p\in PS} pz\} \qquad z \in \Delta$$

Similarly we define the truncated price adjustment correspondence $\rho^k : \Delta^k \to PS$ as

$$\rho^k(z) = \{p^* \in PS : p^* = \arg\max_{p\in PS} pz\} \qquad z \in \Delta^k$$

Lemma III.5 *The truncated price adjustment correspondence $\rho^k(z)$ is nonempty, upper-hemicontinuous and convex valued.*

Proof: We have to show that ρ^k is upper-hemicontinuous and convex valued (nonempty is trivial). To show convexity, take $z \in \Delta^k$ such that $p_1, p_2 \in \rho^k(z)$. Then $p_1 z = p_2 z$. Therefore $\sigma p_1 + (1-\sigma)p_2 \in \rho^k(z)$, and $\sigma p_1 + (1-\sigma)p_2 \in PS$. This means $\rho^k(z)$ is convex.

To show upper-hemicontinuity, take a sequence $z^n \to z, p^n \to p$, such that $p^n \in \rho^k(z^n)$. Then we have that for all n

$$p^n z^n \geq p^n y \qquad y \in \Delta^k$$

Taking limits we have

$$pz \geq py \qquad\qquad y \in \Delta^k$$

which implies that $p \in \rho^k(z)$. Then $\rho^k(z)$ is upper-hemicontinuous. Q.E.D.

III.D.5 Existence of Equilibrium in the Truncated Economy

In subsection III.B.8 we introduced the definition of a competitive equilibrium in the unrestricted economy. Similarly, we define a competitive equilibrium in the k^{th} truncated economy as:

Definition III.4 *A competitive equilibrium in the k^{th} truncated economy is a price vector $p^{*k} \in PS$, and an allocation (d^{*k}, s^{*k}) such that*

(1) $d^{*k}(a) \in D(a, p^{*k}, \lambda^{*k})$ for a.e. $a \in A^k$.

(2) $s^{*k}(f) \in S(f, p^{*k})$ for all $f \in F$.

(3) $\int_{A^k} d^{*k} \leq \sum_F s^{*k} + \int_{A^k} e$

(4) $\int_{A^k} d^{*k} = \lambda^{*k}$,i.e. λ^{*k} is consistent.

Theorem III.2 *There exists a competitive equilibrium in the k^{th} truncated economy, $k = 1, 2, 3, ...$*

Proof: We have shown that the correspondence $\rho^k \times \Lambda^k \times Z^k : PS \times \Omega^k \times \Delta^k \to PS \times \Omega^k \times \Delta^k$ is upper-hemicontinuous and convex valued. The set $PS \times \Omega^k \times \Delta^k$ is compact and convex. We can therefore apply Kakutani's Fixed Point Theorem. Then for $k = 1, 2, ...$there exist a fixed point $(\lambda^{*k}, z^{*k}, p^{*k})$ such that $(\lambda^{*k}, z^{*k}, p^{*k}) \in \Lambda^k(p^{*k}, \lambda^{*k}) \times Z^k(p^{*k}, \lambda^{*k}) \times \rho^k(z^{*k})$. Now we will show that $(\lambda^{*k}, z^{*k}, p^{*k})$ is a competitive equilibrium in the k^{th} economy.

Since $z^{*k} \in Z^k(p^{*k}, \lambda^{*k})$ and

$$Z^k(p^{*k}, \lambda^{*k}) = \int_{A^k} D(a, p^{*k}, \lambda^{*k}) - \int_{A^k} e - \sum_F S^k(f, p^{*k})$$

then there exist integrable functions $d^{*k} : A^k \to \Omega^k$ such that $d^{*k}(a) \in D^k(a, p^{*k}, \lambda^{*k})$ for a.e. $a \in A^k$ and $s^{*k}(f) \in S^k(f, p^{*k})$ for $f \in F$ such that

$$z^{*k} = \int_{A^k} d^{*k} - \int_{A^k} e - \sum_F s^{*k}(f)$$

Now we have that for a.e. $a \in A^k$

$$p^{*k} d^{*k} \leq M^k(a, p^{*k}) = p^{*k} e(a) + \sum_F \alpha(a, f) p^{*k} s^{*k}(f)$$

Integrating over A^k we have

$$p^{*k} \int_{A^k} d^{*k} \leq p^{*k} \int_{A^k} e(a) + \sum_F p^{*k} s^{*k}(f) \int_{A^k} \alpha(a, f)$$

$$p^{*k} \int_{A^k} d^{*k} \leq p^{*k} \int_{A^k} e(a) + p^{*k} \sum_F s^{*k}(f)$$

since $\int_{A^k} \alpha(a, f) \leq 1$ and $p^{*k} s^{*k} \geq 0$. This is to say

$$p^{*k} \int_{A^k} d^{*k} - p^{*k} \int_{A^k} e(a) - p^{*k} \sum_F s^{*k}(f) \leq 0$$

or equivalently $p^{*k}z^{*k} \leq 0$.

Since p^{*k} maximizes pz over PS, then this implies that $z^{*k} \leq 0$. This is so because if $z^{*k} > 0$ for some $i \in \{1, 2, ..., N + M\}$, then we would have $p^{*k}z^{*k} > 0$.

Finally, we have to show that total demand is consistent, in the sense that in equilibrium individuals are maximizing *given* the equilibrium choices of other agents. But this is quite trivial, since by $(\lambda^{*k}, p^{*k}, z^{*k})$ being a fixed point of the correspondence $\Lambda^k \times Z^k \times \rho^k$ we automatically have that $\lambda^{*k} \in \Lambda^k(p^{*k}, \lambda^{*k})$, which is the consistency condition of equilibrium. Therefore the price vector p^{*k} and the allocation (d^{*k}, s^{*k}) represent a competitive equilibrium in the k^{th} truncated economy for $k = 1, 2, ...$ Q.E.D.

III.D.6 Existence of Equilibrium in the Unrestricted Economy

In this section we finish the proof of the main theorem of the paper, i.e. we show that there exist an equilibrium in the unrestricted economy. The strategy of proof will be the following: We take the sequence of equilibria p^{*k} , (d^{*k}, s^{*k}) for $k = 1, 2, ...$, and find a cluster point p^*, (d^*, s^*) for this sequence. Then we show that this cluster point is itself an equilibrium in the unrestricted economy.

For $k = 1, 2, ...$, define the following function:

$$\begin{aligned} g^k(a) &= d^{*k}(a) \text{ if } a \in A^k \\ &= e(a) \quad \text{if } a \notin A^k \end{aligned}$$

Then we have that for $k = 1, 2, ...$

$$\int_{A^k} d^{*k} - \int_{A^k} e(a) - \sum_F s^{*k}(f) \leq 0$$

which implies

$$\int_A g^k(a) - \int_A e(a) - \sum_F s^{*k}(f) \leq 0$$

where $g^k(a) \in D^k(a, p^{*k}, \lambda^{*k})$ a.e. in A^k, $s^{*k}(f) \in S(f, p^{*k})$ for all f.

Without loss of generality assume $p^{*k} \to p$ as $k \to \infty$. Then we claim:

Claim III.1 *There exist functions $g : A \to \mathbf{R}^{N+M}$ (integrable), and $s : F \to \mathbf{R}^{N+M}$, such that $(g(a), s)$ is a cluster point of $(g^k(a), s^{*k})$ for a.e. $a \in A$ and $f \in F$. Moreover, the sequence $\int_A g^k \to \int_A g$ as $k \to \infty$.*

Then we have

$$\int_A g(a) - \int_A e(a) - \sum_F s(f) \leq 0$$

which is condition (3) in the definition of an equilibrium. Call $\lambda = \int_A g$. Then $\lambda^{*k} \to \lambda$ as $k \to \infty$. The proof of the following lemma completes the argument:

Lemma III.6 *(1) $g(a) \in D(a, p, \lambda)$ for a.e. $a \in A$; (2) $s(f) \in S(f, p)$ for $f \in F$; (3) $\lambda \in \int_A D(a, p, \lambda)$.*

Proof: To prove (1), take $a \in A$. Then there is $k^*(a)$ such that for $k > k^*(a)$, $a \in A^k$. Then for $k > k^*(a)$ we have $g^k(a) \in D^k(a, p^{*k}, \lambda^{*k})$. For $k = 1, 2, ...$ we have

$$p^{*k} g^k(a) \leq p^{*k} e(a) + \sum_F \alpha(a, f) p^{*k} s^{*k}$$

then taking limits as $k \to \infty$

$$pg(a) \leq pe(a) + \sum_F \alpha(a, f) ps$$

and therefore $g(a) \in B(a, p)$ a.e in A.

Now take $x \in B(a,p)$. We can have two cases: (i) $px < M(a,p)$, and (ii) $px = M(a,p)$. In case (i) we have that for k sufficiently large

$$p^{*k}x < M^k(a,p^{*k})$$

(since $M^k(a,p^{*k}) \to M(a,p)$ as $k \to \infty$). Then

$$u(a,g^k(a),\lambda^{*k}) \geq u(a,x,\lambda^{*k})$$

Therefore by continuity of u we have that

$$u(a,g(a),\lambda) \geq u(a,x,\lambda)$$

(where $\int_A g^k = \lambda^{*k}, \int_A g = \lambda$, and we have used the result in the claim that $\lambda^{*k} \to \lambda$) what is to say $g(a) \in D(a,p,\lambda)$. In case (ii), take $a \in A/A^p$. Then we can find a sequence $x^i \to x$ such that $px^i < M(a,p)$. This implies that for k sufficiently large

$$p^k x^i < M^k(a,p^k)$$

which in turn implies

$$u(a,g^k(a),\lambda^{*k}) \geq u(a,x^i,\lambda^{*k})$$

Taking limits first with respect to k and then with respect to i continuity of u gives

$$u(a,g(a),\lambda) \geq u(a,x,\lambda)$$

and then $g(a) \in D(a,p,\lambda)$. Therefore the proof of (1) is complete, since we have shown before that A_p is a set of measure zero when the distribution of income is dispersed.

Now we show (2). We want to show that $ps(f) \geq py$ for $y \in Y^f$.Take $y \in Y^f$. For k sufficiently large we have that $y \in Y^{f,k}$. For $k = 1, 2, ...,$ we have $s^k(f) \in S^k(f, p^k)$ and then for k such that $y \in Y^f$ we have $p^k s^k(f) \geq p^k y$. Taking limits as $k \to \infty$ this gives

$$ps(f) \geq py$$

which is (2).

To show (3), note that from the claim above we have that $\int_A g^k = \lambda^{*k} \to \int_A g = \lambda$ as $k \to \infty$. But we have shown in (2) that $g(a) \in D(a, p, \lambda)$. Therefore we have that $\lambda = \int_A g(a) \in \int_A D(a, p, \lambda)$, which is the definition of λ being consistent. Then the proof is complete. Q.E.D.

III.E Conclusion

In the introduction it was noted that there has been scarce work on NE in the General Equilibrium literature. The model presented above is an attempt to fulfill this empty spot.

In markets with NE, the definition and existence of competitive equilibrium are problematic. As it was pointed out before, NE are usually bundled with indivisibilities, which makes existence of competitive equilibrium not trivial. The main contribution of the current research is to prove existence of a suitably defined equilibrium in this situation.

It should be pointed out that nothing has been claimed about the efficiency properties of equilibrium. As it is the case when there are externalities, equilibrium cannot be expected to be efficient. The reason is of course the lack of a market for the externalities themselves. In the current setting, inefficiencies

will typical involve the market to be dominated by an inferior standard, compared to some of the alternative standards in the market. Examples of that situation have been presented in the literature, see for example Katz and Shapiro (1985). One of the examples that have been presented are the home video market that was dominated by the VHS standard when apparently the Betamax standard was superior.

In the paper we have used a key assumption, namely that the distribution of income in the population is dispersed for any price vector different from 0. This assumption gives us two desirable properties: First, it allows us to disregard households with budget lines containing troublesome points, i.e. households in which we could observe some discontinuous demand behavior. And second, it makes aggregate behavior smooth, in the sense that aggregate demand changes smoothly with prices. These two features of the model are essential for the existence result.

Finally, a possible extension would be to investigate rigorously the efficiency properties of the equilibrium in markets with NE. As it was pointed out above, we would not expect equilibrium to be Pareto optimum. Working out these results would represent an interesting extension of the current work.

References

[1] Arrow, K. and F. Hahn (1971): General Competitive Analysis. *Holden-Day, Inc., San Francisco.*

[2] Aumann, R. J. (1966): Existence of Competitive Equilibria in Markets with a Continuum of Traders. *Econometrica* 34: 1-17.

[3] Aumann, R. J. (1976): An Elementary Proof that Integration Preserves Upper-hemicontinuity. *Journal of Mathematical Economics* 3: 15-18.

[4] Berglas, E. (1976): On the Theory of Clubs. *American Economic Review* 66: 116-121.

[5] Besen, S. M. and J. Farrell (1994): Choosing How to Compete: Strategies and Tactics in Standardization. *Journal of Economic Perspectives* 8: 117-131.

[6] Buchanan, J. (1965): An Economic Theory of Clubs. *Economica* 1-14.

[7] Cole, H. and E. Prescott (1997): Valuation Equilibrium with Clubs. *Journal of Economic Theory* 74: 19-39.

[8] David, P. A. (1985): Clio and the Economics of QWERTY. *The American Economic Review Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association* 75: 332-337.

[9] Dybvig, P. H. and C. Spatt (1983): Adoption Externalities as Public Goods. *Journal of Public Economics* 20: 231-247.

[10] Ellickson, B., B. Grodal, S. Scotchmer and W. Zame (1999): Clubs and The Market. *Econometrica* 67: 1185-1218.

[11] Farrell, J. and G. Saloner (1985): Standardization, Compatibility, and Innovation. *Rand Journal of Economics* 16: 70-83.

[12] Farrell, J. and G. Saloner (1986a): Standardization and Variety. *Economics Letters* 20: 71-74.

[13] Farrell, J. and G. Saloner (1986b): Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation. *American Economic Review* 76: 940-955.

[14] Farrell, J. and G. Saloner (1988): Coordination Through Committees and Markets. *Rand Journal of Economics* 19: 235-252.

[15] Hildenbrand, W. (1970): Existence of Equilibria for Economies with Production and a Measure Space of Consumers. *Econometrica* 38: 608-623.

[16] Katz, M. L. and C. Shapiro (1985): Network Externalities, Competition, and Compatibility. *American Economic Review* 75: 424-440.

[17] Katz, M. L. and C. Shapiro (1986): Technology Adoption in the Presence of Network Externalities. *Journal of Political Economy* 94: 822-841.

[18] Katz, M. L. and C. Shapiro (1994): Systems Competition and Network Effects. *Journal of Economic Perspectives* 8: 93-115.

[19] Liebowitz, S. J. and S. E. Margolis (1994): Network Externality, An Uncommon Tragedy. *Journal of Economic Perspectives* 8: 133-150.

[20] Ng, Yew-Kwang (1973): The Economic Theory of Clubs: Pareto Optimal Conditions. *Economica* 40: 291-298.

[21] Ng, Yew-Kwang (1974): The Economic Theory of Clubs: Optimal Tax/Subsidy. *Economica* 41: 308-321.

[22] Rohlfs, J. (1974): A Theory of Interdependent Demand for a Communications Service. *Bell Journal of Economics* 5: 16-37.

[23] Sandler, T. and J. T. Tschirhart (1980): The Economic Theory of Clubs: An Evaluative Survey. *Journal of Economic Literature* 18: 1481-1521.

[24] Scotchmer, S. (1997): On Price-taking Equilibria in Club Economies with Nonanonymous Crowding. *Journal of Public Economics* 65: 75-88.

[25] Starr, R. M. (1999): A General Equilibrium Model of Network Externalities. Preliminary Notes. University of California, San Diego.

[26] Yamazaki, A. (1978): An Equilibrium Existence Theorem without Convexity Assumptions. *Econometrica* 46: 541-555.