UNCOVERING ECONOMIC BEHAVIOR IN HETEROGENEOUS INSTITUTIONAL ENVIRONMENTS: THREE ESSAYS ON APPLIED AND EXPERIMENTAL ECONOMICS

A Dissertation

by

JOSE GABRIEL CASTILLO GARCIA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Li Gan
Co-Chair of Committee,	Alex Brown
Committee Members,	Catherine Eckel
	Dmitry Vedenov
Head of Department,	Timothy Gronberg

August 2015

Major Subject: Economics

ABSTRACT

The three essays in this thesis foster the identification of unexpected influences of different institutional arrangements on economic outcomes. The first two essays, experimental in nature, analyze the impact of two different contexts within Public Goods Games (PGG) environments. The first essay documents exact replications of four classic experiments in PGG and cast unexpected results in contribution behavior. First, it shows how the attenuation effect in replication studies, well documented in other disciplines, is also pervasive in experimental economics. Not all previous findings replicate, and effects found in successful replications are much smaller. Second, it shows that experimental context matters; experimental subjects in Texas tend to contribute more and free ride less, across different experiments.

The second essay analyzes whether democratic institutions have any impact on agency problems where group members face a centralized arrangement of sanctioning power. It offers novel evidence, although a weak effect, of the intrinsic incentives for pro-social behavior attached to legitimacy in democratic institutions to promote collective action and higher economic efficiency.

Finally, the last essay offers an empirical alternative to unravel heterogeneous unobserved traits on credit market customers. Through the use of mixture density estimation methods and rich administrative data, it identifies different *quality-types* of clients for credit demand and default decisions. Credit customers differ in their individual preferences, as well as levels of foresight, strategic behavior; all unobserved by the principal (lender). Accounting for these unobserved traits improves the forecast of potential clients' behavior and offers alternatives for different contracts and risk-pricing strategies to reduce credit rationing.

DEDICATION

This work is greatly indebted to the guidance and support of Li Gan (advisor), Alex Brown (co-advisor) and Catherine Eckel (committee member), with whom I have shared and learned the value and intricacies of academic research, and for whom I have nothing but a profound admiration and respect.

Nothing in this world is reached without the greatness that comes from the most humble human feelings. Victory and defeat will always be ephemeral scenarios through which every human soul thrives. No victory is possible without passion, no defeat is worth without reason; no world exists without love.

To my family, my infinite source of life.

To my love, my source of inspiration.

ACKNOWLEDGEMENTS

I gratefully acknowledge the support in the beginning of the journey in my graduate studies to the scholarship received from the Fulbright Commission and SENACYT of Ecuador, and the Economics Department at Texas A&M University.

The first essay of this dissertation is coauthored with Catherine Eckel and Haley Harwel, from the Economics Department at TAMU. It owes to the support and effort of the staff at the Center for Behavioral and Experimental Economic Science at UT Dallas, especially Wendy Mak Lee, Eric McLester, Sheheryar Banuri, and the many undergraduate research assistants who helped conduct the sessions. Funding was provided by CBEES and the Negotiations Center at UT Dallas.

The second essay is coauthored with Zhicheng Xu (Phil), from the Department of Agricultural Economics at TAMU. We are greatly indebted to Ping Zhang and Jia Yan from Shenzhen University in China, for research funding and outstanding laboratory support.

Finally, the fourth chapter would have not been possible without the support of an anonymous financial institution for which help I am grateful.

I also thank Li Gan, Alex Brown, Catherine Eckel, Dmitry Vedenov, Steve Puller, Marco Palma, Daniel Fragiadakis, Charles Plott, Enrique Fatas, Mark Issac, Manuel Hernández, Roberto Mosquera, fellow members of the experimental research group at TAMU and several other participants of the 2014 ESA North American Meeting in Fort Lauderdale-FL, for their valuable comments and insights in this research.

The corresponding disclaimer applies. All errors are mine.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

The fact that different institutional environments shape economic outcomes is not new in economics. Nevertheless, beyond any theoretical argument, the study of such effects represents an empirical challenge and the understanding of *cause and effect*, in a scientific flavor, is always ambitious due to the complex nature of social interaction. To this endeavor, social sciences have to add the fact that the smallest molecule or unit of study, the individual, as an economic agent, has own will. The intricacies of human behavior, before studied only in psychological sciences, have had a great impact in the study of economics. Methods, theories and a new body of knowledge accounts for a very lively research agenda that has successfully identified some interesting behavioral regularities, as well as inconsistencies, that offer a new perspective to old questions of rational choice and decision making.

One remarkable milestone in the discipline is the rise of experimental and behavioral economics to which most of this work subscribes. Not free of valid criticism, specially due to external validity arguments, the field builds a body of knowledge that ties highly divorced theoretical and empirical research in economics. The experimental nature of the field supports a more scientific treatment not only due to the efforts to isolate economic influences of treatments through randomization, but also due to the replicability of the research, which offers higher consistency on the results.

The three essays contribute to the experimental and applied economic research by identifying unexpected influences of different institutional arrangements on economic outcomes, as well as novel econometric applications. The first two chapters analyze the impact of two different context within Public Goods Games (PGG) environments. The first essay documents exact replications of four classic experiments in PGG and cast unexpected results in contribution behavior. First, it shows how the attenuation effect in replication studies, well documented in other disciplines such as medicine, is also pervasive in experimental economics. Not all previous findings replicate and, when they do, their effects are much smaller. Second, we show that, when it comes to economic decisions, context matters; experimental subjects in Texas, mostly undergrad students, tend to contribute more and free ride less across all experiments.

The second essay studies economic efficiency on two different institutional arrangements of sanctioning power: endogenous versus exogenous power delegation. We analyze whether democratic institutions have any impact on agency problems where group members face a centralized arrangement of sanctioning power. This study offers novel evidence, although partially weak, of the identification of intrinsic individual incentives of democratic institutions towards cooperation. Incentives towards collective action increase due to authority legitimacy, regardless of the actual behavior or performance of the leader; as a result higher economic efficiency is achieved.

Finally, the last essay, non-experimental, offers an empirical alternative to unravel heterogeneous unobserved traits, arguably individual preferences, on the credit market. Following Gan and Mosquera (2008) and Gan, Hernández and Liu (2013), through the use of mixture density estimation methods and rich administrative data, I identify the presence of private information that intuitively can be interpreted as different *quality-types* of clients for default decisions and credit demand. Credit customers differ in their individual preferences, as well as levels of foresight and strategic behavior; all unobserved by the principal (lender). Through out-of-sample performance analysis, I show that the estimation method proposed improves the forecast of potential clients' default behavior over any current regression technique and offers alternatives for different contracts and risk-pricing strategies to reduce credit rationing and attain higher market efficiency.

Understanding unexpected economic consequences of institutional environments remains the most interesting empirical research agenda in applied microeconomics. The three essays of this thesis combine diverse empirical approaches to answer different questions in this common research goal.

2. FOUR CLASSIC PUBLIC GOODS EXPERIMENTS: A REPLICATION STUDY*

2.1 Introduction

This chapter introduces Eckel et al. (2015)*and motivates the discussion over the importance of replications in Experimental Economics, a topic highly overlooked in this relatively new field.

Considerable attention has been focused on the problems of publication bias, selective reporting, and the importance of research transparency in the social sciences, especially in recent years. Publication bias occurs because articles with findings that are statistically significant, theoretically interesting, and novel are more likely to be published than null, dull, or replication studies. Selective reporting means that more interesting or novel findings within a study are more likely to be published, and insignificant or non-intuitive results left in a file drawer. The evidence on reported significance levels suggests a serious bias, leading Ioannidis (2005) to assert that "most published research findings are false." His model predicted "rates of wrongness" in medical research of 80 percent for non-randomized studies, 25 percent of randomized trials (the "gold standard" for experimental research), and ten percent of large-scale randomized trials ("platinum standard"). These correspond roughly to the rates at which medical study results are overturned. In a related observation, Schooler (2011) notes that attempts to repeat studies often result in an apparent decline in treatment effects (such as estimates of drug effectiveness) over time, a phenomenon he has referred to as "cosmic habituation"(Lehrer 2010), and calls for a open repository of unpublished results to help combat this problem. Replication

*Reprinted with permission from Eckel, Harwel and Castillo. "Four Classic Public Goods Experiments: A replication study," *Research in Experimental Economics*, Vol. 18, Emerald Group Publishing, forthcoming.

seems an important element of the response to demonstrated bias in published results. Closer to home, a recent controversy over the replicability of research results in social psychology, and in priming research in particular, is highlighted in an open letter from Daniel Kahneman (Yong 2012), who called on his colleagues in priming research to respond to the criticism.

> I believe that you should collectively do something about this mess. To deal effectively with the doubts you should acknowledge their existence and confront them straight on, because a posture of defiant denial is self-defeating. Specifically, I believe that you should have an association, with a board that might include prominent social psychologists from other field [sic]. The first mission of the board would be to organize an effort to examine the replicability of priming results, following a protocol that avoids the questions that have been raised and guarantees credibility among colleagues outside the field (Yong 2012, supplementary material).

The profession was quick to respond. This effort resulted in a special issue of the Journal of Social Psychology (2014) on replications, which includes, among a number of other papers, a notable effort by the "Many Labs" replication project to replicate 13 major results in the field (Klein et al. 2014). This paper reports the results of experiments conducted in 36 labs with over 6,000 subjects, and finds that ten effects are robust, one is weak, and two fail to replicate (both priming studies, as it happens). While the flurry of replication in social psychology was inspired by a storm of criticism and skepticism, including the discovery of fraudulent behavior by social psychologists such as Diederik Stapel, Dirk Smeesters and Lawrence Sanna, economics research has not come under similar attack. Nevertheless, replicability remains an important aspect of the appeal of experimental research, and is a valuable activity. Although Plott (1982) refers to replication as “the heart of experimental economics,” published replication is not commonplace in experimental economics. Published replications are most often found as a first step in a new research project or extension (e.g., Hung and Plott 2001). In the realm of the public goods experiment, one of the most popular “canonical games” in economics (Eckel 2007), many studies have replicated prior results in somewhat different environments.

With this in mind, we set out explicitly to replicate as closely as possible several important, highly-cited public goods experiments. Students in a PhD course in Experimental Economics selected four papers to replicate: Isaac and Walker (1988); Andreoni (1995a); Andreoni (1995b); and Fehr and Gächter (2000). These papers contain some of the most important results in the study of public goods experiments. The Voluntary Contribution Mechanism public goods game is a simple environment designed to test theories about social dilemmas. In these games subjects make decisions in groups, typically of size 2 to 10, and each member of the group must allocate a fixed endowment between an individual investment, which pays \$1 to the decision maker for each \$1 invested, and a public investment, which pays a lower amount (less than \$1 but more than \$1/n), termed the Marginal Per Capita Return (or MPCR) to each member of the group. The incentive structure mimics that of a public good, and embodies the classic tension between the social optimum of full contributions to the public good, and individual payoff-maximization, which entails free riding on the contributions of others. From the earliest studies using this game, results have shown positive levels of contributions that then deteriorate over time, in most environments (Ledyard 1995). The first of the studies we replicate, Isaac and Walker (1988) tests the effect of changing the MPCR, and group size. In the original study, they find that the MPCR has an important impact on contributions, with a move from .3 to .75 increasing average contributions and decreasing the number of free riders. The effect of changing group size from 4 to 10 is weaker. To our knowledge, no one has replicated their design. Andreoni (1995a, 1995b) conducted two important variations on the public goods game. In the former, he examined whether the contribution level in the game might be due to confusion, since free riding is a dominant strategy, and found considerable evidence that confusion plays a role. In the second, he explored the effect of positive or negative framing on contribution levels. These two studies constitute the second and third replications. In Andreoni (1995a), the experiment consisted of three treatments: regular, a standard public good game with groups of size 5 and an MPCR of 0.5; rank, where earnings depended only on the subjects' relative payoff in the game and there was no payoff to cooperation; and regrank, which provided the same information on relative earnings as the rank condition, but earnings were determined as in the regular condition. He found that approximately half of the total contribution to a public good game could be attributed to kindness and the other half to potential confusion, but that confusion decreased rapidly across the first five rounds, while the kindness remained more or less stable. The presence of confused subjects has been confirmed by Houser and Kurzban (2002) and Ferraro and Vossler (2010), but in somewhat different environments. Both involve a comparison between play against real counterparts with play against computerized players, providing a simpler test of confusion than Andreoni (1995a), and both studies also show that confusion declines over time. The approach taken in Andreoni (1995b) is to vary the description of the game in two treatments, a positive-frame description, in which subjects' contributions to the group account generate a positive externality for others in the group, and a negative-frame description in which contributing to the individual account generates a negative externality for each member of the group. The payoff functions are identical in the two treatments; only the description changes. This manipulation leads to substantially lower average contributions and more free riders in the negative-frame treatment. A number of studies have tested the effect of positive and negative framing in variations on the public goods game, but none has attempted to exactly replicate the original study. Sonnemans, Schram, and Offerman (1998) find a similar, but weaker, effect in a binary, step-level public good experiment; however, Bougherara, Denant-Boemont, and Masclet (2011) show that framing has an even more powerful effect in a provision-point mechanism public good setting. Willinger and Ziegelmeyer (1999) replicate the effect of framing differences in a public good game with an interior Nash equilibrium; they find that while positive framing leads to above-equilibrium contributions, negative framing leads subjects to play the Nash equilibrium. Park (2000) explores heterogeneity in responses to positive and negative framing and show stronger effects for subjects who have and individualistic value orientation, in contrast to this with a cooperative value orientation. In a somewhat different environment, Shanley and Grossman (2007) show that subjects are more willing to contribute to a public good than to refrain from contributing to a "public bad" that, in essence, deteriorates a public good; they find no difference in the number of free riders. Grossman and Eckel (2012) find no effect of positive and negative framing in a "real donation" dictator game study, where donations go to a charitable organization producing public goods in the field. Finally, the fourth study, Fehr and Gächter (2000), is one of the first to test the effect of sanctions on public goods contributions. This study adds an additional stage to each round of the public good game in which subjects can, at a cost to themselves, punish other members of their group. Punishment is introduced in a "partners" setting, where groups are stable across rounds of the game, and a "strangers" setting, where groups are reconstituted each round. Punishment is very effective in enhancing contributions. Since their paper was published, there have been many studies (too many to survey here) testing various aspects of the structure and effectiveness of punishment institutions. (Chaudhuri 2011) provides a useful survey of research up to that point. In a concluding statement he notes that the opportunity to punish in such games generally increases contributions, but given the cost of punishment, does not always enhance efficiency. Besides the issue with efficiency, he suggests two additional cautions. First, he notes that punishment itself is a second-order public good, and could require punishment of the non-punishers (free riders on punishment) as well as the free riders on contributions. Second, most studies show some anti-social punishment, which appears to be counter-punishment of high contributors by low contributors.

In most cases we replicate the pattern in the original data, but the treatment effects are consistently smaller than in the original studies. In some cases, the results of the original study cannot be replicated, in that the treatment effect sizes fall below statistical significance. All studies show a positive Texas effect: the UT Dallas subjects consistently contribute higher amounts to the public good than in the original studies.

2.2 Experimental design and procedures

All of the experiments were conducted in a way that mimicked as closely as possible the original conditions of the experiments, with one exception: To ensure comparability across replications, all experiments were computerized. When possible, the original program was obtained from the authors, but otherwise the games were programmed using z-tree (Fischbacher 2007). Payment was adjusted to reflect current norms of payment for subjects: all received a $5 show-up fee, and average earnings were $13-$19. Subjects for all experiments were recruited using ORSEE (Greiner 2004), and sessions took place at the Center for Behavioral and Experimental Economic Science (CBEES) at the University of Texas at Dallas, in April and May 2012. A summary of the titles and treatments is included in Table 1. For IW, and the two Andreoni studies we collected at least as much data as the original study, but for several of the studies, the show up fee is not reported.

Table 2.1: Original and replication design comparison

	Isaac & Walker		Andreoni Kindness/Confusion		Andreoni Warm Glow/ Cold Prickle		Fehr & Gächter	
	1988	Repl.	1995	Repl.	1995	Repl.	2000	Repl.
Show-Up Fee	n/a	$5	n/a	$5	n/a	$5	15 CHF ($9)	$5
Subjects	84	84	120	120	80	60	112	48
Sessions	12	12	2	12	2	12	5	5
Average Earnings ($)		15	8.68	13.92	8.24	15.28	41 CHF ($25)	18.7
Time (minutes)	n/a	n/a	50	30	50	41	120	70
Program	Plato	z-Tree	None	z-Tree	None	z-Tree	z-Tree	z-Tree

2.2.1 Isaac & Walker, group size effects in public goods provision

The 2x2 factorial design includes two treatments, varying the group size (n) from 4 to 10; and the MPCR from .3 to .75, as in the original study. The design also varies the total endowment of tokens for each individual across groups, with an eye to keeping the maximum earnings relatively constant across treatments. The endowment varied in the four person treatment from 62 tokens with the lower .3 MPCR, to 25 in the .75 MPCR rounds. The endowment also varied with the MPCR in the 10-group treatment. Subjects were endowed with 25 tokens in the low .3 MPCR treatment to 10 tokens in the high .75 MPCR rounds. One treatment is played for ten rounds in stable groups, followed by a surprise restart and a second ten rounds with a different MPCR. For the 4-person groups, sessions included 12 subjects (three groups of 4) in anonymous groups. For the 10-person groups, each session was a single group. Subjects were randomly assigned to a computer station; dividers ensured that decisions were anonymous. The replication used z-tree to code the experiment, where the original authors used PLATO as their computerized platforms. Subjects completed self-paced instructions that were the same as in the original study. At the end of each round subjects were shown a summary of group contributions and their own group and individual earnings from the round. These were totaled after the final round and converted to dollars at the rate of $0.005 per token. The average individual earnings, including a $5 show up fee, were about $15. After subjects completed the first treatment of ten rounds, they were informed they would participate in another ten rounds. To account for sequencing effects, the treatment order was counterbalanced, with six sessions having high then low MPCR, and six the reverse. To be consistent with the work of Isaac and Walker, our data analysis only considers the data from the second period of 10 rounds. Isaac and Walker considered the last ten rounds to be conservative.

2.2.2 Andreoni, kindness or confusion

The experiment replicates the three treatments of the original design: named regular, rank and regrank, which vary the earnings calculation (based on the standard game or based on rank) and information (standard or standard + rank), as explained above. The design is between-subjects, with each session consisting of a single treatment played in groups of five persons for ten rounds, with two anonymous groups per session. The endowment is 60 tokens, and the MPCR is 0.5. The experiment was coded in z-tree; the original study was conducted by hand. The regular condition is the standard game. In the rank condition, payoffs are calculated based only on the rank of the subject's earnings in the game, removing any incentive to contribute . Effectively, this treatment "makes a zero-sum payoff game out of a standard public goods game." (Andreoni, 1995). Any remaining contributions could be attributed to confusion. Because these two treatments differ in both the earnings structure and the information given to the subjects about rank, the third treatment, regrank, removes the confound in the design by using the standard payoff calculation but providing rank information. Earnings were about $14 including a $5 show-up fee.

2.2.3 Andreoni, warm glow versus cold prickle

The design consists of two treatments: a standard game, which is a "positive" frame mentioning the positive externality of a contribution to the group account on the other group members, and a negatively framed game, which emphasizes the negative externality associated with a contribution to the individual account. The design is between-subjects, with each session consisting of a single treatment played in groups of five persons for ten rounds, with two anonymous, rematched groups per session. The endowment is 60 tokens, and the MPCR is 0.5. Like the Andreoni (1995a) study, the experiment is coded in z-tree; the original study was conducted by hand. Earnings were about $15 including the $5 show up fee. The positive frame had standard instructions. In the negative treatment subjects were again given 60 tokens in each round and were told they may allocate them however they wish between the two accounts. However, subjects were also given automatic earnings totaling 120 additional tokens each round. They were told that investing in the individual account offered a one-for-one payoff for the subject but also led to a deduction of one-half a token from the earnings of all the other group members.

2.2.4 Fehr & Gächter, cooperation and punishment

The 2x2 design varies two elements of the game: whether there is punishment, and whether groups are stable (partners) or reconstituted each period (strangers).

Table 2.2: Punishment cost function

Punishment Points	0	1	2	3	4	5	6	7	8	9	10
Cost of Punishment	0	1	2	4	6	9	12	16	20	25	30

The group composition is between-subjects, and the punishment element is within-subjects. In each session subjects in groups of four first complete ten rounds of punishment, followed by ten rounds without, or vice versa. Each session consists of 12 subjects (three groups of four). Subjects received an endowment of 20 tokens and the MPCR of the public account is 0.4. The experiment is based on a translation of the original study and is programmed in z-tree. Earnings were about $19, including the show up fee. In the no-punishment condition, the standard game is played for ten rounds. The punishment treatment adds a second stage where subjects can simultaneously punish one another, at a cost to themselves. The cost of punishing is the same as in the original paper.

2.3 Results

2.3.1 Isaac & Walker (1988)

Tests of treatment differences using aggregate results from Isaac and Walker (1988) and the Texas replication are shown in Table 3. The levels of contributions and proportion of "strong free riders" (subjects with contributions less than 1/3 of their endowments) can be thus analyzed. For example, the first row of Table 3 tests whether low and high MPCR generate different outcomes in groups of four players; the second performs the same test for groups of 10. While in the original study, high MPCR leads to significantly higher contributions for both group sizes, we are unable to replicate that result in the Texas data. The impact of group size is never significant in our data, though it is in Isaac and Walker for low MPCR. However, we do replicate three out of four comparisons on the number of strong free riders in each treatment.

Table 2.3: Isaac & Walker (1988): non-parametric tests and treatment differences within sample, both studies

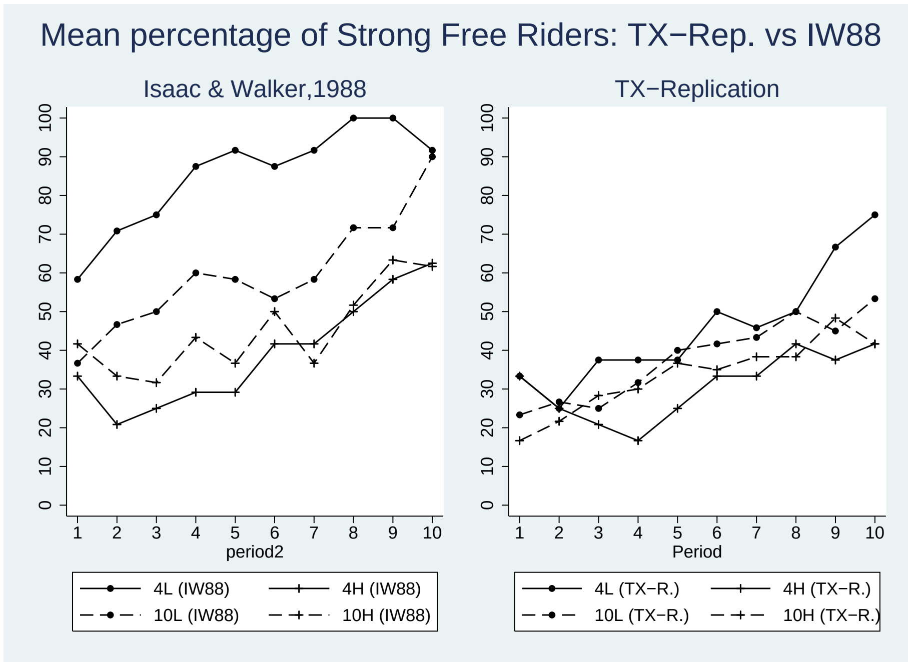
Non-parametric tests for equal means (N=12 for 4L and 4H; N=40 for 10H and 10L)		**Isaac and Walker** z	**Isaac and Walker** p-value	**TX-Replication** z	**TX-Replication** p-value
Contributions	**4L vs. 4H**	4.107	0.000	0.691	0.490
	10L vs. 10H	3.200	0.001	1.218	0.223
	4H vs 10H	0.451	0.652	0.010	0.992
	4L vs. 10L	4.666	0.000	0.000	1.000
Number of Strong free-riders	**4L vs. 4H**	5.101	0.000	2.959	0.003
	10L vs. 10H	2.888	0.004	1.115	0.265
	4H vs 10L	5.428	0.000	5.158	0.000
	4H vs 10H	5.379	0.000	5.158	0.000
	4L vs. 10L	4.933	0.000	4.272	0.000

Two-tail test from a normal approximation at 5% significance level. Tests are performed at the individual level (one observation per individual) using a Wilcoxon rank-sum test.
In contrast to the original study, tests include the 20 periods (2 sequences) of the experiment.
Strong free-riders are those with $contributions < 1/3 * endowment$.

Figures 1A and 1B show the pattern of results over time in the Isaac & Walker and replication data. In both, the left panel graphs the Isaac & Walker data, and the right the Texas replication. It is easy to see that the treatment differences are much weaker in the Texas data: neither changes in MPCR or group size appear to have substantial impacts on contribution levels.

Figure 2.1: Comparison of Isaac & Walker (1988) with Texas replication

(a)



(b)

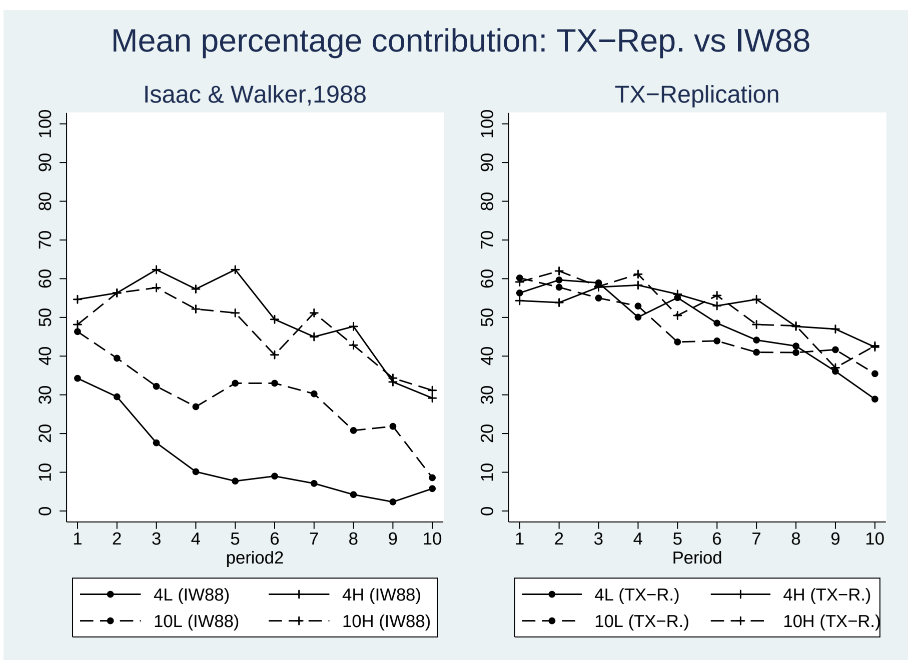


Table 2.4: Determinants of contributions, regression results
(TX-Replication + Isaac & Walker (1988))

Dependent variable: Percentage Contributions			
	RE	**RE_Int**	**RE_Int_Lag**
Texas (replication=1)	0.139***	0.353**	0.353**
	(0.0385)	(0.149)	(0.150)
MPCR (high=1)	0.207***	0.370***	0.276***
	(0.0435)	(0.0161)	(0.0131)
Group Size (high=1)	0.0785	0.165***	0.114***
	(0.0755)	(0.0198)	(0.0176)
MPCR * Group Size	-0.0961***	-0.197***	-0.132***
	(0.0263)	(0.0188)	(0.0189)
Texas * MPCR		-0.325***	-0.224***
		(0.0448)	(0.0505)
Texas * Group Size		-0.172	-0.433***
		(0.143)	(0.0960)
Texas*MPCR*Group Size		0.202***	0.102
		(0.0519)	(0.0577)
Other's Average Contribution-OAC(t-1)			0.276***
			(0.0304)
Texas *OAC(t-1)			-0.201***
			(0.0336)
Constant	0.235***	0.128***	0.0810***
	(0.0634)	(0.00826)	(0.00433)
N (observations)	3360	3360	3192
N (subjects)	168	168	168
r2	0.0761	0.0970	0.128

Random Effects estimation. Standard errors clustered at group level in parentheses.
Percentage contributions: contribution/endowment by experiment.
Significance at: * 10% level, ** 5% level, *** 1% level.

Regression analysis on the percent contributions shows significant treatment effects for both MPCR and group size in the pooled data (Table 4). The first model also reveals negative interaction term indicating the treatment effect is different for the larger groups. Notably, there is a positive, significant coefficient on a dummy variable indicating the Texas replication data. In the second model we add Texas interaction effects for the treatment variables. The treatment effects are clearly different in the replication; indeed, the interaction effects indicate that there are no significant treatment effects for the Texas data, the interaction coefficient offsets the main effects. The results are unchanged when we also control for the contributions of others, using a variable capturing the lagged contributions of others in the group ($OAC(t-1)$).

The data for the first replication study fail to replicate the original study results. While the pattern of results is consistent with the original study, the effect sizes are quite different from the original study.

2.3.2 Andreoni (1995a): kindness or confusion

Tests for aggregate treatment effects for Andreoni (1995a) are summarized in the table below. This table shows that we were unable to find a significant difference between treatments reg versus regrank in the replication data. The difference between regrank and rank are marginally significant, and we do see significant differences between the reg and rank treatments. The per-period averages for each treatment follow a similar pattern to the original results with contributions for regular > regrank > rank, and free riders for regular < regrank < rank, but smaller differences between the treatments.

Table 2.5: Andreoni (1995a): non-parametric tests and treatment differences within sample, both studies

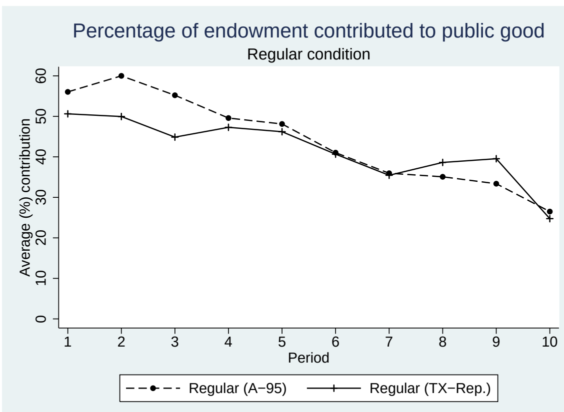
Non-parametric tests equal means		**Andreoni**			**TX-Rep.**	
(n=40 for both studies)		**z**	**p-value**	**z(A95)**	**z**	**p-value**
Contributions	**Reg. vs. Reg.Rank**	3.766	0.0003	3.772	1.473	0.1407
	Reg.Rank vs.Rank	3.602	0.0077	3.58	1.638	0.1014
	Reg. vs.Rank	5.091	0		2.995	0.0027
Number of pure free riders	**Reg. vs. Reg.Rank**	1.817	0.0692	2.281	1.705	0.0883
	Reg.Rank vs.Rank	2.88	0.004	2.42	0.492	0.6224
	Reg. vs.Rank	3.602	0.0003		1.971	0.0488

Two-tail test from a normal approximation at 5% significance level.
Tests are performed at the individual level using the Wilcoxon rank-sum test.
z(A95): corresponding tests reported in the original paper.

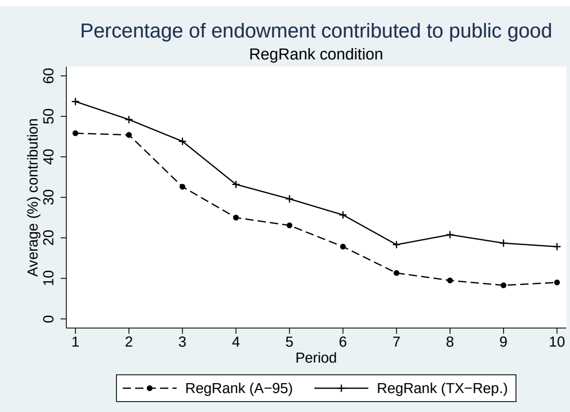
The pairwise comparisons for the treatments for all 10 periods are shown in Figure 2 A-C below. (We illustrate these separately because putting all in one figure is overly cluttered.) The regular condition shows very similar behavior in the two studies. The regrank and rank conditions show higher contributions in Texas than the original.

Figure 2.2: Comparison of Andreoni (1995a) with Texas replication

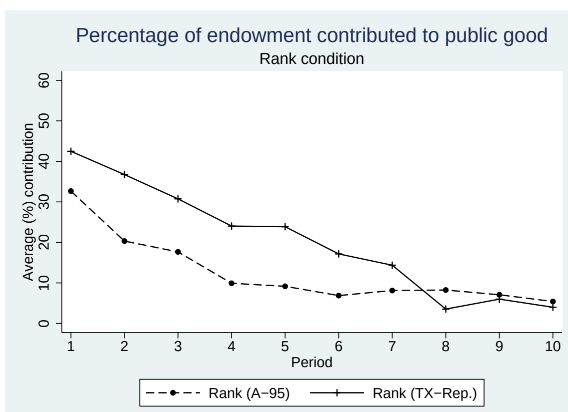
(a)



(b)



(c)



In the original paper, confusion is measured as the difference between regrank and rank . In both treatments subjects receive information about rank, but only in the latter treatment does rank alone determine earnings. Andreoni argues that the comparison between these two provides a clean measure of confusion. In the Texas data, confusion appears to play a larger role, as contributions are somewhat higher in the rank treatment as compared with the original. While the level of confusion declines over time in both data sets, the levels of confusion are higher in the Texas data and the levels of kindness lower. However the proportion of subjects that couldn't be classified as motivated by either is larger in Andreoni's data.

Table 2.6: Determinants of contributions, regression results
(TX-Replication + Andreoni (1995a))

	Dependent variable: Contributions			
	RE (se: robust)	**RE (se: cl)**	**RE (se: cl)**	**RE (se: cl)**
Texas (replication=1)	-1.375	-1.375	-0.450	-3.120
	(1.639)	(3.231)	(1.597)	(2.336)
Treatment 2: RegRank	-18.93***	-18.93***	-9.503**	-10.59*
	(1.418)	(2.123)	(2.575)	(1.508)
Treatment 3: Rank	-12.78***	-12.78**	-6.899*	-7.62**
	(1.435)	(3.933)	(2.643)	(1.94)
Texas * RegRank	6.025***	6.025	2.289	3.941*
	(2.032)	(4.011)	(3.062)	(1.457)
Texas * Rank	6.352***	6.352	2.948	3.963
	(2.152)	(5.374)	(2.926)	(1.886)
Other's Avg. Contribution-OAC(t-1)			0.381***	0.339***
			(0.0570)	(0.00912)
Texas * OAC(t-1)				0.0741
				(0.101)
Constant	26.46***	26.46***	11.63**	13.17***
	(1.152)	(2.059)	(2.970)	(1.111)
N (observations)	2400	2400	2160	2160
N (subjects)	240	240	240	240
r2	0.104	0.104	0.206	0.207

Random Effects estimation. Standard errors clustered at group level in parentheses for models 2,3 and 4.
Contributions: tokens per period, 60 each.
Excluded treatment: Regular.
Significance at: * 10% level, ** 5% level, *** 1% level.

Table 6 shows a regression analysis, pooling the original and replication data. The first and second models are the same except in the way the standard errors are calculated (robust versus clustered at the group level). Using the more conservative second model, this illustrates the treatment effects in the original study, with the regrank treatment reducing contributions from the regular treatment, and rank reducing it still further. However the Texas interactions partially offset the treatment effects, showing that in the replication data the effect sizes are smaller. In some cases these fall below statistical significance. Texas contributions are higher in both the rank and regrank treatments.

In sum, we see the same pattern of results in the replication as in the original data, but the effect sizes are smaller. Confusion as measured in the Andreoni's study appears to be somewhat higher in the replication. Texas contributions are higher in two out of the three treatments.

2.3.3 Andreoni (1995b): warm glow versus cold prickle

The results for Andreoni (1995b) are summarized in table 7 below. The treatment difference between the positive and negative frame is successfully replicated, with significant differences in both contributions and the number of free riders.

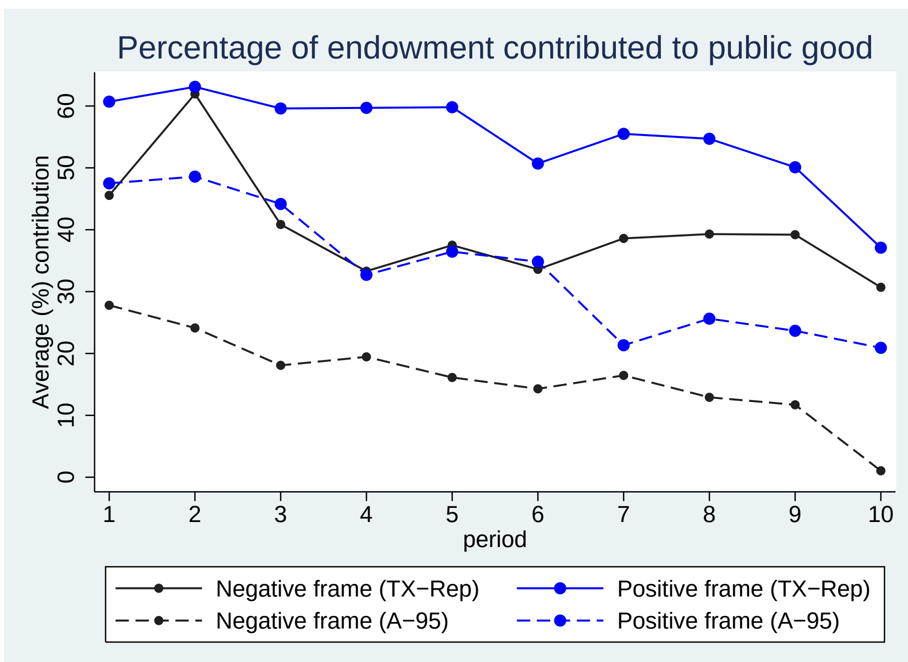
Table 2.7: Andreoni (1995b): non-parametric tests and treatment differences within sample, both studies

Non-parametric tests for equal means		**Andreoni**			**TX-Replication**	
		z	**p-value**	**z (A95)**	**z**	**p-value**
Contributions	**Pos. vs. Neg. frame**	3.51	0.0005	3.44	2.611	0.009
Number of pure free riders	**Pos. vs. Neg. frame**	3.79	0.0002	3.5	3.121	0.002

Two-tail test from a normal approximation at 5% significance level.
Tests are performed at the individual level using the Wilcoxon rank-sum test.
z(A95): corresponding tests reported in the original paper.

The tables from the original study were replicated (information available upon request), and include contributions for each period of play. Figure 3 illustrates the findings and shows the pattern of behavior over time. The original data (dashed lines) clearly show the treatment effect, and the Texas data replicate the effect. Notably, the replication shows a substantially higher level of contributions for both treatments. Additional tests of statistical significance for the Texas effect were performed and are available upon request.

Figure 2.3: Comparison of Andreoni (1995b) with Texas replication



Regression analysis in Table 8 confirms these results. Using robust or clustered standard errors, the treatment effect for positive frame (in contrast to negative frame) is positive and statistically significant. There is a pronounced Texas effect in the data, with Texans contributing on average about twice as much as in the original study.

Controlling for the contributions of others shows a familiar "matching" effect, with higher contributions by others associated with higher own contributions (Croson 2007). The interactions effects in the fourth model show that there is no significant difference in the treatment effect between Texas and the original study, and that the Texans do not respond differently to the information about prior contributions of others.

Table 2.8: Determinants of contributions, regression results
(TX-Replication + Andreoni (1995b))

	Dependent variable: Contributions			
	RE (se:robust)	**RE (se:cl)**	**RE (se:cl)**	**RE (se:cl)**
Texas (replication=1)	14.31***	14.31***	10.58***	9.607**
	(1.578)	(0.661)	(0.527)	(2.675)
Treatment (Positive frame)	10.43***	10.43**	7.22***	7.504**
	(1.350)	(1.825)	(1.041)	(1.516)
Texas * Treatment	-1.408	-1.408	-1.063	-1.610
	(2.218)	(1.982)	(1.407)	(2.381)
Other's Avg. Contribution-OAC(t-1)			0.223**	0.202*
			(0.0473)	(0.0802)
Texas * OAC(t-1)				0.0412
				(0.117)
Constant	9.720***	9.720***	5.757**	6.053**
	(0.853)	(0.887)	(1.399)	(1.545)
N (observations)	1400	1400	1260	1260
N (subjects)	140	140	140	140
r2	0.146	0.146	0.175	0.175

Random Effects estimation. Standard errors clustered at group level in parentheses for models 2,3 and 4.
Contributions: tokens per period, 60 each.
Excluded treatment: Regular.
Significance at: * 10% level, ** 5% level, *** 1% level.

In sum, the treatment effect of a positive versus a negative frame on identical public goods experiments is successfully replicated, with one key difference: subjects in Texas make substantially higher contributions than in the original study.

2.3.4 Fehr & Gächter (2000)

The results for the Fehr and Gächter (2000) paper largely replicate the original study. Recall that the treatments vary group matching (Strangers versus Partners), punishment (No-Punishment versus Punishment), and order. Table 9 shows the average treatment effects for the replication, alongside the same tests with the original data. In all cases the treatment effect is replicated in the Texas data, though treatment effect sizes are smaller ($p < .0002$).

Table 2.9: Fehr & Gächter (2000): non-parametric tests and treatment differences within sample, both studies

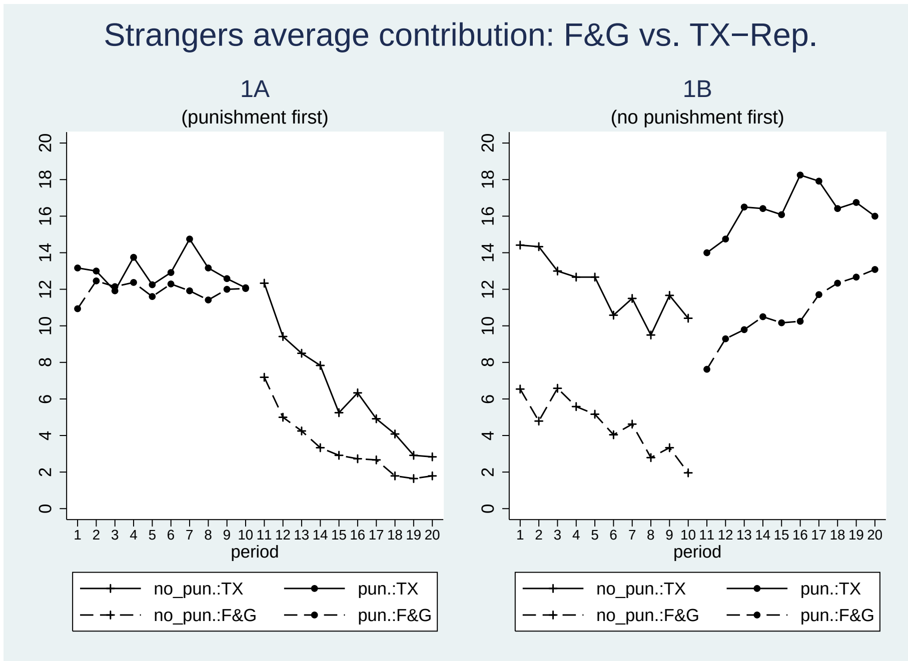
Non-parametric tests:	**Texas Replication**		**F&G Original**	
No-punishment vs. Punishment	**z**	**p-value**	**z**	**p-value**
Strangers	3.959	0.0001	7.161	0
Partners	1.992	0.0464	2.803	0.0051

Two-tail test from a normal approximation at 5% significance level.
Tests are performed at the individual level. Wilcoxon match pairs test reported.
Aggregate results for F&G2000 calculated from original data set.

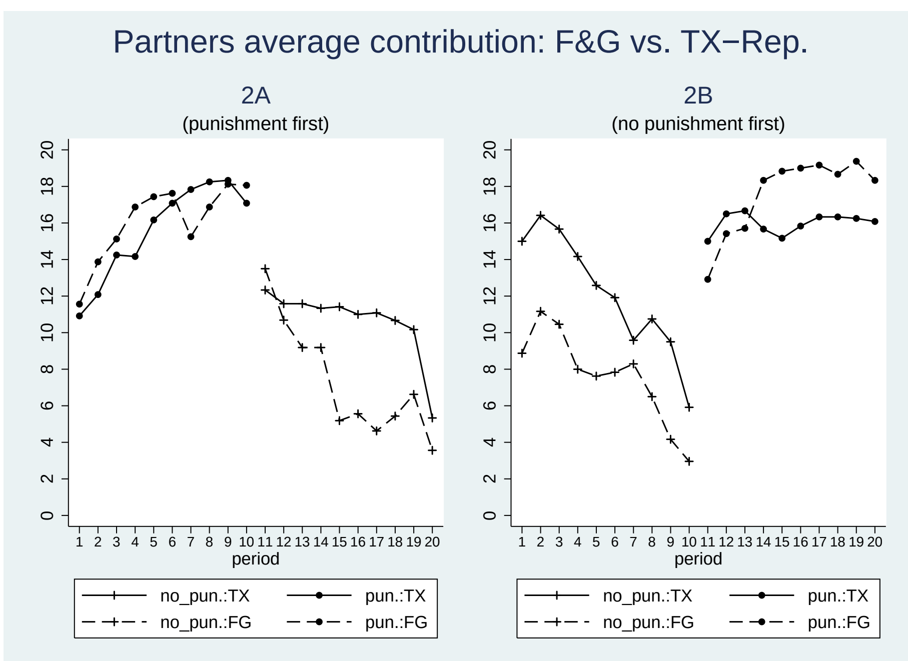
Figure 4 (a) shows the aggregate effects for the strangers treatment. The Panel A shows the results when ten rounds with punishment precede ten without, and vice versa for panel B. When punishment is first, the Texas data are very close to the original; when punishment is removed, the decay in contributions is slower for the Texas data. When no-punishment is first, the level of contributions in the Texas data is considerably higher, but the treatment effect is replicated.

Figure 2.4: Comparison of Fehr & Gächter (2000) with Texas replication

(a)



(b)



We next turn to the partners treatment. Figure 4 (b) shows the aggregate data when punishment rounds are played first (panel A) and when no-punishment rounds are played first. In panel A, as in the strangers treatments, the Texas data for punishment rounds matches the original data, and the decline is slower once punishment is removed. In panel B, the no-punishment rounds again are above the original; the punishment rounds are slightly below.

Table 2.10: Determinants of contributions, regression results
(TX-Replication + Fehr & Gächter (2000))

	Dependent variable: Contributions		
	RE (se: robust)	**RE (se: cl)**	**RE (se: cl)**
Texas (replication=1)	4.788***	4.788**	2.334**
	(0.402)	(1.371)	(0.725)
Treatment (punishment)	8.095***	8.095***	6.148***
	(0.295)	(0.435)	(0.270)
Texas * Treatment	-3.628***	-3.628**	-2.392**
	(0.514)	(1.349)	(0.902)
Partner (=1)	3.245***	3.245**	1.547
	(0.359)	(1.133)	(0.772)
Treatment * Partner	0.843*	0.843	0.657
	(0.462)	(1.025)	(0.469)
Other's Avg. Contribution-OAC(t-1)			0.497***
			(0.0249)
Received punishment points-RPP (t-1)			-0.756***
			(0.185)
Constant	3.919***	3.919***	1.548***
	(0.209)	(0.390)	(0.326)
N (observations)	3200	3200	3040
N (subjects)	160	160	160
r2	0.338	0.338	0.454

Random Effects estimation. Standard errors clustered at group level in parentheses, for models 2, 3.
Contributions: tokens per period, 20 each.
Significance at: * 10% level, ** 5% level, *** 1% level.

Finally, we conduct regression analysis with the pooled data from the original and the replication. These are shown in Table 10. Once again, there is a pronounced

Texas effect, with Texans contributing on average 3 tokens (about 40%) more than in the original study. The regression confirms that punishment increases contributions in both pairings - partners and strangers - but is larger in the partners treatments. In addition, the interaction between Texas and the punishment treatment is negative and significant, indicating a smaller (though still significant) treatment effect in the replication.

From this analysis we conclude that the treatments are replicated. In all conditions, punishment enhances contributions. However, there are two clear differences from the original study. The first is that there is a "Texas effect" - with subjects in Texas contributing significantly more on average than in the original study. Second, the treatment effects are smaller than in the original study, but remain statistically significant.

2.4 Conclusions

We conducted replications of four highly-cited research papers in public goods experiments. For the most part we are able to replicate results from prior studies. In the Isaac and Walker study, we do not find significant treatment effects on contributions for the MPCR in groups of size 4 or 10, but we do find significant effects echoing the original study on the number of strong free riders. For Andreoni (1995a), we replicate the pattern of results showing a strong presence of confusion, and, if anything, show a higher proportion of confusion-based decisions in the replication than in the original. For Andreoni (1995b) we find a strong effect of framing on decision making, as in the original study. Finally, for Fehr and Gächter, we show that punishment increases contributions in both partners and strangers pairings, though our effect sizes are smaller than in the original study. The most notable feature of our results is that in all cases the treatment effects are smaller in the replication than in the original study. Table 11 below summarizes this finding. The left side of the table shows, for each treatment combination, the effect size for the original study (difference between treatments), compared with the effect size for the replication. In all cases these amounts are comparable between the original and the replication, as we replicated the parameters in the originals. In every case but one (Andreoni, regrank - rank), the replication effects are smaller than in the original studies. The largest differences are for the Isaac and Walker paper, and the smallest for Andreoni's warm glow/cold prickle framing study. This decline in treatment effects was predicted by Schooler (2011), who noted a pronounced decline in treatment effects across a wide variety of types of studies. It seems as if economics experiments are not immune from this effect. The right side of the table makes a different point. Here for every treatment we have calculated the difference between the level of contributions in the original study and the level of contributions in the replication. In all but two cases, contributions are higher in the replication. That is, across all studies we observe a Texas effect, with Texans giving at higher levels than in the original study. We can only speculate as to the causes. The subject pool at UT Dallas may differ from those of the earlier studies, but this is impossible to test as the original studies contain no demographic information. The other studies were conducted for the most part at large state universities (Arizona, Wisconsin). UT Dallas is not the flagship university in Texas (that would be UT Austin). In comparison, UT Dallas is smaller, it has the highest entering SAT scores among Texas universities (because of its focus on computer science and engineering), and has a larger fraction of ethnically Indian and Chinese students (48% - though many of these are second-generation), reflecting the predominance of tech-related industries in the north-Dallas area where the university is located. This is in comparison to Arizona and Wisconsin, where about 10 percent of students were minorities in 1995. We are unaware of any studies showing

Table 2.11: Treatment effects and the "Texas effect:" differences in the average percentage contribution to the public good

	Treatment Effects: Difference in Treatments			**The Texas Effect**	
	Original	Replication	(Rep.-Original)	(Rep.−Original)	
Isaac-Walker (1998)					
4H-4L	36.99	4.46	-32.53	4L	35.26
10H-10L	17.28	4.96	-12.32	4H	2.73
10L-4L	16.47	-0.78	-17.25	10L	18.01
10H-4H	-3.24	-0.28	2.96	10H	5.69
Andreoni K/C (1995a)					
Reg-regrank	21.3	10.71	-10.59	Reg	-2.29
Regrank-rank	10.24	10.79	0.55	Regrank	8.3
Reg-rank	31.54	21.5	-10.04	Rank	7.75
Andreoni W/G (1995b)					
WG-CP	17.38	15.03	-2.35	Positive	21.51
				Negative	23.86
Fehr-Gächter (2000)					
Strangers	39.00	26.85	-12.15	Strangers-NP	27.80
Partners	47.50	22.00	-25.5	Strangers-P	15.65
				Partners-NP	19.50
				Partners-P	-6.00

Units reported are the average percentage contribution with respect to each experiment's endowment.
P-NP: Punishment first followed by No Punishment, and the corresponding combinations.

systematic differences in contributions associated with these factors.

Higher giving in the game might simply be due to higher levels of personal generosity among Texans. It is also possible that confusion plays a role, as evidenced by the somewhat higher level of confusion shown in the replication of Andreoni (1995a). However, it is not unreasonable to think that Texans might be more generous, stereotypes to the contrary. (Whenever we tell folks that we have a pronounced Texas effect in our public goods games, we ask them to guess which way the effect goes, and inevitably they assume the worst of Texans.) Texas is strongly conservative, with Republicans controlling all statewide offices, dominating the state legislature, and holding both US Senate seats and 25/36 seats in the US House of Representatives. The Republican candidate has won the presidential race in all elections in the last three decades. The major Texas cities, including Austin, Dallas, Houston, and San Antonio, usually support Democrats, while their suburbs are heavily Republican. UT Dallas draws largely from Dallas and its suburbs, and so has roughly equal proportions of students who identify themselves as Republican or Democratic. We did not collect political information in this study. Claims are often made of greater giving by conservatives. For example, Brooks (2007) claims that conservatives make 30 percent higher donations of money than liberals (even controlling for income), give more blood and even donate more of their time. Texas ranks 13th among the states in donations among those who itemize tax deductions, according to the Chronicle of Philanthropy (August 19, 2012), with donations of 5.1% of discretionary income. However, most studies using national surveys show little difference in giving by political orientation (Anderson, Mellor, and Milyo 2005; Margolis and Sances 2013), and most lab experiments show no difference in secular giving by party affiliation (Fowler and Kam 2007) or religiosity (Eckel and Grossman 2004).

3. INSTITUTIONAL LEGITIMACY AND PUBLIC GOODS GAMES: A LABORATORY EXPERIMENT ON THE DISTRIBUTION OF SANCTIONING POWER

3.1 Introduction

> "Many forms of Government have been tried and will be tried in this world of sin and woe. No one pretends that democracy is perfect or all-wise. Indeed, it has been said that democracy is the worst form of government except for all those other forms that have been tried from time to time." *Winston Churchill, Speech in the House of Commons (November, 1947)*

One of the main challenges for any society is to establish, motivate and sustain collective action through its institutions, while, at the same time, balancing individual vested interests that are not so rarely antagonistic. In the presence of *social dilemmas*, voluntary contributions to a public good can reflect the level of collective commitment, especially considering that the contributor faces an opportunity cost of the use of his resources (e.g., effort, time, money). The way complex and modern societies have managed to efficiently enforce collective action is through mechanisms of power delegation. To enhance collaborative efforts towards social goals, the *leader* has among his responsibilities and tools the use of the *carrot* or the *stick*, i.e., reward or punishment. Provided his decisions are consequential to the set of possible social outcomes, abundant research across disciplines concentrates on the analysis of the leader's characteristics and the influence of his decisions or policies (Cartwright, Gilletand van Vugt 2013, Brandts and Cooper 2007, Chaudhuri and Paichayontvijit 2010). Nevertheless, the effects of many pre-established institutional arrangements, such as the leader's selection mechanism, are less understood. Resent research in experimental and political sciences have turn its efforts towards this goal. This paper contributes to this research by testing the performance of exogenous versus endogenous arrangements of sanctioning power in a Public Goods environment. Under exogenous power-distribution the leader is randomly chosen by the experimenter (*Leviathan*), while, under the endogenous case, group members can choose their leader democratically (*Democracy*). Our results show that the Democracy outperforms the Leviathan in promoting collective action towards the socially efficient outcome. Given our parsimonious experimental environment, the observed leader's behavior in both arrangements, and controlling for several characteristics of participants, we argue that the mechanism through which the *Democracy* dominates is through authority *legitimacy*.

Social dilemmas, such as those found on Common Pool Resources and Public Goods Experiments and sanction institutions, go beyond the political arena. Firms and organizations face similar challenges to enforce cooperation. Firms need to set productivity goals that will reflect their economic performance and spend a significant share of their budget on monitoring systems and supervision to ensure their managers and work force commit to the same goals. Beyond the principal-agent problem and the design of incentive-compatible contracts that promote efficiency, there are instances where peer-supervision is desirable, and potentially reduces the risks and costs of unwanted outcomes, e.g. selecting a supervisor or team leader from the workers' pool (an *insider*) to enforce the principal's agenda. Think of the dilemma that any college's Dean faces when choosing a Department Head. Should he choose from within or outside the faculty?; which delegation strategy would result in higher commitment to educational goals? Public Goods, in theory, are not efficiently provided under voluntary contribution due to opportunities for strategic behavior where the dominant strategy is to *free ride*. Paradoxically, abundant experimental literature in laboratory and field studies (Ledyard 1995, Ostrom 1992) has shown that individuals consistently deviate from such pessimistic predictions. Nevertheless, cooperation tends to deteriorate towards the non-cooperative equilibrium over time. Hence, the study of institutional arrangements that enhance and sustain cooperation remains a desirable goal.

Beginning with the seminal experiment by Fehr and Gächter (2000), most related experiments study the effects of decentralized sanctioning institutions or peer-punishment, in which every participant has the power to punish or reward others within the group (see Chaudhuri 2010). An opposing argument, that resembles a *Hobbessian* view (1651), leans towards the centralized authority of the state as a more efficient public goods provider. Not much attention has been devoted to the study of centralized institutional arrangements. Recent efforts in this line of research include: Baldassari and Grossman (2011) (BG2011, hereafter), Andreoni and Gee (2012), Brandts et al. (2013), Kocher et al. (2013), Carpenter et al. (2012) and O'Gorman et al. (2009).[1] Apart from BG2012, none of the studies have focused exclusively on the effects of power-distribution in a punishment setting and their designs typically involve other treatments that simultaneously affect contribution outcomes (e.g leadership, communication).

This paper provides the first experimental evidence in the laboratory that isolates the effects of exogenous versus endogenous power-distribution mechanism in contributions for social dilemma environments. We look at an incentive mechanism to improve cooperation in social dilemma situations by looking at the contribution levels to a PGG under two different centralized institutional arrangements: endoge-

[1] A recent summary of related literature can be found in Van Lange, Rockenbach and Yamagishi 2014.

nous versus exogenous power distribution. After a first phase for a standard linear Voluntary Contribution Mechanism (VCM, henceforth) we introduce (unexpectedly) a sanctioning institution. The sanctioning power is centralized to one group member, e.g, the manager, for the rest of the session. In the exogenous mechanism (*Leviathan*) one participant is randomly selected from the group as a manager (leader/punisher) by the computer, while in the endogenous mechanism (*Democracy*), one participant is elected by a simple plurality voting rule among the group members. In the second phase, each period has two stages. The first stage consists of a VCM followed by a second stage where a tax/punishment mechanism is implemented. The manager has two decisions: whether leaving subjects payoff unchanged (not to punish) or to punish one (and only one) group member; and, if punishment is assigned, then choose who to punish, along with the intensity of the punishment. Similar to real world institutional schemes, a centralized punishment is costly for the society as a whole (e.g., bureaucracy and administrative costs), hence, punishment assigned is costly for all group members (including the manager) and is deducted from the punished subject, while unassigned punishment is returned equitably to each group member.

Similar to previous results in peer-punishment (decentralized) institutions (Fehr and Gächter 2000) we find that in a centralized arrangement the presence of punishment opportunities increases contribution levels. Contrary to theoretical predictions, the power delegation mechanism is not an innocuous feature of a Public Goods Game (PGG, henceforth). The *Democracy* treatment enhances cooperation levels over the *Leviathan* treatment, suggesting the former as a preferred dominant strategy to enhance collective action. Furthermore, considering that punishment is individually and socially costly, higher welfare levels (measured as net payoffs) are reached in the *Democracy* treatment. These findings provide important insights for organizations that aim to induce its goals in lower management levels by allowing the *legitimacy* of the authority to kick in, a feature disregarded in the classical *principal-agent* model, but greatly acknowledged by cooperative firms (firms where stakeholders share ownership, work and administrative responsibilities). Furthermore, on the political arena, it supports the importance of *electoral legitimacy* in political systems to promote collective action.

The remainder of the paper is organized as follows. Section II gives a brief overview of the related literature. In section III, we describe the experimental design and procedures in detail. Section IV discusses the different theoretical predictions and the working hypotheses. We report the experimental results in section V. Section VI concludes the paper. Instructions for the experiment are included in the document's Appendix.

3.2 Literature review

Our paper contributes to several literatures. In the experimental literature, a growing number of studies have concentrated on the effects of sanctioning opportunities in social dilemmas since the seminal study by Fehr and Gächter (2000) They examine the effectiveness of monetary punishment in promoting contributions to a PGG under a decentralized peer-punishment scheme. Their experimental setting is enriched by the fact that punishment is socially costly. Hence, punishment use is not a credible threat to deter free riding behavior; and, assuming a purely rational agent, the subgame perfect Nash equilibrium (SPNE) in this game predicts zero contributions and zero punishment. They show that the existence of punishment opportunities increase significantly the average contribution levels; however, cooperation cannot be maintained if the threat of punishment, or willingness to punish, is not credible. At the same time punishment can be ineffective in the presence of reputation formation and revenge opportunities (mainly from free riders). Nikiforakis

(2008), for example, introduces opportunities for anti-punishment into a PGG and shows that the willingness to punish free riders decreases significantly, thereby further undermining the effectiveness of punishment in promoting cooperation. Our design builds on the importance of sanctioning institutions in cooperation and explores a sanctioning scheme with neither reputation nor revenge opportunities.

This paper also relates to the literature on democratic participation in social dilemmas. The importance of democratic institutions has been extensively discussed in long-run development (e.g., North 1981, LaPorta et al. 1998, Acemoglu et al. 2005). However, the experimental study of democratic institutions is relatively new in social dilemma games, in most of which the institutions' incentives are exogenously imposed by experimenters. A natural development is to investigate endogenously chosen institutions. In general, when given the chance to vote for the institutional environment, subjects seldom vote for sanctioning institutions and prefer rewarding arrangements, although not the most effective; and if a sanction environment is already in place, they tend to restrict punishment levels and often punish the free riders (Decker et al. 2003, Botelho et al. 2005, Ertan et al. 2009, Sutter et al. 2010). Another different approach to endogenous institutions is to allow voting over the set up and intensity of the incentives. Putterman et al. (2011) introduce a dynamic voting scheme where subjects choose whether to punish contributions over the private or the public good, as well as the parameters of intensity. They show that most subjects (89%) are in favor of penalizing contributions to the private account and there is learning behavior over optimal sanctioning schemes. Nevertheless, subjects might also prefer a sanction-free environment for many reasons that include the rational recognition of the welfare costs involved in punishment. Also, there is still ambiguity about self-governed group preferences. Most of the previous studies allow decentralized punishment (peer-to-peer punishment), which can result in *excess* punishment and higher short-run inefficiencies. We reduce the potential *misuse* of punishment by allowing the punishment of one (and only one) of the group members, without restricting the punishment intensity or who receives it.

A vast body of literature argues in favor of direct effects of democracy on economic performance and policy selection.[2] Positive effects are found independently of the information and sophistication of the subjects, for example, Dal Bó et al. (2010) show that the effects on cooperation on a prisoner's dilemma game, by modifying the payoffs, is greater when the policy is chosen endogenously; further, these effects are independent of the sophistication of subjects and information provided about the voting stage. We minimize possible leaders' confounders by setting a communication-free environment. The only available information is group members' contributions, which is provided equally in both treatments: endogenous and exogenous; and it is presented to the subjects without identifiers to avoid reputation formation. Furthermore, our experimental design avoids potential biases due to institutional endogeneity and selection bias, i.e. more cooperative subjects are more prone to democracy, or policies (choices) within a democratic institution promote higher cooperation. We do not observe significant differences in the manager's characteristics between mechanisms although democratically elected managers tend to act more "responsible" and show slightly higher cooperation levels.

From an evolutionary perspective, modern societies overcome these caveats of social dilemmas through a centralized punishment power or delegation of power.[3] Recent experimental literature on PGG concentrates on this institutional arrange-

[2]For an excellent review and summary see Dal Bó 2014.

[3]Note that centralization of punishment does not guaranty higher efficiency. Dictators tend to overuse and abuse their power towards non altruistic goals. In experimental studies there is certain ambiguity in the efficiency of centralized versus decentralized punishment arrangements. O'Gorman (2009) and Carpenter et al. (2012) find contradicting results in public goods experiments, however differences in the designs might explain the results. For a discussion see Van Lange et al. 2014.

ment, in particular on third-party punishment institutions. Andreoni and Gee (2012) show that a *Paladin* ("a gun for hire") can effectively deter free riding behavior and improve welfare in PGG when compared to peer-to-peer punishment. Guillen et al. (2006) show how centralized sanctioning institutions have a significant *educational* effect in promoting cooperation. Nevertheless, there is no conclusive result when it comes to the effects of voting in a centralized institutional environment. Since human leaders can effectively promote social norms and cooperation by promoting an agenda (Levy et al. 2011), the effects of endogenous arrangements are often clouded by the fact that outcomes are not independent of the leader's characteristics. Elected leaders may have contradicting incentives to those of their constituencies and make counterproductive decisions, while pro-social leaders can affect outcomes via their own contribution, group status or their signal credibility (Brandts and Cooper 2007, Hamman et al. 2011, Kocher et al. 2013, Brandts et al. 2013, Eckel et al. 2007). The experimental design proposed in this paper allows us to isolate the results from those in the leadership literature by looking at the direct effect of the manager selection mechanism on cooperation. We come back to this point on the results section.

A study that is closest in nature to ours, although in a *lab-in-the field* setting, is the study by Baldassari and Grosman (2012). They study the legitimacy of the authority on cooperation in a PGG by comparing random versus a secret ballot selection of a third-party punisher in different producer organizations in rural Uganda. They find that subjects contributed to the public account about 9% more in the endogenous mechanism. This evidence is relevant considering that decisions in natural environments summarize a series of unobserved factors (e.g., iterative interaction, reputation, identity, culture). Nevertheless, in our perspective, this same reason blurs the possible causal inference. Further, third-party punishment might bias results provided leaders payoff function does not depend on the group results. To our knowledge, no other laboratory study has concentrated exclusively on the effects on voting on centralized punishment institutions, and we believe that the evidence in our design, in a controlled environment, overcomes several caveats from confounding and contaminating effects originated in simultaneously imposed treatments in previous studies.

3.3 Experimental design and procedures

The experimental design is based on the canonical VCM for a standard linear Public Goods Game. Table 1 shows the structure and timeline of the experiment. We test two treatments, each in two phases to analyze between and within-subjects mean contribution differences.

Table 3.1: Experimental timeline description

Treatment	Phase 1 ($t \in [1, 10]$)	Manager Selection	Phase 2 ($t \in [11, 20]$)	Participants	Sessions
Leviathan	VCM	Random	Punishment	75	3
Democracy	VCM	Voting	Punishment	80	3
Total				155	6

3.3.1 The voluntary contribution mechanism

Let $I = \{1, 2, ..., n\}$ denote a group of n subjects who participate in a Public Goods Game repeated T periods. In each period $t \in T$, each participant $i \in I$ receives an endowment w, which is common to all participants (hence we drop the index) and will be allocated to either a private account or a public account. We denote the contribution of individual i into the public account in period t as c_{it} which must satisfy $0 \leq c_{it} \leq w$. Let C_t denote the total sum of all the group members' contributions in period t, i.e. $C_t = \sum_{j=1}^{n} c_{jt}$. Hence, the monetary payoff of individual i in period t is given by:

$$\pi_{it} = w - c_{it} + \alpha C_t \tag{3.1}$$

α is the Marginal Per Capita Return (MPCR) of the contribution, which range satisfies $0 < \alpha < 1 < n\alpha$. Full contribution to the public good is socially efficient if $n\alpha > 1$. However, the *social dilemma* appears due to the domain of the MPCR, one unit contributed to the public good implies a return of $\alpha < 1$ and the best strategy for a self-interested rational subject is to free ride, i.e. invest their endowment on their private accounts.

The parameterization is as follows. Participants are randomly divided into groups of $n = 5$ members that remain fixed ("partner matching") throughout the whole experiment. We choose this number to minimize the possibility of ties in the voting mechanism. Nobody knows which other participants are in their group, and nobody is informed of the group composition. Each session includes two phases under different regimes. The first phase is a VCM standard Public Goods Game lasting $T = 10$ periods. Every period, each participant receives an endowment of $w = 20$ experimental currency units (ECUs). Every participant decides how to allocate his endowment to a "group account" or a "private account." The total amount in the group account is doubled and divided equally among all group members, hence $\alpha = 0.4$. Participants observe how many points fellow group members contributed to the group account at the end of each round. To preclude reputation effects, participants are provided with information about the other group member contributions and payoffs in random order. Subjects receive the same amount/type of information, hence information effects become neutral in this setting and differences can be attributed to the treatment mechanism (Dal Bó et al. 2010).

3.3.2 The punishment environment

The second phase comprises the presence of punishment opportunities. It lasts for another 10 periods, and each period is composed of two stages. During the first stage subjects face the standard VCM with the same parameters previously described. In the second stage, one subject, the manager (punisher)[4] in each group, decides to impose punishment on his fellow group members. To do so, an amount of tax ($x = 4$) is automatically collected from each group member and put into a *management account* in each period. After the *manager* acknowledges the tax, he has two decisions at hand; on the *extensive margin* he has to decide whether to punish any of his fellows or not, i.e. to discipline or forgive observed behavior; while on the *intensive margin* he has to decide whom to punish and the size of the punishment, by using the 20 points available each period from the management account. The *manager* is free to leave all group members' earnings unchanged (i.e. not to punish), in which case the collected tax is returned to every subject; if partial punishment is assigned, the remaining points are returned equally to each member. Decisions faced by the punisher are simpler than in previous schemes in the sense that he has to choose only one group member. The decision of whether to punish a group member and whom to punish, can reveal some additional information about the punisher's preferences. We return to this point in the results section.

We assume a linear 1:1 cost function.[5] At the end of each period, the punisher's

[4]During the experiment, we subscribe to a neutral language and avoid the use of words such as "punisher", "punishment," and "tax;" we choose "manager", "reduction," and "points collected" instead.

[5]The intensity of reward and punishment varies in different experiments. For instance, Fehr and Gächter (2000) use a strictly increasing and convex cost function. In Sefton et al. (2007), the ratio of cost for punisher and target is also 1:1. Nikiforakis (2008) provides the strongest punishment, one unit of which cost target three units. Sutter et al. (2010) discussed the effect of intensity. They compared the effectiveness with or without leverage, i.e., the ratio of 1:1 or 1:3. Or see the

decisions (punished subject and intensity) are reported on each member's terminal.

A general expression for subject i's payoff function in period t, considering within group unconstrained punishment, can be written as follows:

$$\pi_{it} = \underbrace{w - c_{it} + \alpha C_t}_{\text{VCM}} - \overbrace{x - p_{it} + \frac{1}{n}(nx - p_{jt})}^{\text{tax/punishment mechanism}} \tag{3.2}$$

where p_{it} is the punishment imposed on subject i in period t.

This expression is useful to analyze different efficiency levels of the tax/punishment mechanism;[6] nevertheless, to simplify the intuition of the environment and restrict the punishment assignment to only one group member (j) and minimize punishment inefficiencies, the payoff function for individual i becomes:

$$\pi_{it} = \begin{cases} 20 - c_{it} + 0.4C_t & \text{, if no punishment is assigned (VCM)} \\ 20 - c_{it} + 0.4C_t - (0.2p_j) & \text{, if punishment assigned to } j \text{, or } j \neq i \\ 20 - c_{it} + 0.4C_t - 1.2p_i & \text{, if punishment assigned to } i \text{, or } i = j \end{cases} \tag{3.3}$$

Two aspects of this expression are worth noticing. Costs incurred in the tax / punishment mechanism are binding not only on the group members, but also on the leader's payoff. Hence, this environment reproduces the equilibrium incentives from Fehr and Gächter (2000) where the best strategy is to free ride ($c_{it} = 0$) and not to punish ($p_j = 0$); while, at the same time, acknowledges a natural feature

discussion in the survey by Casari (2005) for the design. Our design is similar with the leverage case, since the cost of punishment is equally shared by all the members.

[6]Possible extensions of this analysis can involved specific modifications of this simple environment; for example, changing the tax parameter we can analyze the deterrent influence of potential punishment. Also, punishment efficiency in this environment can be simply assessed by using different cost functions, something that has been analyzed before.

of certain organizations where costs of any policy implemented are shared among every member.[7] A second aspect is that there is not a *bankruptcy rule*. Bankruptcy ($\pi_{it} < 0$) is possible if, in a particular period t, a high contributing member (not the manager) is in a group composed mostly by free riders and receives *exemplary* (high) punishment. We considered this to be a rather extreme case and pursue the present mechanism for its simplicity, without further restrictions.[8]

3.3.3 The treatments

The PGG with punishment is played under two different arrangements for the distribution of power: exogenous versus endogenous.

In the exogenous power distribution (XPD), the *Leviathan*, one participant in each group is randomly selected as a punisher (manager) by the computer.[9]

The endogenous power distribution (NPD), the *Democracy*, is based on the democratic election of a leader (punisher/manager) by a simple *plurality voting* rule. The computer breaks the tie in case it rises (i.e., 2:2:1).[10] Everyone can submit one vote

[7] BG2011 implemented a third party punisher who plays an external monitoring role. His payoff function is different from anyone else in the group in the sense that it does not depend on the group contributions (he does not contribute to the PGG) but rather on his decisions over punishment assignment. We think this potentially generates different incentives for the manager. Other regarding preferences become a confounder for self maximizing behavior in third party punishment. For example, the manager's perception over levels of inequality aversion might be stronger when being an outside observer than when facing the cost of any punishment decisions that affects his final payoff.

[8] Throughout our experiment we did not observe bankruptcy, i.e. no one's payoff was negative in any period and punishment was typically assigned to the free riders. Overall, punishment used fluctuates in the whole range, however, 93% of punishment assigned is equal or lower than 10 Points (mean= 2.23, median= 0). Different tax/punishment calibration can be tested in replications, and we encourage this practice.

[9] The randomization procedure can raise some concerns, in particular, subjects may not trust the computerized randomization procedure. The extent to which this might affect our results is uncertain; however, it is common practice in current experimental settings and we subscribe to this practice.

[10] We use a *plurality voting rule* that does not require further selection stages. In the experimental results we did not observed the case in which each subject receives one vote (although it is a latent possibility). We control the differences in votes the managers received in the subsequent econometric analysis.

but is not allowed to vote for oneself. Subjects cast their votes simultaneously and anonymously. The only signal voters have at hand come from the historical contribution record of each group member on the first phase.[11] They were informed of this being the only chance for them to be able to link the information about contributions to a fellow group member.

Finally, To avoid strategic contributions between phases, subjects were not informed about details of the second phase until the end of the first one. They only know there will be some changes proposed during the development of the experiment.

Implementation: The experiment was conducted at the Economics Research Laboratory (ERL), at Shenzhen University (SZU), in China. We used "z-Tree" (Fischbacher, 2007) for the computer programing and interface. The sample used includes 155 subjects, all are students from several majors at SZU. The sample includes 47% of women; 39% were economics undergraduate students. Participants had never participated in a Public Goods experiment. Each session lasted about 90 minutes. Before entering the experimental laboratory, participants were told that they would receive a show-up fee of 10 RMB (Chinese Yuan)[12] which was paid in cash upon completion of the session. As usual, they were also informed that they would have other opportunities for extra payoffs based on their performance. They were not provided any other experimental details before hand.

After being randomly seated at separate computer terminals, subjects received written instructions that were also read aloud by the experimenter. Experimental in-

[11] The information includes total and per period contributions. Some concerns may raise due to the available information. The information independence comes from the fact that subjects observe contributions in each period in both treatments, hence any information about group composition is equally transparent. One fair criticism that needs further scrutiny is the fact that authority/institutional legitimacy on the endogenous versus exogenous power distribution can be confounded with trust in better chosen leaders, e.g., never a free rider, provided they do select leaders based on previous contributions. We come back to this point in the leadership section.

[12] At the time of the experiments the exchange rate was 6.10 RMB per USD 1.00.

structions, included in the Appendix A, use neutral language, avoiding terms such as "public good," "contribution," "punishment," "leader," and "democracy." To ensure complete understanding by all subjects, a set of basic test questions were presented and had to be correctly answered before each phase began. At the end of the experiment subjects filled out a post-experimental questionnaire, including demographic questions, as well as attitudes about experimental procedures and payoff.

3.4 Theoretical predictions

The basic behavioral hypothesis from classical non-cooperative game theory, built on the assumption of a self-interested rational subject, predicts that "free riding" behavior is a dominant strategy in a sanction-free VCM. Further, introducing punishment opportunities in the PGG does not affect the equilibrium prediction of pure free riding behavior. In the second stage of each period, due to the punishment cost, it is not a "rational" strategy for the leaders to impose any punishment on a fellow group member. Thus the threat of punishment is not credible and the subgame perfect Nash equilibrium (SPNE) remains the zero contribution for each participant. This predictions are unconditional on any mechanism of power distribution (leader's selection procedure).

Nevertheless, empirical and experimental evidence have shown consistently the failure of such strict predictions of standard non-cooperative game theory (Ledyard 1995, Chaudhury 2010) and the research agenda has expanded to explore alternative explanations for observed contributions. Social preferences and other regarding preferences hypothesis have greatly enriched the behavioral aspect of the theory.[13] The social-preference-based story suggests that leaders will punish free riders due

[13] Social preference includes altruism and warm glow in Andreoni (1990) and Andreoni (1995), inequality aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), positive and negative reciprocity (Charness and Rabin 2000, Dufwenberg and Kirchsteiger 2004), fairness (Rabin 1993, Roth 1995), among others.

to considerations that go beyond individual strategic behavior.[14] If the punishment observed is *altruistic* (Fehr and Gächter 2000) in the sense that it pursues social improvements in future interaction, then it is a "rational" strategy in repeated PGG and we should observe improvements in cooperation as the game develops. A driving force behind the existence of altruistic punishment are the emotions (negative emotions) triggered by free riding behavior (Fehr and Gächter 2002). Consistent with these hypothesis, our experimental subjects contribute around 40% of their endowment in the initial period of the VCM and contributions decrease to around 20% in the final period of the first stage. However, once punishment is at play contributions rise to around 55% of the endowment in the first period and the pattern of contributions is consistently increasing throughout the second phase, independently of the leader's selection treatment. The final stage represents a one shot game where punishment cannot be enforced anymore, cooperation deceases but returns to around the same initial level. We also observe that the most likely punished member of the group is the free rider.

As in previous studies, punishment can also come from inequality aversion arguments. Furthermore, contributions in a punishment environment can come from aversion to be the *underdog*. If altruistic punishment is at play and leaders impose the average contribution as a social norm, the probability to receive punishment for the free rider is high in our design (the second to last free rider does not receive any punishment). Hence a "race to the top" incentive mechanism has also a role in our results. No one wants to be the only one punished.

However, different from peer-to-peer punishment incentives, in a centralized power distribution, emotions and altruistic punishment should run through the leader who

[14] A large body of literature discusses the driving forces to punish free riding in PGG. Reciprocity and fairness (Falk et al. 2005; Carpenter 2007; Casari and Luini 2012); justice and equity (Kahneman et al 1986), among others.

is the one called to impose a social norm, hence we focus on the relationship between the leader's selection mechanism and cooperation levels of all group members. Theoretical and empirical arguments in social psychology, sociology and political sciences, support more optimistic outcomes in the endogenous power-distribution institution. Institution or authority *legitimacy*, has been argued, plays an important role in social choice (Weber 1948, Lipset 1959). However it remains a category vaguely defined (Brandts et al. 2014) and difficult to isolate in experimental settings. The legitimacy of an authority or an institution may not come from an specific political organization but from the group's perception over the way its preferences (values) converge with the ones of the political institution in place (Lipset 1959, p.86-87). In other words, institution or authority legitimacy can arise under heterogeneous power delegation mechanisms.

In view of these arguments we hypothesize the following:

Hypothesis 1: The contributions' performance among institutional arrangements follows the following order: NPD>XPD>VCM.

As mentioned, the fact that sanctioning institutions improve cooperation in the public goods provision justifies the last inequality. Beyond this well established finding, we argue that *legitimacy* through the manager's selection mechanism has a direct a role in enhancing collective action. The channel through which legitimacy plays a role in the effectiveness of the institutional environment is subject to debate. The legitimacy of the institution entails an objective evaluation over the appropriateness of the political structure to pursue specific social goals, as well as an affective assessment that members of a group perform, consciously or unconsciously, over a range of beliefs and perceptions over what those social goals are (Lipset 1959). In the same line of argument, democratic institutions can affect economic performance through two main channels: institutional incentives (i.e. institution's performance, leader's performance and path dependency) or individual behavior (i.e. influencing pro-social behavior). Both channels coexist in a "real world" setting and, although we do not disregard the effects of the former we favor and concentrate on the latter and analyze the extent to which this channel affects the economic performance. Incentives towards collective action go beyond "rational" arguments and are routed deeply in other aspects of intrinsic human behavior.

We predict the *Democracy* (endogenous) treatment will be more effective in promoting contributions than the *Leviathan* (exogenous). If there is evidence of changes in pro-social individual behavior on the institutional performance, it must be the case that this argument holds despite of the manager's contribution behavior or punishment frequency. In other words, we should expect most group members, regardless of their hierarchic position within the group, to contribute more in the democracy.

Hypothesis 2: There is heterogeneity in the effectiveness of punishment between treatments. NPD should be higher than in XPD at initial periods since elected managers may feel more responsible to discipline the free riders than their exogenously selected counterparts. Even though, the dynamics of punishment should show a decreasing pattern in both treatments, on latter periods it should be the case that the NPD falls below the XPD, provided the institutional effectiveness converges faster to the social norm in the NPD and higher cooperation levels are in place.

Hypothesis 3: Higher efficiency, measured as net earnings (payoffs), is reached in the endogenous power distribution, i.e. the net interaction of the previous two hypothesis results in better welfare outcomes in the NPD versus the XPD.

3.5 Results

The main goal of this study is to test the effect of endogenous versus exogenous power distribution, the *Democracy* versus the *Leviathan* treatment, on individual contributions in the PGG; further, we argue that the only mechanism behind this behavior is *institutional legitimacy*. The results are discussed as follows. The performance across institutional regimes is analyzed in Section 5.1. Section 5.2 discusses the econometric results. Section 5.3 analyses whether or not there is a leadership effect, i.e., test whether the performance of the leader (punisher) differs across regimes.

3.5.1 General results

Table 2 (Panel A) summarizes the overall average of contributions, punishment behavior and profits under the three regimes. The lowest average contribution levels are observed in the VCM in the baseline for both treatments; subjects contribute on average 7.08 out of 20 ECU in each period (35% of the endowment) By comparing the between treatments baseline we do not find they are statistically different (p=0.18 for a Man-Whitney *U* test). As in previous studies in the Fehr and Gächter (2000) flavor, our baseline also reproduces evidence of deviation from the standard theoretical prediction of free riding behavior, in fact contribution patterns match closely.

Subjects face the punishment environment in period 11, after the groups where assigned a leader by one of the two proposed mechanisms. Contributions in both treatments increase when punishment opportunities are at hand. The *Leviathan* treatment increases contributions to around 58% of the original endowment (11.60 ECU), while the *Democracy* treatment improves in average to 67.5% of endowment, to 13.50 ECU, almost 91% higher than in the baseline VCM. It is worth noticing that the existence of punishment shapes the initial behavior, without individuals having any previous experience with their leader's influence in the game, contributions start well over the baseline initial point. The Leviathan treatment's average contribution begins at about half of the endowment (10 ECU) and ascends slightly over time up to period 18 where it reaches around 13 ECU before declining again back at 10 ECU.

Table 3.2: Average performance comparison

	Panel A: Contribution (Points)		
Treatments	Baseline: VCM	Punishment	Punishment-VCM
Leviathan	6.920 (0.259)	11.604 (0.248)	4.684 (0.359)
	($N = 750$)	($N = 750$)	($p = 0.000$)
Democracy	7.230 (0.229)	13.049 (0.237)	5.819 (0.330)
	($N = 800$)	($N = 800$)	($p = 0.000$)
Democracy-Leviathan	0.310 (0.345)	1.445 (0.343)	**DD:** 1.135 (0.390)
	($p = 0.182$)	($p = 0.000$)	($p = 0.000$)

	Panel B: Punishment (Points)		
	Leviathan	Democracy	Democracy-Leviathan
	4.207 (0.220)	4.706 (0.244)	0.500 (0.165)
	($N = 750$)	($N = 800$)	($p = 0.764$)

	Panel C: Profit (Points)		
Treatments	Baseline: VCM	Punishment	Punishment-VCM
Leviathan	26.920 (0.259)	29.921 (0.262)	3.001 (0.369)
	($N = 750$)	($N = 750$)	($p = 0.000$)
Democracy	27.230 (0.229)	31.164 (0.268)	3.934 (0.353)
	($N = 800$)	($N = 800$)	($p = 0.000$)
Democracy-Leviathan	0.310 (0.345)	1.242 (0.376)	**DD:** 0.932 (0.427)
	($p = 0.140$)	($p = 0.000$)	($p = 0.029$)

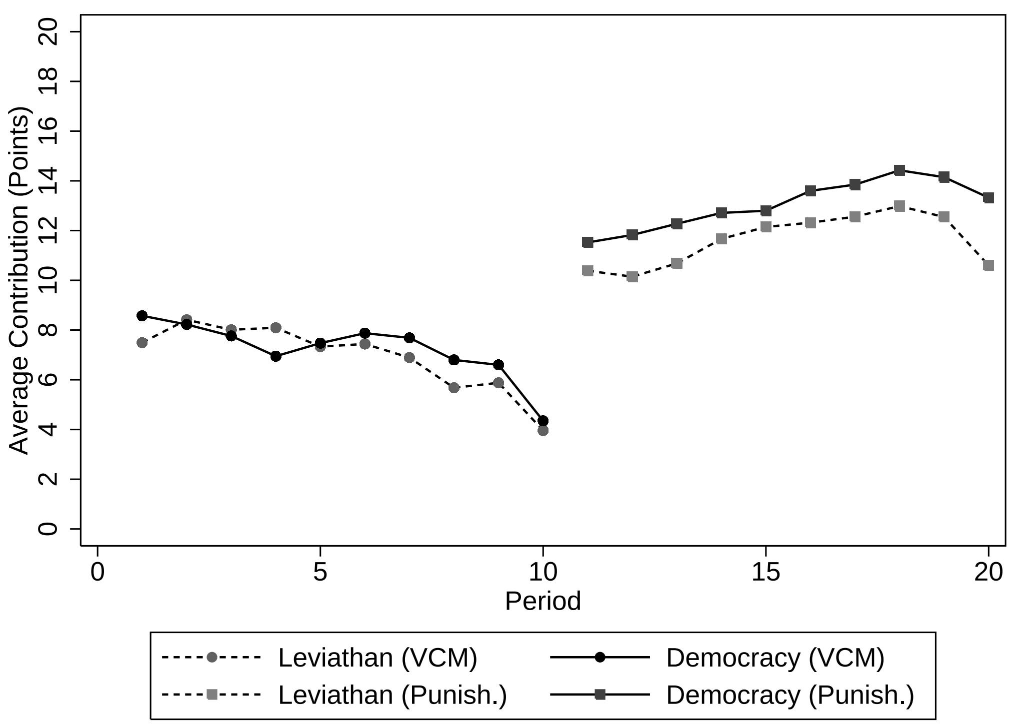
Standard errors are in parentheses. p values are reported for a Mann-Whitney U tests. Two-sided t-tests report similar results.

The Democracy treatment's average contribution begins at a higher level (about 12 ECU) and also ascends to its highest level (slightly higher than 14 ECU) in period 18 and declines to 13 ECU at last. Figure 1 displays the time trend of contributions by institutional arrangement.

In the between treatments analysis, we found evidence that shows how collective action is improved by allowing group members chose their leader (punisher). Contrary to a SPNE solution concept, subjects react contributing significantly higher when they choose their leader as opposed to when he is imposed. According with our first hypothesis, the difference between the Democracy and the Leviathan is positive; and, unconditional on other factors, it ascends to around 1.14 ECU (5.7% of the endowment). The result is highly significant for nonparametric Mann-Whitney U test (p=0.00).[15]

Result 1. *Punishment opportunities in both treatments, endogenous (Democracy) and exogenous (Leviathan) power distribution, trigger significantly higher contributions than the standard VCM. Subjects contribute at consistently higher levels in the Democracy treatment than in the Leviathan. (H1: $NPD > XPD > VCM$)*

Figure 3.1: Average contributions by treatment



Punishment represents a private and a social cost, hence its use can potentially

[15]t-Tests report similar results.

lead to overall economic inefficiencies (welfare losses). Table 2 (Panel B) also reports the general results of the experiments in these two aspects. The average punishment ascends to 4.21 ECU and 4.71 ECU in the Leviathan and Democracy, respectively. However, this difference is statistically insignificant (Mann-Whitney U test, $p =$ 0.76), i.e., there is no difference between punishment behavior across treatments. Nevertheless, it is worth noting the difference in the declining pattern. While both treatments decrease parallel for most of the phase, the Leviathan treatment shows some ambiguous punishment, its levels increase sharply in the final two periods to up to 6 ECU. In the Democracy treatment punishment continues its declining trend to almost zero.

Nearly 80% of the leaders chose to use punishment in the initial period for the Democracy treatment, while only 60% of leaders do it in the Leviathan. Differences in punishment trends are negligible between treatments; however, the likelihood of punishment is almost zero in the Democracy for the last period, but about half of punishers continue to punish in the Leviathan treatment. Consistent with our second hypothesis, even though average contributions in the Democracy treatment are higher (by about 1.1 ECU units), on average, elected leaders react more severely to free riding behavior, using around 3 punishment points more than the punishment applied by randomly-chosen leaders. This partially supports the second hypothesis. In the same line of argument, punishment in the Democracy treatment is more consistent in its decreasing pattern, while average punishment has a U-shaped pattern in the Leviathan. We found these results rather striking considering that contributions are mostly parallel between treatments (except for the last period). There is no apparent reason for a randomly chosen leader to change punishment behavior dramatically in the last periods, although the very last period might represent some form of altruistic punishment, specially considering there is no further stages to the game. Figure 2 displays detailed information about temporal patterns of intensity and frequency of punishments.

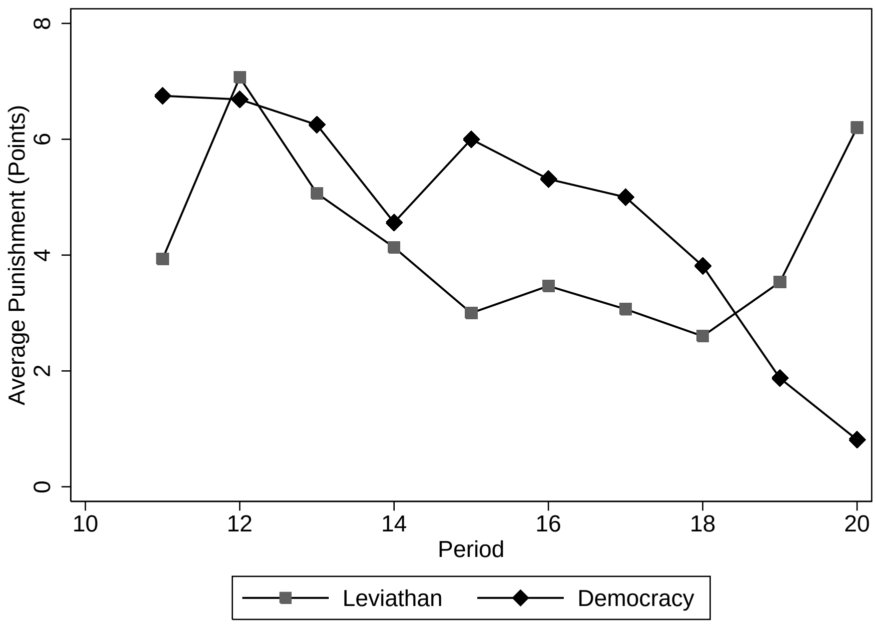
Result 2. *Frequency and average intensity of punishment are not significantly different between the endogenous and exogenous distribution of power, although we find heterogeneous punishment patterns. In particular, leaders democratically elected show higher levels of altruistic punishment. Randomly chose leaders react somewhat arbitrarily and show an average intensity of a U-shaped form.*

Finally, when it comes to economic efficiency, measured simply as the profit in each period, i.e. the period earnings minus the punishment, [16] we naturally find very similar patterns as the contributions (see Figure 3a). The average earnings in the VCM (first phase) is 27.08 ECU, while the average payoff in the Leviathan and Democracy is higher (and statistically different from each other, p=0.029). Average payoff reaches 29.92 ECU and 31.16 ECU, respectively. Payoffs are always higher in the Democracy treatment except for period 15; these differences accentuate more if we exclude the punished subjects (mostly free riders, see Figure 3b). The punishment differences between the erratic behavior in the Leviathan treatment versus a more consistent declining trend in the Democracy treatment affect the payoff results shown. Payoffs are the net effect between contribution behavior and punishment applied (or not used), hence, as argued in our hypotheses, we find evidence that centralized punishment institutions enhance social efficiency. Furthermore, higher welfare levels are reached in the endogenous power distribution (more compliance with social norms and less punishment needed). The last result confirms our third hypothesis.

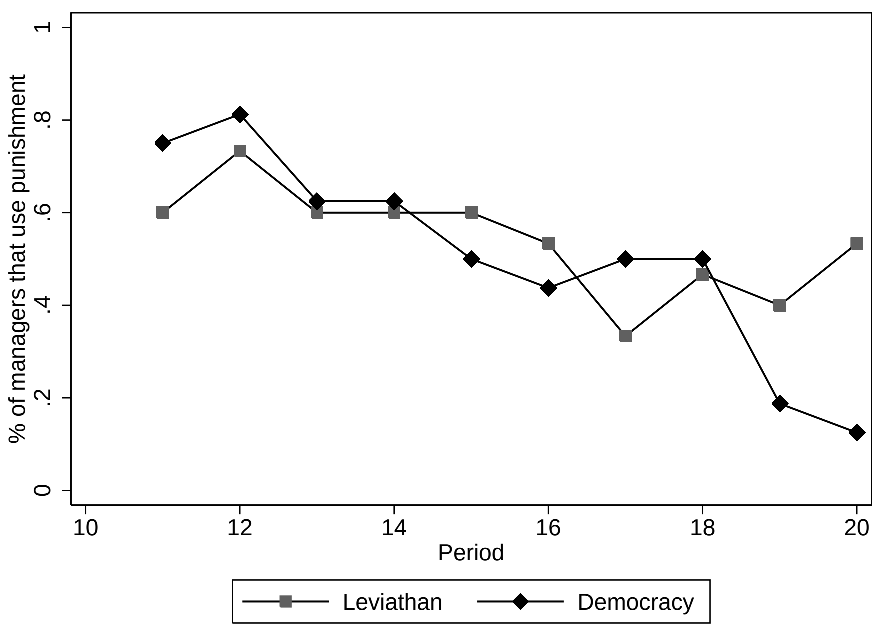
[16] Efficiency in this setting can be measured also in relation with potential earnings in each period (Eckel et al. 2010). Conclusions are the same.

Figure 3.2: Punishment behavior

(a) Average punishment assigned



(b) Punishment frequency

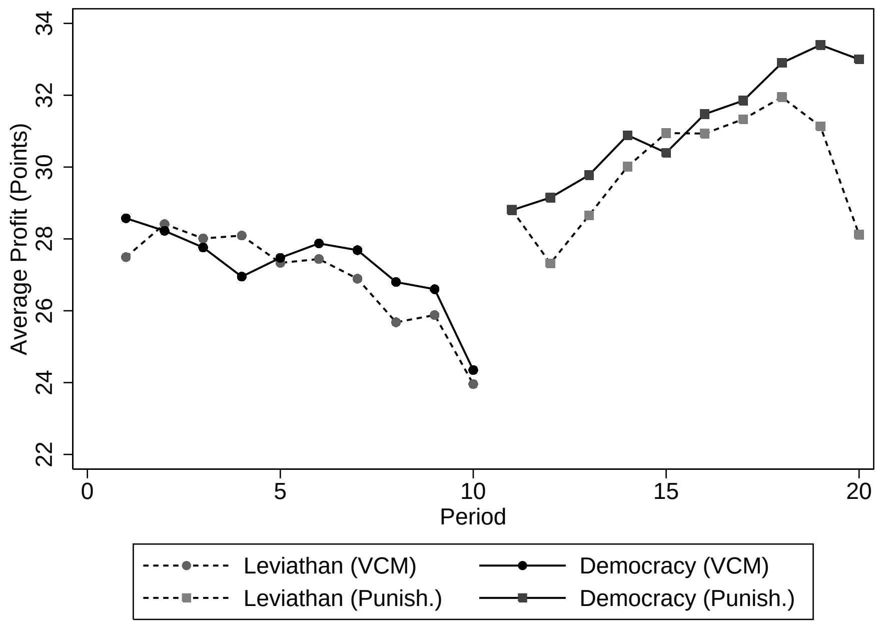


Result 3. *Higher efficiency is reached in the endogenous distribution of power (Democracy treatment); net payoff is significantly higher than in the Leviathan treatment.*

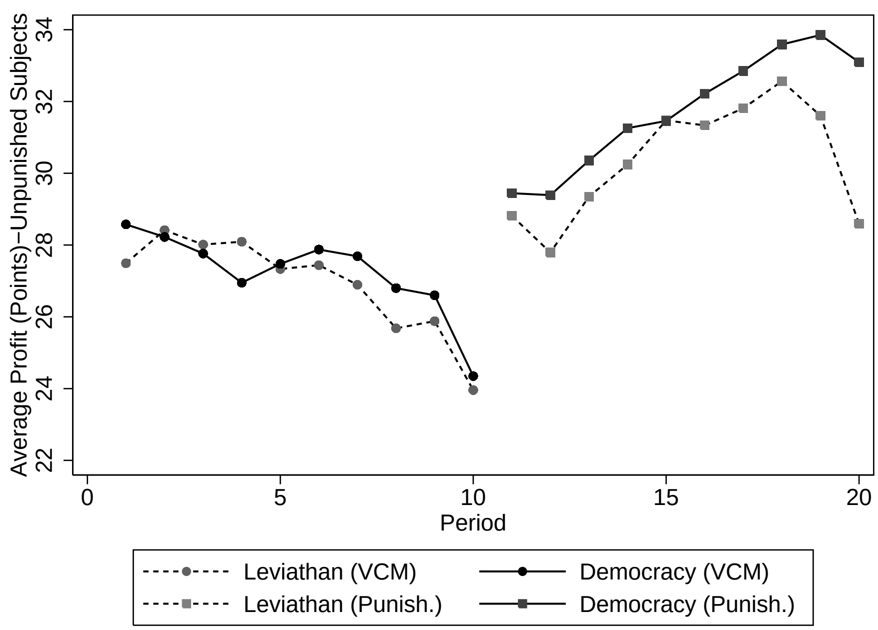
Fehr and Gächter (2000) find that punishment initially causes a payoff loss relative to the VCM, both in the strangers and the partners treatment. Yet there are relative payoff-gains in the last two periods out of ten in the strangers treatment, whereas the relative payoff-gains are positive from period 4 to the end in partners treatment. The relative payoff-gain in the final period is about 20% in the partners treatment and 10% in the strangers treatment. In punishment regimes with high (3) and low punishment leverage (1). Sutter et al. (2010) finds no significant gains relative to the standard VCM. Hence regarding the average overall payoff gains in the final period, the Leviathan arrangement is *at least* no worse than their partners treatment, whereas the Democracy clearly outperforms the others. These differences, we argue, come from the decentralized punishment institution where everyone can punish any other group member, leading to higher inefficiencies (less effectiveness, or over punishment for a similar effect). In our design only one subject can be punished in each period, the likelihood of being effectively punished is lower conditional on being close to the social norm (group's average contribution). At the same time, the free riding behavior becomes more salient for the punisher, hence there are incentives to push contributions upwards to avoid being the underdog (a "race to the top" effect) This two forces at play show that behavioral results are relevant even in such difficult institutional conditions, which strengthens our argument.

Figure 3.3: Subjects' payoff

(a) Overall profits



(b) Profits excluding free riders (punished subjects)



3.5.2 Econometric results

We perform a formal econometric analysis to support the results shown, and analyze the determinants of contributions (C_{igt}) in each institutional arrangement conditional on other factors (observed and unobserved) that might affect subjects performance throughout the experiment.

A natural extended regression specification to analyze the results is the following:

$$C_{igt} = \gamma_1(D{*}P)_{it}+\gamma_2 Democracy(D)_i+\gamma_3 Punishment(P)_t+Z'_{ig}\Phi+X'_i\mathit{\Gamma}+\alpha_g+\lambda_t+\varepsilon_{igt} \tag{3.4}$$

where the subindex i represent the experimental subject, g the corresponding group and t the time period.

The coefficient of interest that captures the average treatment effect of the endogenous power distribution is $\hat{\gamma_1}$. Similar to a diff-in-diff approach, the regressors of interest, the product of the Democracy for those subjects in this treatment and the Punishment institution in the second phase, represents the difference between the Democracy and the Leviathan treatment.

To capture the deterrence effect of punishment and compliance to the social norm, the model includes Φ, which contains $\overline{C}_{-ig,t-1}$, the lag of the average contribution of other fellow members in group g, $P_{g,t-1}$, the lag of the general amount of punishment executed by the punisher within the group. $P_{i,t-1}$ represents the punishment received by individual i, if he was punished in the previous period, and the trend t within each phase.

Finally, X_i is a vector of individual demographic controls that include: gender, major, math score in the National Entry Exam,[17] year in college, whether or not

[17]National Higher Education Entrance Examination-NCEE is an academic test held annually for

they received an scholarship, urban or rural area, family income, ethnicity, whether they have a job, party membership, whether or not a single child in family, and trends within each phase.[18] As usually assumed in experimental analysis, due to randomization (and the Law of Large Numbers), $\varepsilon_{igt} \sim NIID[0, \sigma_\varepsilon^2]$, nevertheless, we tested several model specifications that include individual,[19] group (α) and time (λ) fixed effects. As corresponds, we dropped the time invariant controls for comparison on the results shown.

Our identification strategy relies on random assignment of the treatment, in other words, there are no systematic differences between the control and treatment groups (i.e., Leviathan and Democracy in our experiment) that would have caused the outcome levels to be different in the absence of treatment; furthermore, if any differences arise they are observed and we can control for them.[20]

Several specifications of equation (4) are presented in Table 3. Column 1-2 show the results from GLS regression with random effects.[21] Columns 3 show the results for an individual fixed effects model.

Similar to the raw (unconditional) results shown in Table 2, the coefficient of interest has exactly the same magnitude ($\gamma_1 = 1.135$) and is strongly significant. Also, consistent with previous findings (see Gürerk et al. 2006), just the presence of punishment opportunities increase the contribution levels, in average, punishment opportunities increase contributions by around 4.7 ECU, when unconditional on

senior high school students to apply for higher education.

[18]Party membership refers (redundantly!) to the Communist party. Due to technical difficulties, these controls were not available in one session. We tested the results with the reduced sample and found no significant differences.

[19]Dropping covariates X_i and estimating a parameter δ_i instead.

[20]The natural estimation strategy that follows is parallel to a *a difference-in-differences* approach. In the context of laboratory experiments, the identifying assumption is satisfied, and although some violations to the randomization process are documented (see Table 6), we controlled for individual fixed effects which is the most flexible specification to control for unobserved (*time-invariant*) individual heterogeneity.

[21]Tobit panel regressions yield the same qualitative results.

Table 3.3: Determinants of contributions, regression results

	Dependent variable: Contributions (ECU)		
	Random effects (GLS)		Fixed Effects (ind.level)
	(1)	(2)	(3)
Democracy vs.Leviathan (PxD)	1.135***	0.623*	0.642*
	(0.390)	(0.348)	(0.346)
Punishment (P)	4.684***	2.200***	2.280***
	(0.280)	(0.272)	(0.272)
Democracy (D)	0.310	0.0239	$0^{\dagger}$
	(0.711)	(0.526)	(.)
Other's Avg.Cont. (t-1)		0.666***	0.647***
		(0.0210)	(0.0218)
Punishment received (t-1)		-0.0911**	-0.0718*
		(0.0406)	(0.0405)
Punishment in the group (t-1)		0.0108	0.00417
		(0.0218)	(0.0219)
Trend (within phase)		-0.236***	-0.230***
		(0.0323)	(0.0322)
Constant	6.920***	3.446***	3.560***
	(0.511)	(0.438)	(0.256)
R2 (overall)	0.136	0.412	0.411
N	3100	2945	2945

Standard errors in parenthesis. Sample differences are due to the use of lags.
† Variable dropped. Time invariant once controlled for individual FE.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

other factors. Note that the differences in treatment, once controlling for the changes in the institutional arrangement, are not significant.

The regressions analysis allows us to evaluate these results conditional on other factors, that are part of the game setting, and that might affect subjects' behavior throughout the experiment. When controlling dynamically for information available to the subjects in each period, the average treatment effect decreases to almost half of the unconditional result ($\hat{\gamma}_1 = 0.623$). Its statistical significance is also affected (the standard errors are about the same, hence it is the magnitude of the coefficient what drives the attenuation effect), however, it checks out for a 10% significance level. The effect of others' average contribution, which represents the social norm imposed within the group, is positive; in other words, subjects react to the within group behavior adjusting their posterior contributions. This effect is significant and explains an increase of about 0.7 (ECU). The deterrent effect of punishment is contradictory but very weak (0.09 units), which suggest that most results are driven by the punishment opportunities and not necessarily the way punishment is executed. Recall that in our design only one subject can be punished, hence, we argue that the general incentive for contributions withing the second phase is a "race to the top," i.e., no one wants to be the only one punished in the group. Finally, the average trend within phase is decreasing and significant.

We collected and test several controls withing our study. We found that men contribute consistently more and also that party membership (Chinese Communist Party) also plays a role in contribution behavior by increasing contributions. Nonetheless, the most flexible approach to control for time invariant observable and unobservable individual characteristics is to use a fixed effects estimation. We tested several additional specifications.[22] Column 3 includes the same estimation of equation 4 allowing for individual fixed effects. Results are robust for various specifications in all variables analyzed.

When it comes to assignment of punishment it becomes clear that punishment is assigned based on a reference for the social norm (the non free riders average contribution). Table 4 shows some related regression results. Columns 1 and 2 show the estimates for a linear probability model (panel, random effects),[23] in which the dependent variable is an indicator of whether or not the subject is punished. Column 3 and 4, analyze the amount of punishment received by subjects. Recall that the

[22]We tested for observed controls, group fixed effects, time fixed effects and combinations. Table 7 contains some of the specifications. Also, we estimate Tobit models for similar specifications; results remain consistent and conclusions are the same, but report RE and FE models for easiness of interpretation. Results available upon request.

[23]Similar conclusions are drawn from a panel Probit estimation. For easiness of exposition we show the linear probability version. Other results are available upon request.

leader has up to 20 point available each period, hence the intensity of punishment reveals the response sensitivity of the leader to the observe violation. Both results show the same pattern. Deviations from the social norm, and in particular, free riding behavior (see the negative deviation) increase the probability to receive punishment. Furthermore, the average additional punishment ranges from 0.4 to 0.7 points. More importantly, there is no evidence for different punishing behavior between treatments, the Leviathan and the Democracy. This has important implications in terms of efficiency in the endogenous distribution of power.

Table 3.4: Punishment decisions

	Random effects (GLS)			
	Dependent variable			
	Punished (= 1)		Punishment received (Points)	
	(1)	(2)	(3)	(4)
Democracy	0.0058	0.0046	0.269	0.255
	(0.0184)	(0.0164)	(0.217)	(0.199)
OMC absolute deviation $\lvert C_{it} - \bar{C}_{-it} \rvert$	0.030***		0.396***	
	(0.0023)		(0.0256)	
OMC negative deviation $\lvert C_{it} - \bar{C}_{-it} < 0 \rvert$		0.056***		0.695***
		(0.0026)		(0.0283)
OMC positive deviation $\lvert C_{it} - \bar{C}_{-it} > 0 \rvert$		-0.0024		0.0292
		(0.0028)		(0.0307)
Constant	0.0185	0.0275**	-0.341**	-0.240
	(0.0149)	(0.0134)	(0.173)	(0.159)
R2 (overall)	0.086	0.246	0.115	0.272
N	1550	1550	1550	1550

OMC: Other group members' contribution.
Standard errors are in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

3.5.3 An inquiry on leadership behavior

The extent to which the type of leader influences collective action deserves some attention in social dilemmas and sanctioning institutions. In this section we analyze whether it is the type of leader, as opposed to the institutional framework, what drives our results.

The fact that leadership plays an important role in cooperation, coordination and collective action has been well established in experimental research. In this literature, a *leader* has specific roles and, in general, takes a publicly observable action so the *followers* can acknowledge it and react accordingly by best responding. Leaders exercise their influence in many ways: *leading by example* (i.e., choosing first), communication (e.g., sending an encouraging message), networks (e.g., by their location within a network), etc. There is also evidence on the influence of the leader's *social status* where weakly induced status, normally by a trivia quiz and a public *star* recognition (Ball et al. 2001), differentiates contribution or cooperation behavior in non trivial ways (Eckel and Wilson 2007, Kumru and Verteslund 2010, Eckel et al. 2010). High-status leaders influence followers more effectively. In social dilemma games and sanctioning institutions, actions taken by the leader are very salient for the group members, to the point that, other things equal, the type of leader determines the final outcomes. Hence, it is natural to think that the leader's selection mechanism in these environments favors the endogenous institutional framework (voluntary leaders), although their level of influence is debatable (Arbak and Villeval 2007, Brandts et al. 2014).

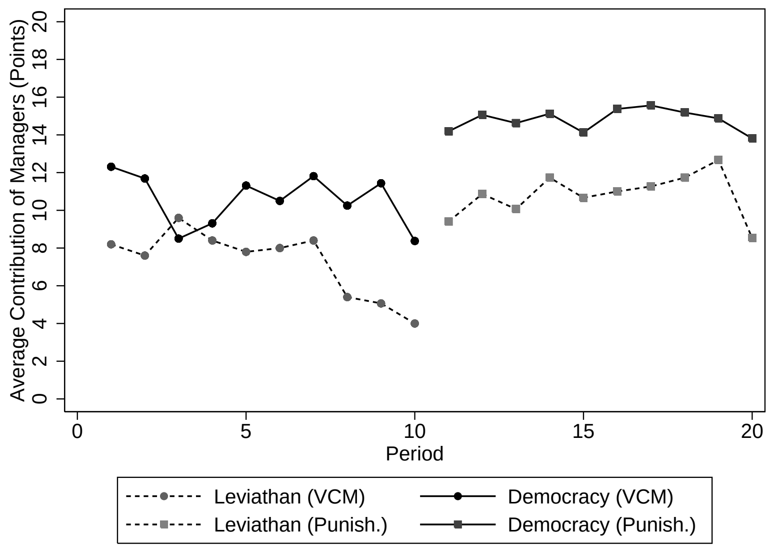
We take discrete distance from this literature in important ways. The leader in our environment is either randomly imposed or democratically selected for the whole sanctioning institution phase (second phase), however, any leader's action is anonymous and simultaneously taken with the rest of the group members, i.e., there are no *followers* in terms of decisions. Furthermore, provided that contribution information is presented to the players in random order, it is not ex ante clear for a subject whether the leader of the group contributes more or less than any social norm, i.e., the leader's action (contribution/punishment) cannot be identified from the actions of any other member. Many priors might come at play in both treatments, however; since the leader cannot inflict self punishment, his incentives to contribute are in fact lower during the second phase. Also, since the leader's punishment decisions are socially costly, punishment actions are binding for his payoff just in the same way they affect the other group members.

Table 5, Panel A, reports the comparison of contribution levels between leaders and other group members within treatments. The first section reports these differences for the first phase (VCM with no punishment). It provides evidence that the randomization mechanism worked in the Leviathan treatment, i.e. those leaders chosen do not behave differently from the others; while, in the Democracy treatment, elected leaders typically are high contributors. More importantly in these results, notice that our design gives the leader natural advantage over the rest of the group members. Different from decentralized punishment environments, the selected leader keeps its appointment throughout the rest of the experiment, hence, he is not at risk of receiving any punishment. The leader can choose to free ride safely (anonymity ensures no other concerns). Nevertheless, the power distribution mechanism triggers some positive "responsibility" impulse on the leaders (see Figure 4). In other words, those called to impose the social norm, feel compelled to comply, regardless of the selection mechanism.

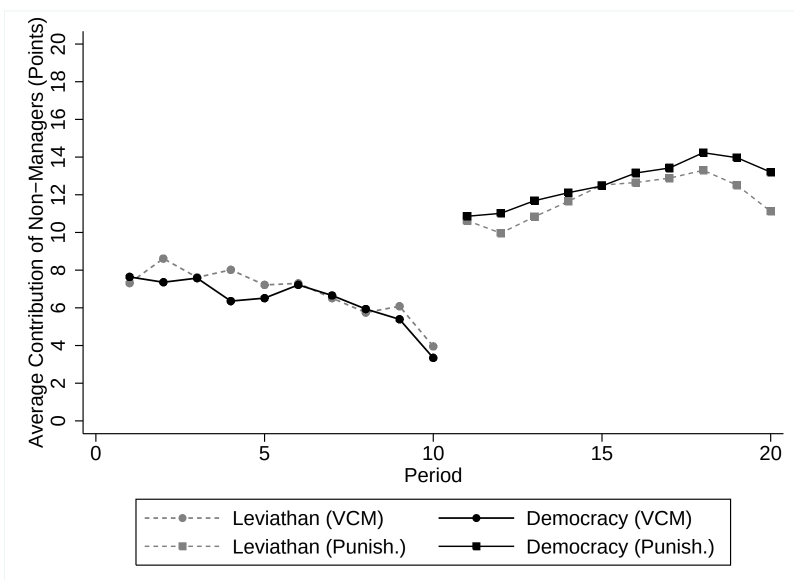
It is worth noticing, however, the differences across treatments. In the first period of the second phase (punishment institution) leaders in both treatments react positively, however, leader's contributions in the Democracy treatment are more consistent throughout the whole phase than in the Leviathan, where contributions drop dramatically on the last period. The second section shows that overall difference in average contributions in the second phase (tax/punishment institution) is not sta-

Figure 3.4: Average contribution by group hierarchy

(a) Managers' average contribution



(b) Other members' average contribution



tistically significant (12.86 versus 12.22, $p = 0.051$, MW U test). Differences arise between treatments. Leaders in the Democracy treatment contribute significantly more than their group fellows (14.80 versus 12.61, $p = 0.00$, MW U test); this difference is not significant in the Leviathan treatment ($p = 0.139$, MW U test).

Table 3.5: Leadership

Panel A			
	Managers	Others	p-value Mann-Whitney U test (H_0: equal means)
Baseline-VCM ($\bar{C}_{t\leq 10}$, Points)			
Overall	8.952 (0.184)	6.612 (0.434)	0.000
Leviathan	7.247 (0.639)	6.838 (0.283)	0.900
Democracy	10.550 (0.564)	6.400 (0.239)	0.000
Punishment ($\bar{C}_{t>10}$, Points)			
Overall	12.859 (0.416)	12.223 (0.189)	0.051
Leviathan	10.793 (0.631)	11.807 (0.267)	0.139
Democracy	14.794 (0.503)	12.613 (0.265)	0.000
Punishment-VCM			
Overall	3.906 (0.464)	5.610 (0.002)	0.000
Leviathan	3.547 (0.718)	4.968 (0.356)	0.031
Democracy	4.244 (0.598)	6.213 (0.342)	0.001
Panel B			
	Leviathan	Democracy	p-value Mann-Whitney U test (H_0: equal means)
Baseline-VCM ($\bar{C}_{t\leq 10}$, Points)			
Overall	6.920 (0.259)	7.230 (0.229)	0.182
Manager	7.247 (0.639)	10.550 (0.564)	0.001
Others	6.838 (0.283)	6.400 (0.239)	0.600
Punishment ($\bar{C}_{t>10}$, Points)			
Overall	11.604 (0.248)	13.049 (0.237)	0.000
Manager	10.793 (0.631)	14.794 (0.503)	0.000
Others	11.807 (0.267)	12.613 (0.265)	0.049
Punishment-VCM			
Overall	4.684 (0.319)	5.819 (0.299)	0.001
Manager	3.547 (0.718)	4.244 (0.598)	0.203
Others	4.968 (0.356)	6.213 (0.342)	0.002

Standard errors in parentheses.

Panel B in Table 5 reports the same information organized to test for differences of managers and non-managers between treatments. Again, when looking at contribution levels on the first phase (VCM-NP), the voting artifact shows the positive selection of leaders for those on the Democracy treatment, while contributions are not statistically different for other members, across treatments. When the punishment institution is in place (second phase), we reject the hypothesis of equal mean contributions either for leaders ($p = 0.00$) or for other group members ($p = 0.049$); in other words, the Democracy treatment induces significantly higher contribution levels than the Leviathan for both, managers and non-managers. Furthermore, given the fact that subjects in the Democracy treatment are given a signal about their leaders, an skeptic reader might argue that this evidence supports the idea of a confounding leadership effect.[24] Even though this skepticism has some room in our design, in the third section of Panel B we test for the differences in the changes in behavior between phases, by treatment. Consistently with our main hypothesis[25] contribution differences between phases are significant for other group members, while changes in contribution behavior across phases and treatments are statistically the same ($p = 0.203$)

The channel through which leadership plays a role is still somewhat ambiguous and requires further exploration. However, we argue that there are intrinsic incentives that need to be considered, a *self commitment behavior* in response to being allowed to choose a leader, the authority *legitimacy*; and, therefore, observed behavior is independent of leadership qualities. Regardless of the channel of influence, this is suggestive evidence that supports the idea that endogenous distribution of power

[24]We are somewhat sympathetic with this skepticism, and further efforts will be directed to isolate this effect.

[25]Hypothesis 1: the Democracy treatment outperforms the Leviathan and differences are unconditional on the leaders' behavior

(Democracy treatment) favors contribution levels and outperforms other institutional arrangements in social dilemma games with sanctioning opportunities.

3.6 Concluding remarks

We report new evidence of a laboratory experiment using a linear public goods game that focuses on endogenous sanctioning power. Departing from most related experiments under endogenous institutional arrangement, in which participants are allowed to vote for their preferred incentive mechanism in a decentralized institution (Gürerk et al. 2006; Sutter 2010), our experiment compares the effectiveness of exogenous versus endogenous distribution of sanctioning power under a centralized institutional arrangement.

Similar to previous findings, we find that sanctioning opportunities, either endogenous or exogenously assigned, outperform the standard voluntary contribution mechanism (VCM), both in terms of higher average contribution and earnings (social efficiency). Beyond the punishment effects on contributions, our experimental results present novel findings. We find that the endogenous power-distribution institution, the Democracy, promotes higher cooperation and efficiency than the exogenous power-distribution institution, the Leviathan treatment, in a centralized power environment. A possible explanation, the one we attribute our results within this analysis, is authority/institutional *legitimacy*, in the sense that presented evidence is driven by intrinsic incentives within the endogenous power distribution, conditional on the existence of punishment opportunities, i.e., there is a "self commitment" mechanism that activates in a democratic electoral framework. Whether this behavior is confounded by leadership or inequality aversion requires more in-depth analysis. We discard the pure leadership argument since, at least in our environment, managers do not change contributions differentially between phases, while differences observed in other group members are significant, unconditionally of treatment. Furthermore, leader's actions are not distinguishable from other's, and are simultaneously taken.

A final point should be made in terms of the external and internal validity of these results. Can these results be driven by the sample being drawn in China, a country with a centralized government? Our initial reaction is that this seems very unlikely. Without prejudice of further research on different cultural and institutional environments, as well as a call for replications that confirms or debates previous results in the discipline;[26] if anything, the fact that a population that faces a centralized power distribution reacts positively to an institutional arrangement that involves a democratic election, only reaffirms our suspicion that there are intrinsic motivations in a democratic institutional framework that affect effective choices. How this institution is linked to contributions is a research question that we try to address in this research. Nevertheless, it is still an open empirical question to define how context dependent experimental results are.

In summary, the introduction of a democratic participation mechanism into sanctioning power institution for a public good provision, reduces the free-riding problem in social dilemmas and promotes higher cooperation without sacrificing social efficiency (cost/effective). These results expand on the agency problems and the contracts that enforce cooperation disregarding the agent/manager selection mechanism. Furthermore, they offer valuable insights to explain differences in governance outcomes of common pool resources and provision of public goods whether in the lab or in the field. On the political arena, our results support the faith in liberal democracies, conditional on having an institutional framework in place that holds their members accountable.

[26]See Eckel et al. (2015) forthcoming.

4. INDIVIDUAL PREFERENCES AND CREDIT BEHAVIOR: AN EMPIRICAL INQUIRY ON CONSUMER UNOBSERVED HETEROGENEITY

4.1 Introduction

One known fact about the discounted utility model, rarely discussed, is the indistinguishable nature of the psychological motivations for intertemporal choice. Although its simplicity provides interpretative advantages, a large body of literature has investigated its behavioral limitations (see Frederick et al. 2002). Empirically, structural approaches that rely on particular utility functions would not recover any generalizable information over these motivations. Beyond the choice dynamics, the most important feature of credit markets is the agency problem derived from the imperfect information framework. Typically, one of the incumbents in any credit contract, namely *the agent*, has more information over its willingness and capability to repay a loan, information that is not necessarily revealed by the clients or captured by very rigorous credit assessment procedures. By means of a finite mixture density approach, I investigate the possibility to recover (at least in a partial sense) some information over individual preferences, strategic behavior and willingness to repay a loan, by uncovering the unobserved *quality-types* of clients on the credit market, and show some of the statistical advantages of modeling credit behavior by identifying a latent process.

In the core of the neoclasical model, subjects face different decisions over consumption, savings, investment and credit demand, over a life cycle. Intertemporal choice over these variables depends mainly on forward looking considerations over future income and budget constraints, what is known as the Permanent Income Hypothesis (Friedman, 1956). The *free access* [1] to credit, allows agents to allocate resources intertemporally, only now they need to consider the financial burden associated with the use of credit.[2] Following the "rational behavior" hypothesis, financial decisions and, in particular, loan's demand, are the result of rational choice and individual maximizing behavior; in other words, economic agents, consciously or not, rely on some form of cost/benefit analysis to determine their intertemporal allocation of consumption; i.e. define the stream of credit loans/payments that best serves their long term interests. However, a feasible choice in the short run is to "default," that is, fail to honor the loan payments on a particular period to keep current liquidity, and bear whatever consequences might follow, which are far from the mere financial costs. The traditional way to summarize such behavior in most credit markets is the *credit score (or score-card)* which has become a powerful tool to signal reputation and creditworthiness of clients. The influence of this information mechanisms can not be underestimated, and now days it goes beyond basic financial restrictions, from affecting future access to credit, interest rates and borrowing limits; to less obvious areas including auto or home insurance prices, employment screening and tax compliance (Fisher and Lyons, 2010).

Traditional credit score practices are built either on available debt balance or, in the best case scenario, through the use of micro level data on historical information about consumption and demographic characteristics, leaving out a wide range of unobserved factors that affect behavior. That is the case of models that explore the Default Probability (DP) plagued with endogeneity problems due to the lack of information over individual preferences and selection bias. How to account for

[1] By *free access* in this case I mean free entry, i.e. no liquidity constraints.

[2] The foundations of the *representative agent model* used in macroeconomics rest on this intertemporal (dynamic) approach. See Deaton (1992) for other insights over the implications of time consistency in other macro models.

the particular idiosyncrasies of credit users is an important empirical and theoretical question. Experimental practices have developed a wide range of methods (elicitation methods) as revelation mechanisms for individual preferences, some more successful than others; yet its applicability outside the laboratory is cumbersome and even questionable. A suitable empirical alternative, proposed by Gan and Mosquera (2008) that I revisit in this paper, is to add some basic statistical structure to the intertemporal choice problem described in order to identify client's heterogeneous types. This approach offers a good compromise between the estimation of structural parameters which, combined with some additional information that can be considered orthogonal to the population errors, support the identification of the model.

From a policy perspective, knowledge of the individual heterogeneity of clients is highly desirable. On one hand, reducing the information asymmetries in the lender/borrower relationship may lead to improvements on the credit risk management of the company by reducing the adverse selection problem, i.e. selection of riskier clients. The importance of adequate private risk management in financial institutions cannot be stressed enough; if credit risk becomes systemic, the social consequences are catastrophic, it is enough to see how bad risk management in the so called *subprime* loans have caused in the financial crises in US and Europe in recent years. Aside from this dark perspective, knowledge of credit types can also result in more efficient credit allocations, potentially reducing *credit rationing* conditions in competitive markets. As supported by equilibrium models under rationing conditions, having multiple contracts based on signaling mechanisms can potentially improve (Pareto) market efficiency.

The paper's contributions can be summarized as follows. First, it proposes an empirical approach to identify the existence of private information in the credit market that results in different *quality-types* of clients. The estimation procedure offers better withing and out-of-sample predictions of delinquency rates/default (better relative fit) for a two types equilibrium; in other words, accounting for unobserved heterogeneity there are important improvements on screening techniques applicable to the financial industry. When it comes to the credit demand, due to higher variance, we identify up to five *quality-types* of clients and the specification helps to interpret behavioral differences in the elasticities of the economic variables. Second, it expands the reach of previous studies (Gan and Mosquera, 2008; and Gan, Hernández and Liu, 2013) by extending the domain to the analysis of education and labor information, with potentially higher relevance for type identification. Finally, it supports the use/collection of additional information and experimental elicitation mechanisms for individual preferences to disentangle the problem of heterogeneous types identification on a mixture density framework. Additionally, it includes a discussion of the partial identification problem present in such models and how the experimental elicitation mechanisms can help improve the consistency of the results.

The paper is organized as follows. Section 2 summarizes some of the main results and previous findings from the relevant literature. Section 3 covers a theoretical discussion about the model that offers some intuition over the results. Section 4 describes the empirical approach adopted for the type-consistent estimation. Section 5 summarizes the main results and out-of-sample performance. Section 6 concludes.

4.2 Literature review

4.2.1 Individual preferences and financial decisions

When it comes to credit many factors are involved in the decision making process and, depending on the institutional setting, they can be analyzed as part of an individual or a group decision problem. In micro-finance programs (joint liability), most studies concentrate on assessments of group homogeneity (Cassar et al., 2007) and trust (Karlan, 2005). In general, *trustworthy* individuals[3] and homogeneous groups are significantly less likely to default.

Among individual preferences, those related to risk and time are ubiquitous in economic analysis and financial literature, and rigorous experimental analysis is relatively scarce and challenging mainly due to the lack of access to detailed micro level data from financial institutions. On the other hand, technical efforts over available information are doubtfully *causal* in a strict sense and are forcefully far from a *Randomized Control Trial*. The difficulties of such approaches are obvious, either for budgetary, commercial or moral reasons. Yet, some of the *correlational* studies available offer some compelling evidence and constitute a body of knowledge in itself that shows how both types of preferences, can be considered determinant factors that shape economic behavior and financial decisions. In this line, Arya, Eckel, and Wichman (2011) analyze how impatience and risk aversion of individuals help explain the FICO credit score.[4] They find that measures of impulsivity, time preferences and trustworthiness are highly correlated, however, they found no compelling evidence about risk aversion. Meier and Sprenger (2010) study the *present bias* phenomenon for a field experiment with low and moderate income families in Boston. They show that present-biased individuals, i.e. people that have strong desire for immediate consumption (or instant gratification), keep higher credit balances than dynamically-consistent individuals. Eckel et al. (2005) find evidence from a sam-

[3]Note that "trustworthy" in experimental work involves trust received by the members of the group receiving the loan, not trustworthiness in a general sense, which can be misunderstood as "creditworthiness."

[4]The FICO (Fair, Isaac, and Company) credit score, is the pioneer and most widely credit score used in the US and is based on clients information from three major credit bureaus: Experian, Equifax, and TransUnion. It is a holistic measure of creditworthiness based mainly on payment history (35%), revolving debt ratios (30%), among other factors: length of credit history, type of credit, new credit lines, etc.; and its index goes from 300 to 850.

ple of Canadian working poor individuals and show that "...risk-averse individuals are more present-oriented;" also, present bias and preferences over short/long term consumption/saving decisions are highly correlated. Eckel et al. (2007) study how *debt aversion*, risk attitudes and time preferences are correlated with investing and borrowing decisions about post-secondary education in a diversified working class sample. Even though they find no significant evidence for debt aversion, they show that measures of risk and time preferences, used as control variables in their analysis, are significant factors for the human capital investment decisions.

Important extensions of the time preferences analysis are models that account for time inconsistency and present bias, known as Hyperbolic Discount models. This approach offers important insights over consumer and credit behavior. Laibson et al. (2007) show that subjective discount factors are far different depending on the time horizon of credit card debt and asset accumulation for retirement. In particular, discount rates are higher for short-term goals than for long-term ones, which implies that subjects have conflicting time preferences, i.e. they act patiently and impatiently at the same time, depending on the time horizon for each decision. Importantly, when accounting also for risk aversion in the simulation they show lower discount factors. The indivisible aspect of risk and time preferences is currently an empirical challenge and some new experimental methods address this issue (Andreoni and Sprenger 2012). Nevertheless, the multilevel problem of individual preferences is still far from definitive and needs to be analyzed also from a broader psychological perspective. A wide body of research in psychology and marketing focuses on self-control measures, impulsivity, mental accounting and other traits; and shows how they offer some explanatory power over general financial behavior (Shrefrin and Thaler 1988 and Prelec and Simester 2001).

4.2.2 Consumer credit analysis

Consumer credit analysis, much as any economic problem, can be addressed from the supply or demand side of the problem, not always successfully distinguishable from each other, mostly due to the constraints over data availability and known identification challenges. Observational studies over the intricacies of consumer decisions are abundant.[5] In the US, most empirical studies support the idea that the Permanent Income Hypothesis fails to account for the presence of biding liquidity constraints in the credit market, and offer some evidence of alternative theoretical approaches, namely, models of *buffer-stock savings* (Carrol 1997) or precautionary savings. Grant (2007), using the US Consumer expenditure survey (1988-1993) finds that around 31% of US households face credit constraints; furthermore, demographic variables play an explicative role and the incidence of such constraints is higher for single females, well educated households and middle income households. Racial differences also appear important although they seemingly come from the demand. Gross and Souleles (2002), using administrative data for credit card users, show the existence of binding liquidity constraints in the credit market. By looking at changes over the credit limits, balances and consumer individual interest rates, they estimate a long-run Marginal Propensity to Consume out of liquidity of around 10-14%, which is heterogeneous depending on whether subjects start closer to the credit limit. They also estimate an average long-run elasticity of debt to interest rate of around -1.3, again heterogeneous conditional on the relative starting position with respect to the credit limit. Additionally, around 90% of households keep very liquid assets (savings and checking accounts) at very low interest rates while at the same time keeping credit card balances (Bertaut and Haliassos 2005). This evidence supports the idea

[5] See Bertola et al. 2006 for a detail summary.

of strategic behavior over credit management.

4.2.3 Credit analysis and finite mixture models

On the empirical side there is not a one-size-fits-all methodology for credit analysis. Statistical methods have a long tradition that dates back to the *discriminant analysis* proposed by David Durand in 1941. Much ground has been covered since and, nowadays, methods vary from ordinary linear regression analysis, OLS, probit / logit estimation, to some more sophisticated that include: nonparametric smoothing, mathematical programming, discriminant analysis, data mining, Markov chains models, neural networks; among others.[6] In essence, all methods have the same objective, separate the "Good" from the "Bad" in terms of creditworthiness.[7] As simple as this may sound, there are several challenges related to modeling credit behavior. Even with detailed information, individual tastes and preferences are traditionally out of the reach of any *underwriting* technique, hence; by ignoring them, lenders face important credit allocation risks.

Finite mixture models, within the more general set of *latent class models*, have a long history in statistics and offer a flexible approach for a type-identification strategy. This estimation method has been applied in different areas of sciences[8] predominantly in marketing and consumer analysis, where they are commonly used for market segmentation (See Tuma and Decker, 2013).[9] In economics, pioneering work is Heckman and Singer (1984) who analyze the relationship of a Non-parametric

[6]See Hand and Henley (1997) and Thomas (2000).

[7]It is important to keep in mind that default *per se* is not an inconvenient action from the point of view of the lender, as long as the debt is honored eventually; this allows for interest charges that result in higher profit. A "bad" action would be not paying at all.

[8]Some related approaches include: Zero-Inflated Poisson, Zero-Inflated Negative Binomial, Poisson Hurdle Model, Gaussian Mixture Model, Regime(Markov)-Switching Model.

[9]Other areas include medicine, where applications fluctuate between drug heterogeneous sensitivity to genetic unobserved/uncontrolled factors (See Schlattmann, 2009); and psychology, where these models contribute to the analysis of behavioral differences over preferences and tastes (See Lubke et al. 2005).

Maximum Likelihood Estimator within a labor search and duration model environment and showed how this problem can be reduced to a discrete mixing distribution. Other related work includes Keane and Wolpin (1997) for a labor and human capital inquiry; Gan, Huang and Mayer (2008) who identify the presence of private information in insurance markets and Feinstein (1990) who identifies heterogeneous responses for law violations and crime detection.

Applications over credit markets are scarce. Alfo et al. (2005) use similar techniques to analyze firms classification based on creditworthiness. Gan and Mosquera (2008) (GM08 hereafter) investigate the default probability for credit card users that encompasses traditional scoring techniques and a mixture density estimation to identify the existence of two types of clients. I revisit their approach in this paper and analyze some caveats and extensions. Another contribution is the work of Gan, Hernández and Liu (2013) who analyzed the repayment behavior in group lending (*joint liability*) and show how accounting for *group types* improves the predictive power of traditional models of default probability. Nevertheless, depending on the problem at hand, the use of this estimation approach comes at the cost of weak and *ad-hoc* identification arguments; the identification sections offers some insights over this problem.

4.3 Credit decisions, asymmetric information and signaling: theoretical discussion

One of the main allocative inefficiencies that a competitive credit market faces is *credit rationing.* Credit rationing occurs when the lender limits the amount of loans offered (credit supply), at a given interest rate, even though there are borrowers willing to accept higher interest rates to access the credit market. In other words, there is an unsatisfied excess demand for credit due to the fact that rational lenders are not willing to incur in higher credit risk levels that could lead to lower profits (credit risk management). Credit rationing manifests in several ways. On one hand, there are clients excluded from the market although they are indistinguishable (on observables) from others receiving credit, at any interest rate; or on the other, reliable clients either receive lower credit limits or hold lower balances than they would otherwise. In a perfectly competitive market, the interest rate (the price of credit) should increase to eliminate the excess demand, however, the nature of imperfect information and uncertainty avoids such simplified market response and interest rates tend to be sticky. A seminal work that describes the equilibrium and credit rationing conditions under imperfect information is given by Jaffee and Russel (1976) (JR76 hereafter). They describe the existence of a (Nash)equilibrium under credit rationing for a single (pooling) contract, i.e. a pair $\langle L, R \rangle$, that can be strictly preferred to the non-rationing (competitive) equilibrium. Furthermore, they argue that deviations from the credit rationing equilibrium, to a multiple (separating)-contract equilibria are intrinsically unstable and there is no Nash equilibrium possible. Stiglitz and Weiss (1981) offer an alternative characterization for credit rationing equilibria where the market interest rate, chosen by lenders over a profit maximization argument, does not clear the market although it is at an equilibrium level. In their setting, credit rationing occurs because rising the interest rate (or collateral requirements) would negatively impact the pool of clients that enter the market (adverse selection), inducing better clients to leave the market due to the costs while at the same time riskier clients accepting the conditions. Furthermore, higher interest rates might negatively affect investors' incentives (moral hazard) who might choose riskier projects to compensate for the financial costs of the loan. Lenders have no incentive but to choose an average interest rate that diversifies is pool of clients and maximize their profits. Milde and Riley (1988) (MR88 hereafter) add uncertainty conditions to the JR76 basic approach and describe how, if signaling is allowed for loan-size and interest rates, a multiple-contract Nash equilibrium arises as a possibility in the market.

A detailed analysis of the theoretical development of these arguments is beyond the purpose of this paper. I summarize in this section some of the main insights and challenges of the theoretical approach that motivates the empirical work.

4.3.1 The competitive model

Assume an intertemporal utility function for 2 periods,[10] that is additively separable (time separable); the maximization problem of any agent i (dropped for notational convenience) is:

$$\begin{aligned} \max_{\langle c_1, c_2 \rangle} \quad & U(c_1, c_2) = U(c_1) + \alpha U(c_2) \\ \text{s.t.} \quad & c_1 = y_1 + L \\ & c_2 = y_2 - RL \end{aligned} \tag{4.1}$$

Where $U(.)$ is (any) concave function[11] and a time discount factor $\alpha = 1/(1+\delta)$ for a δ subjective time preference parameter. L represents the loan principal, y_t is income in period $t = 1, 2$ and $R = (1+r)$ represents an interest factor. Individuals are identical and face an interest rate r from a competitive credit market. Any borrower faces this maximization problem and chooses over the consumption smoothing pattern, taking everything else as given (i.e. exogenously defined). Since this is essentially a sequential argument, the budget constraint can be written as: $c_2 = y_2 + R(c_1 - y_1)$. The use of this (binding) constraint in the maximization problem implies *not defaulting*. JR76 defines the borrower that restricts his consumption to this condition, as the *honest* type.

This can be translated into the following unconstrained optimization problem

[10] I concentrate in a two periods intertemporal choice problem for exposure simplicity, however the main theoretical conclusions extend to a dynamic finite/infinite horizon set up. See Deaton (1992) and Attanasio (1999) for the basic model details.

[11] As usual, concavity implies $U' > 0$ and $U'' < 0$.

where the agent chooses over the loan amount:

$$\max_{L} \quad U(L+y_1, y_2 - RL) = U(L+y_1) + \alpha U(y_2 - RL) \tag{4.2}$$

From the FOC of the maximization problem (Euler equation) we derive the Marginal Rate of Substitution:[12] $MRS_{c_1,c_2} = \frac{U'(c_1)}{U'(c_2)} = \alpha R$; hence the *loan-size-choice* of a client is determined by the ratio between the interest rate and the agent's time preference $\left(\alpha R = \frac{1+r}{1+\delta}\right)$; in other words, how much the individual is willing to give up of consumption in period two, to consume a unit in period one.[13]

Up to this point, it is clear that the consumption pattern (consumption *smoothing*) depends on the relative size of the subjective time preference coefficient and the market interest rate. If we are willing to assume some level of risk aversion, as is common in economic studies, i.e. concavity of $U(.)$; we add a layer over the optimization problem. It is common ground on theoretical and experimental work to recognize that these preferences are interrelated, in fact, as Frederick et el. (2002) mentions, they "... create opposing forces in intertemporal choice: diminishing marginal utility (concavity of utility function or risk aversion) motivates a person to spread consumption over time, while positive time preference motivates a person to concentrate consumption in the present."[14] Ignorance of this relationship affects any possible related inference. Relatively recent work on experimental economics has shown that joint elicitation of risk and time preferences helps adjust parameters previously identified

[12] The MRS between periods represent the underlying preferences independently of the utility function chosen (preference invariance). This can be seen by applying any monotonic transformation $v(u)$ to the utility function (Varian, 1992).

[13] In experimental studies, when *time preferences* are measured independently, the subjective implied *discount factor* (typically $1/(1+r) = 1/R$) is obtained from a sequential multiple price list (see Frederick et al. 2002 and Harrison et al. 2008 for a summary); the higher the discount factor (the lower the interest rate), the more inclined towards present consumption an individual is. Such results assume risk neutrality.

[14] Frederick et al. (2002), p.359. Text in parenthesis is mine.

using independent assumptions over risk or time neutrality.[15]

A full characterization of consumers on only these two dimensions: time and risk preferences, would consider the continuum of both parameters involved (all possible combinations of risk seeking/neutral/averse and patient/time neutral/impatient subjects). Such characterization is useless for analytical and empirical purposes, just in the same way that infinitesimal differences in consumption patterns are not informative about human behavior. "People are different" however, in terms of understanding economic behavior, we are more interested in how people are similar.

It is straightforward to derive a *loans' demand function* from the FOC, which is a function of the interest rate level R, and the individual preference parameters (we assume a determined /certain income level for both periods):

$$L^* = L^*(R, \alpha, \kappa) \tag{4.3}$$

From the concavity of the utility function we can directly infer $\partial L/\partial R < 0$,[16] i.e. a downward sloping demand function in the $\langle L, R \rangle$ space for the competitive market case (see Figure 1). Relaxing this assumption and allowing fluctuations of the risk parameter on the *risk seeking* range, highly complicates the analysis; that is, regardless of the interest rate level, we still need to say something about the other parameters, and specifically, the sign of $\partial L/\partial \kappa$ and $\partial L/\partial \alpha$.

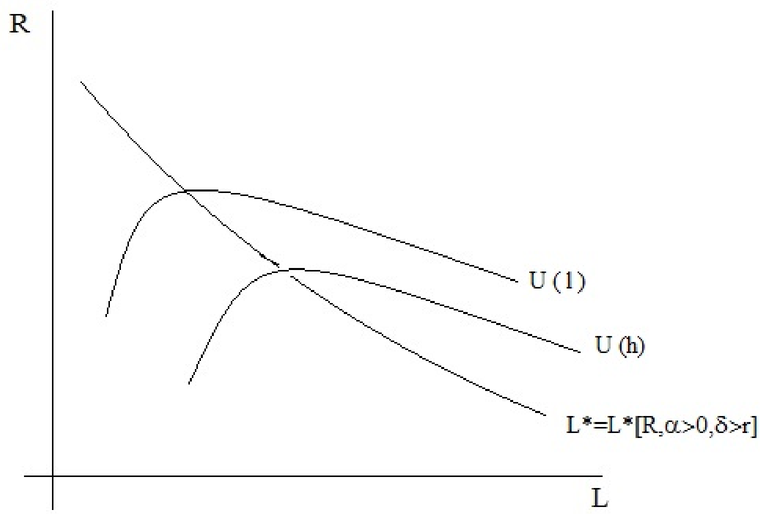
4.3.2 The Milde & Riley / Jaffee & Russel model revisited

Milde and Riley (1988) offer theoretical conditions for *Pareto efficiency* in an equilibrium model with separating contracts, as opposed to a single (pooling) contract, which; under capital costs for the lender, results in a *rationing equilibrium.*

[15]See the work of Harrison et al. (2008) and Andreoni and Sprenger (2012).

[16]Jaffee and Russel (1976) showed these conditions for a general case.

Figure 4.1: Loan's demand and indifference curves



4.3.2.1 Borrowers "quality" type-identification

Following MR88 by adding an uncertainty factor to the second period consumption, we can restate the original constraints in the general model (equation 1) ; hence we have:

$$
\begin{aligned}
c_1 &= y_1 + L \\
c_2 &= max\left\{Z, \tilde{Y}_2 - RL\right\}
\end{aligned}
\tag{4.4}
$$

where, $\tilde{Y}_2 = \tilde{u} + \theta$ and $E(\tilde{u}) = 0$.

The term $\tilde{u}$ is an stochastic component (second period income exogenous shock) for which we assume the existence of a *cdf*, $G(u)$; differentiable and strictly increasing in $G \in (0, 1)$ (see assumption A1 in MR88). θ is a type component that is a function of (observed) income level on the second period (Y_2). In this paper, I interpret

Z as a "bankruptcy threshold" [17] and corresponds to a heterogeneous unobserved individual factor;[18] thus the second period consumption, hence the default decision for any subject j will be determined by:

$$\tilde{u} + \theta_j - RL \lesseqgtr Z_j \tag{4.5}$$

The uncertainty factor affects the optimization decision. Agents now optimize over the loan expected return. To explicitly allow for heterogeneous levels of risk aversion, without loss of generality, assume the following: i) $U(.)$ has the canonical CRRA form, ii) $Z = 0$ (which reduces the problem to the two types case) and iii) let the default decision hold with equality at u^*; by direct use of this constraint on the agents utility function, the unconstrained optimization problem becomes:

$$\begin{aligned}
\max_L \quad U(c_1, c_2) \quad &= U(L + y_1, max\,\{Z_j, \tilde{u} + \theta - RL\}) \\
&= U(L + y_1) + \alpha \int_{u^*}^{\infty} U(u + \theta_j - Z_j - RL) dG(u) \\
&= \frac{(L + y_1)^{1-\kappa}}{1-\kappa} + \frac{\alpha}{1-\kappa} \left(\int_{u^*}^{\infty} (u + \theta_j - Z_j - RL) dG(u) \right)^{1-\kappa}
\end{aligned} \tag{4.6}$$

Gan and Mosquera (2008) derive the FOC and characterize the equilibrium conditions for a similar problem. They showed that, for a competitive credit market with flexible interest rates and a risk averse agents ($\kappa > 0$); $\partial L^*/\partial \kappa < 0$, in other words, "the amount of loan demanded is lower for those individuals with higher risk aversion."[19] One technical note on their approach is that they assume *full uncertainty* in the second period income, thus imposing an explicit independence assumption be-

[17] MR88 would define this term as a *legally determined minimum income.*

[18] A somewhat appealing interpretation of such heterogeneity, at the cost of generality, comes from JR76 where a similar uncertainty factor comes from a *penalty default* where $Z = \bar{Y}_2 - Y_2$.

[19] That is to say that for $L > L^* \Rightarrow \partial R/\partial L < 0$ and for $L < L^* \Rightarrow \partial R/\partial L > 0$, due to the concavity of $U(.)$

tween the types and their income level. This is a restrictive assumption (somewhat implausible) considering that income signals the type of an individual involved in the credit market, just in the same way that it does so in the human capital literature.

As mentioned in the more general model, individual identification of both parameters is cumbersome theoretically as well as empirically.[20] Nevertheless, to account for such preferences explicitly, experimental measures can help identify the presence of private information. I come back to this point in the identification section and do not extend on the mathematical tractability of these results since, as mentioned, they impose restrictive assumptions that lead the interpretation over particular types of preferences that cannot be distinguished empirically without further information.[21]

For the two types case, it can be seen that in equation 5, accounting for the *bankruptcy threshold*, $Z_j = 0 \Rightarrow \tilde{u} + \theta_j - RL = 0$. We defined the stochastic shock as $E(\tilde{u}) = 0$, therefore the expected second period income level, defined by the type of the individual, leads to zero consumption ($E(\tilde{Y}_2) = \theta = RL$).

Abusing the notation, we can extend the analysis to *j-types* by the parameter θ_j for $j = 1, ..., J$. *"Types,"* in this regard, are a combination of underlying factors or individual characteristics that together represent different *quality-clients*. Using the *bankruptcy threshold* (exogenous and unobservable), now by type of agent Z_j, we can generalize the default identification conditions. Intuitively, every *type* of agent has a different Z_j, i.e. the agent's type determines his second period consumption level,

[20] See Gan and Mosquera's (2008) appendix. They derive the corresponding conditions for risk aversion, using a CRRA utility function, and for time preference, using a quasilinear utility function, independently from each other. In their approach, admittedly, there is no explicit way to identify these preferences from each other. In theory, whether it is the traditional principal/agent problem, or asymmetric information in mechanism design, types can be any two antagonist categories based on an underlying model specification; theoretical characterization is, in this regard, loose.

[21] GM08 do consider a parameter β that works as a discount factor in the model, however they do not offer any interpretation as to its roll in the FOC. Taking this into account, in our interpretation (from equation 5) their results will translate into the *risk dominance* ground, i.e. time preferences are strictly dominated by the risk factor and the range of their analysis restricts to the "patient type" i.e. $\delta > r$ within this framework.

thus his choice, over default or not, depends on his subjective threshold. As a result, for J types of agents the selection into types comes from the following identity:

$$\tilde{u} + \theta_j - RL = Z_j \tag{4.7}$$

Under very general conditions[22] *high quality* individuals signal by accepting smaller loans since they face reduced uncertainty in the second period, either from better ex-ante beliefs over the success of their investment (the actual use of the loan) or from their better income expectation (due to their type); while *lower quality* subjects at a given interest rate, signal by asking for higher loans. From equation 7, we can define the probability of default of a particular type as $G(RL + Z_j - \theta_j)$, hence the repayment probability is $[1 - G(RL + Z_j - \theta_j)]$, and combining the demand side of the analysis, the main conclusions for the equilibrium model can be summarized as follows:

- $L^*(\theta_j)$ is a decreasing function in θ_j; and,
- $[1 - G(RL + Z_j - \theta_j)]$ is an increasing function of θ_j.

Whether it is for the heterogeneous external "moral" or "economic" costs of default faced by each agent (JR76), or due to existence of private information related to individual preferences; the *quality-type* corresponds to an *ex-ante* unobserved combination of factors, hence any argument towards the *ex-post* classification of subjects requires some intuition over explicit information available.

4.3.2.2 Lender and market behavior under competitive conditions

The equilibrium conditions from the supply side of the problem are straightforward. Following JR76 and MR88, assume a large number of lenders obtain their

[22]See MR88, proposition 7, p.113. A central assumption for this result is the additive form of the utility (production) function.

resources from the capital market at an interest rate i (or more generally the opportunity cost of the resources), for whatever loan amount the lender has to pay $(1+i)$ over the principal. Considering that lenders are risk neutral they maximize their loan expected profits, accounting for potential outcomes that involve the uncertainty shock $\tilde{u}$ and the unobserved types θ_j, the maximization problem of the lender becomes:

$$E\left[\pi_{\theta_j}(L,R)\right] = RL\left[1 - G(RL + Z_j - \theta_j)\right] - (1+i)L \tag{4.8}$$

In words, the expected profit from the lender corresponds to the net expected loan repayment.

The supply curve, that is the set of loan contracts in the $\langle L, R\rangle$ space that satisfy the competitive market condition of $E\left[\pi_{\theta_j}(L,R)\right] = 0$ (the *iso-profit* curve), considering the repayment probability give the client's types, can be derived implicitly from the FOC:

$$R = \frac{(1+i)}{(1 - G(RL + Z_j - \theta_j) - RLg(RL + Z_j - \theta_j))} \tag{4.9}$$

The supply's slope can is as follows:[23]

$$\frac{dR}{dL} = \frac{(1+i)}{L\left[1 - G(RL + Z_j - \theta_j) - RLg(RL + Z_j - \theta_j)\right]} - \frac{R}{L} \tag{4.10}$$

This last equation is useful for our intuition. Under perfect competition, $E\left[\pi_\theta\right] = 0$ and $R = (1+i)$; thus the slope is equal to zero only if the difference between the repayment probability and the *value at risk* of the total loan is equal to 1 (the denominator of the function). If the market faces no risk, i.e. the probability of

[23]Using the implicit function theorem. See Gan and Mosquera (2008) for similar results.

default is zero (hence the *value at risk* is also zero), then there is no rationing and the market supply covers all the demand, the competitive equilibrium. Back to the types characterization, the competitive equilibrium can only be possible if in the second period the loan size of repayment is below the bankruptcy threshold for every type, that is, beyond this point, and under mild risky conditions the slope of the supply curve is positive up to a point where the interest rate compensates for the credit risk due to indistinguishable clients (all clients are covered, and everyone pays back the loan). The sign of the slope at this point needs additional assumptions over the probability distribution; however, under no rationing conditions, as JR76 argue, "the particular shape of the supply function does not affect (the) main results."

A multiple-contracts equilibrium is also possible in this setting and it helps identify types of clients allowing for a preferable allocative equilibrium where less clients default due to better match of loan sizes, yet; high quality clients would increase their utility level at a lower loan and interest rates.

In summary, from the supply side of the market, the main conclusion over types' identification is intuitive: the expected profit of the lender improves with a pool of better clients.

- $E\left[\pi_{\theta_j}(L,R)\right]$ is an increasing function in θ_j

Under certain set of assumptions[24] MR88 argue that the loan signaling process derives from the *exit* conditions in the market, specifically; the higher the loan size the risky the environment the lender faces. As a result under a pooling contract where there is no signaling, lenders that exit the market are those at the right end of the iso-profit function, those with higher/riskier loans. Separating contracts, that is to say, contracts that allow for heterogeneous signaling (through R and L

[24] Being the most important the form of the "production function" of the loan in MR88 model. See section III, pg.111 for details.

in a competitive market, or simply L in a more restrictive environment) reduce the rationing possibilities due to lender's exit since there is more information over the actual risk the lender faces. In the core of the MR88 analysis is embedded the existence of an stable equilibrium with multiple contracts, that is to say, borrowers at bod ends of the quality spectrum signal by their loan size and lenders identify such idiosyncrasies and exploit them through different contract conditions to manage their overall market credit risk and profits.

When it comes to interest rates, in most credit markets[25] they are given by the lender, or at least there is an scarce room for negotiation between the incumbent parties in financial contracts, specially in consumption credit markets.[26] Even though price competition is possible, it can be limited for several reasons. In the presence of informational asymmetries, the adverse selection problem implies that *low quality* individuals are more willing to accept higher interest rates to access the credit market, signaling also through higher loan amounts; as a result, the credit risk that the lender faces when reducing interest rates to attract more clients is higher. Hence there are lower incentives for price competition, however; even the adverse selection problem remains present (Stiglitz and Weiss 1981).

Relaxing the environment for imperfect competition conditions, we can think of other reasons for heterogeneous elasticities on the credit market. Current clients (incumbent) have particular contract conditions due to the private information that the lender has collected through the credit history; as a result lenders can offer

[25]It is the case in Ecuador, the country in which the empirical section is based.

[26]In the US, credit cards take advantage of information from credit bureaus to offer some promotional rates of even heterogeneous rates depending on the product and client's credit history. Also, some heterogeneity of interest rates can be observed on contracts in specific credit markets where there is more information, such as housing credit or even corporate credit market where companies face different interest-rate-schedules depending on amounts and other contract conditions such as collateral or mortgages. Nevertheless, for the most part, in consumer credit (e.g. credit cards), interest rates, although heterogeneous, are given by financial companies.

preferential treatments for creditworthy clients or any other non-pricing strategy as a reward for their "loyalty" (Sharpe 1990). The switching costs could simply imply that clients are more comfortable with the policies of the current lender. On the other hand, there are also *search costs* involved. Calem and Mester (1994) argue that *search costs* are higher for high quality clients since their loan's demand elasticity is lower, i.e. they are less sensitive to interest rate changes, and doing an effort to locate information is relatively more costly. Either due to search costs, switching costs or costumer *irrationality* (Calem and Mester 1994) imperfect competition is a more realistic framework. Imperfect competition only requires that $r > i$ for the lender to enter the market, however, since the market faced in the empirical approach is closely defined by a constant interest rate ($\bar{r} > i$), in the $\langle L, R \rangle$ space it implies a flat supply curve.

The arguments summarized in this section provide some basic theoretical framework to interpret the type's identification coefficient. They also capture closely the credit card market structure. From the lender's perspective there is a trade-off in the selection process. Clients that always pay their monthly balance in full (never default) are the most "safe" choice for a lender in terms of risk, although not necessarily the most profitable. Credit companies profit out of two main sources: interest charged over revolving balances and fees charged to the establishments for assuming their clients' debts. This latter source, namely *credit card processing fees*, fluctuates with market, type of establishment, consumption category including some room for "loopholes," negotiation and markups. In the US, for example, these fees go from 1.5 % to 3% of the actual purchase plus some additional flat amount depending on the nature of the transaction. Always-payers represent only one source of profits for the credit card company through the processing fees. Clients that keep a revolving balance but pay eventually, are the most profitable source since they increase lender's profits through the two channels. In the lender's perspective a good combination of these two types is the most profitable. Finally, clients that default permanently represent the higher risk to the company and, if no debt renegotiation is in place, they may become effective losses. This last category is the kind of clients to avoid. The types identification offers an suitable alternative to improve over the adverse selection problem.[27]

It is common ground the fact that informational asymmetries and uncertainty generate market inefficiencies, probably beyond credit rationing, due to the presence of adverse selection problems and moral hazard. Reducing these market inefficiencies is an empirical challenge and the industry has undertaken important efforts to minimize the informational asymmetries, however there is still much room for improvement. The identification of consumers' types, that is to say the identification of a combination of client's characteristics that respond to different unobserved motivations towards decisions (*ceteris paribus*), allows for strategic solutions that foster market efficiency improvements.

4.4 Empirical approach

4.4.1 Data

For the types identification I use rich administrative data of actual credit cardholders. The data set belongs to anonymous administrative records of clients from a credit company in Ecuador in 2014. It contains individual historical records of credit behavior and demographics that include: age, marital status and household's family members. Job type information during the application process and profession,

[27]Ryan Guina (2011) (cashmoneylife.com) cites Barry Nalebuff for describing these three types of clients as *Maxpayers*, *Revolvers* and *Deadbeats* while including two other categories as the *Arbitragers* or those that get credit and earn interest in the market, a practice that is illegal in many countries; and the *Reformed Credit Card Users / Non-Users*; those that quit using credit cards after bad experiences.

available on the data, are used to infer variables about education and responsibility conditions.

One traditional criticism in the use of financial data for scoring purposes is the fact that, inevitably, information is available only individuals already selected in the screening process to enter the credit market; not from a random sample of clients. As Hand and Henley (1997) argue, this is a widespread practice and the inference derived, referred as *reject inference*, depending on the characteristics of the data set, is reasonable if "the new score-card (based on current clients) is based on a super set of the characteristics used in the original score-card (used in the admittance selection process)."[28] Whether or not this creates a big *selection bias* is an open empirical question, furthermore; some of the limitations in the types identification could result from this lack of information if excluded clients correspond to a particular *type* untraceable in the estimation process. The strictest the admittance rules of a credit company, the higher the bias. The company that provided the information has a fairly reasonable selection process, and according with personal interviews with credit analysts, current selection policies have relaxed some of the previous years' requirements. Once accounting for age restrictions (people over 18 years or older), and a particular requirement related to *college degree* for young professionals, the selection process is based on general signals of creditworthiness that are common practice in finance: employment and banking history, salary and assets.[29] Without

[28] My own notes in parenthesis.

[29] According to Chatterjee et al. (2005), FICO credit score is based mainly on payment history (35%), amounts owned (30%) and length of credit history (15%). This type of score lacks a lot of the information available in financial markets. Access to some of this information for creditworthiness assessment is prohibited by law in the US since the seventies (Equal Credit Opportunity Act and Fair Housing Act, 1976), on the grounds of potential prejudice and discrimination over: age, race, color, religion, national origin, gender, marital and family status (number of children) and whether or not you receive public assistance; while other information is traditionally confidential either for personal, institutional or identity protection: age, assets, salary, occupation and employment history. Availability and use of the data for credit underwriting or screening depends on the standing legal framework, different country wise; yet, more information allows for

overlooking this aspect, traditional scores are based on available data from current clients and this paper follows this path.

The data set has up to 171,044 observations randomly partitioned in two samples. 60% of the observations to be used in the estimation model, or *control sample*, and 40% to be used in the out-of-sample analysis, *treatment sample*. 41% are women, the overall average age of cardholders is 45, 33.5% have a college degree (according to profession) and the average income is USD 1840.83. Around four thousand observations (2.6%) were eliminated due to inconsistencies.[30]

Table 4.1: Median income, net-worth and debt by client's status (USD)

	Overall sample				Default Cardholders			
	Income	Net-Worth	Tot.CC.Debt	Balance/Income	Income	Net-Worth	Tot.CC.Debt	Balance/Income
By age								
age<35	1300.00	0.00	2402.92	1.05	1286.00	0.00	2980.12	1.54
age $35-50$	1578.00	20000.00	4054.87	1.19	1500.00	16000.00	4776.75	1.55
age $51-65$	1700.00	50000.00	3998.38	1.31	1500.00	45000.00	4577.62	1.67
age>65	1500.00	64241.50	2606.11	1.12	1500.00	59200.00	3077.10	1.57
By income quintiles								
Q1	900.00	10000.00	2271.25	1.07	900.00	9000.00	2753.70	1.38
Q2	1200.00	9000.00	2687.43	1.21	1200.00	7000.00	3342.81	1.58
Q3	1500.00	15000.00	3179.95	1.18	1500.00	11800.00	3944.05	1.63
Q4	2000.00	28500.00	4215.68	1.22	2000.00	21995.00	4995.29	1.64
Q5	3000.00	50000.00	6457.17	1.22	3000.00	42082.00	8224.78	1.76
By education								
High school	1500.00	25000.00	3647.72	1.21	1500.00	18000.00	4269.45	1.58
College	1500.00	16000.00	3123.79	1.12	1400.00	12000.00	3654.91	1.57
By gender								
Male	1600.00	25000.00	3636.47	1.15	1500.00	17000.00	4165.00	1.53
Female	1400.00	17990.00	3235.47	1.22	1319.00	13500.00	3935.20	1.62
By civil status								
Single	1300.00	0.00	2622.98	1.01	1300.00	0.00	3180.62	1.51
Married	1600.00	31000.00	3874.68	1.24	1500.00	25000.00	4471.17	1.58
Widower	1400.00	55000.00	2821.53	1.30	1200.00	45514.50	4101.23	1.91
Divorced	1500.00	29000.00	3775.57	1.31	1486.50	25000.00	4429.07	1.76

Balance/Income is the ratio of monthly pending balance / monthly income.

more accurate assessments, and possibly a more efficient allocation.

[30] Inconsistencies include: reporting less than USD 700 salary (misreporting, according to credit analysts), younger than 20 years old, more than 7 family members and missing data for assets and payments.

The main analysis concentrates on identifying the types of clients on *default* behavior for those with pending balances of over 30 days (10.31% of the sample). This is a relatively stringent condition, however it captures a credit card delinquency behavior that goes beyond possible payment mistakes. Because of the nature of the sample it allows us a good compromise towards the applicability of the estimation procedure.[31]

The second part of our results extends the application to the credit demand and types identification. Using the information from individual clients I analyze their extended credit card demand from all sources available, i.e. total credit card demand.

4.4.2 Empirical model

4.4.2.1 Default probability

Estimation of the *default probability* is a traditional latent variable problem where we only observe whether an individual has defaulted or not, and not the actual probability; hence, the dependent variable is a dummy for $D = \mathbb{1}\,[\text{balance} > 30\text{days}]$. Direct estimation of binary outcome models, either by OLS, Probit or Logit strategies, inevitably disregards biases derived from ignorance of the individual preferences, i.e. ignorance of types in our framework. More generally, the default behavior (outcome) might be the results of two different data generating processes non identifiable by traditional techniques.

To account for the unobserved individual heterogeneity we come back to the

[31]Scores can consider different periods for credit card delinquency. In general, one day default implies an individual misses his minimum payment over the consumption of the last cycle (31 days). Although missing one payment can be an involuntary mistake, 30 days after it certainly involves some deliberate action. According to interviews with the credit analyst, message warnings and telephone calls start from the second day of delinquency and legal collection activities typically start after 90 days, however they can start as soon as the second delinquency day, depending on the history of the client.

types characterization described previously in the theoretical section. Assume the existence of J types and define the unconditional probability of a client of being from a particular type as $Pr(\theta^j = \theta^1) = p^1$, $Pr(\theta^j = \theta^2) = p^2$,..., $Pr(\theta^j = \theta^J) = (1 - \sum_{j=1}^{J-1} p^j)$.

From equation 5 we know that a client of a particular type j defaults if its second period consumption is lower than his bankruptcy threshold, hence we have:

$$D = \mathbb{1}\left[\tilde{u} + \theta^j - RL \leq Z_j\right] \tag{4.11}$$

Note that there are two sources of heterogeneity explicit in this expression (exogenous and unobserved). On one hand θ^j summarizes the heterogenity that comes from human capital factors that help explain individual's income (e.g. cognitive skills). On the other, Z_j represents those factors related to individual's expectation or preferences towards minimum consumption (e.g. keep consumption status). Loosely speaking, both sources are jointly capture in the individual's *quality-type* and are, in this framework, indistinguishable from each other. To simplify the empirical argument in this section and abusing notation, I summarize both in one only *quality-type* parameter $\theta^j = (\theta^j - Z_j)$.[32] Thus, the *conditional expectation* of default, i.e. the probability of default conditional on being a particular type, for a known *cdf* function $G(.)$, can be defined as follows:

$$E(D|\theta^j) = Pr(D = 1|\theta^j) = Pr(\tilde{u} + \theta^j - RL \leq 0) = G(RL - \theta^j) \tag{4.12}$$

By the Bayes' theorem, the *unconditional probability of default* for an individual

[32] An alternative interpretation, as in GM08, would be to assume that individuals have homogeneous minimum consumption expectations ($Z_j = 0$), reducing the problem to have positive net income on the second period.

i and J unobserved types is given by:

$$\begin{aligned} Pr(D_i = 1) &= \sum_{j=1}^{J} Pr(D_i = 1, \theta_i = \theta_i^j) \\ &= \sum_{j=1}^{J} Pr(D_i = 1 | \theta_i = \theta_i^j) Pr(\theta_i = \theta_i^j) \end{aligned} \tag{4.13}$$

$$Pr(D_i = 0) = \sum_{j=1}^{J} \left[1 - Pr(D_i = 1 | \theta_i = \theta_i^j) Pr(\theta_i = \theta_i^j)\right] \tag{4.14}$$

The default probability of an agent in this setting depends directly on two general aspects: the contract choice given the conditions in the $\langle L, R\rangle$ space, and the intrinsic *quality-type* of the client. Taking into account the client's (observable) information available, we can approximate the probability in equation 12 by a linear function of the individual characteristics related to creditworthiness (X_i) or observed credit behavior and loan conditions (L_i). Hence, we have that:

$$G(\bar{R}L^j - \theta^j) \approx G(\varphi + X_i\beta_1 + L_i\beta_2) \tag{4.15}$$

A similar approximation is assumed for the unconditional type probability where we introduce characteristics identified (somewhat arbitrarily) as correlated to the individual preferences, among others. On the next section we extend the discussion about this point. We let the function for types be a function of *observed* factors (Z_i) that help identify the individual's type e.g. individual characteristics, human capital information, experimental measures for individual preferences and other related covariates:

$$\theta_i^j \approx Z_i\gamma_j + \upsilon_i \tag{4.16}$$

Again, if a well behaved *cdf* function $F_i(.)$ exists for v_i, then the unconditional probability is a function of the available information for types, hence:

$$Pr(\theta_i = \theta_i^j | Z_i) \approx F(Z_i \gamma_j) \tag{4.17}$$

Finally, we replace these equations on the default/non-default probability equations 13 and 14, and get:

$$Pr(D_i = 1) = \sum_{j=1}^{J} G(\varphi^j + X_i \beta_1^j + L_i \beta_2^j) F(Z_i \gamma_j) \tag{4.18}$$

$$Pr(D_i = 0) = \sum_{j=1}^{J} \left[1 - G\left(\varphi^j + X_i \beta_1^j + L_i \beta_2^j\right) F\left(Z_i \gamma_j\right)\right] \tag{4.19}$$

Two things are worth noting about the equations. On one hand, the estimated coefficients may vary by type, i.e. we allow for complete flexibility of the coefficients, hence; we should expect heterogeneous elasticities on the probability, conditional on the type.[33] Another important aspect of the statistical procedure is that the set of variables described in X_i and L_i do not need to be fully disjunctive with respect to Z_i, however for partial identification some *exclusion restrictions* are needed. Without further arguments there is certain level of arbitrariness about the information used in each set of variables. Similar to Gan and Mosquera (2008), I include socio-demographic variables such as age, gender, number of children and marital status. Additionally, I extend the identification to labor and education information of potentially higher relevance for the type identification. Finally, I also include an instrument for whether the client enter the credit market during periods of interest

[33] GM08 present the same model for the two types case. I expose this extended version for generalizability, however, due to identification restrictions, the empirical application reduces to the same two types for the probability model.

rate volatility versus a more recent price stability. [34]

We can easily derive the Maximum Likelihood function directly from the probability equations for a Bernoulli distribution. Then, for a *type-consistent* estimation for a subject i we maximize the corresponding *log-likelihood* function:

$$\ell = ln(L) = \sum_{i=1}^{n} D_i ln\left[Pr(D_i = 1)\right] + (1 - D_i) ln\left[Pr(D_i = 0)\right] \tag{4.20}$$

4.4.2.2 Credit demand

Credit demand, same as default decisions, is an intrinsic choice problem for an agent and, as suggested in the theoretical section, such choices are also type contingent. Traditional supply and demand analysis require a good identification for the simultaneous equation problem, e.g. using instrumental variables that help distinguish an exogenous source of variation of prices (Gross and Souleles 2002, Grant 2007, and Crook et al. 2007). Interest rates in our case are relatively constant in the credit market (sample from Ecuador) for the consumer segment. Although there is some heterogeneity of r, competition by prices is not very active. Furthermore, differently from the US, within a particular credit card company, clients face the same consumption interest rate; normally the legal maximum (see Figure 2).

Given little to non variation in interest rates, elasticities estimation is constrained. However, conditional on having a credit card debt ($debt > 0$) to capture the *type-consistent* coefficients of the credit demand we rely on a more *mainstream* finite mixture model were, opposed to our previous approach, has as a continuous dependent variable, the logarithm of *total credit card debt* registered on the Credit Risk Bureau, i.e. credit card debt accumulated over all the client's credit cards. Given

[34] Gan, Hernández and Liu (2013), when analyzing group types, used variables such as literacy, land ownership, housing conditions, occupation and caste (for rural India).

that types are not *ex-ante* observed by the lenders, significant types variation, I argue, can only come from the credit demand behavior.

The estimation process, for J types, is similarly performed through maximum likelihood of the following function:

$$\ell = \sum_{i=1}^{n} \left\{ ln \left[\sum_{j=1}^{J} p^j(Z_i\gamma_j) g_j \left(y_i | \varphi^j + X_i\beta_1^j + L_i\beta_2^j \right) \right] \right\} \tag{4.21}$$

Similar to our previous approach, $g(.)$ could take any functional form.

I present results based on $g(.)$ being a normal function (standard normal in the case of G for the default probability); however, changes in functional form, such as an alternative *logit* function, do not alter the inference.

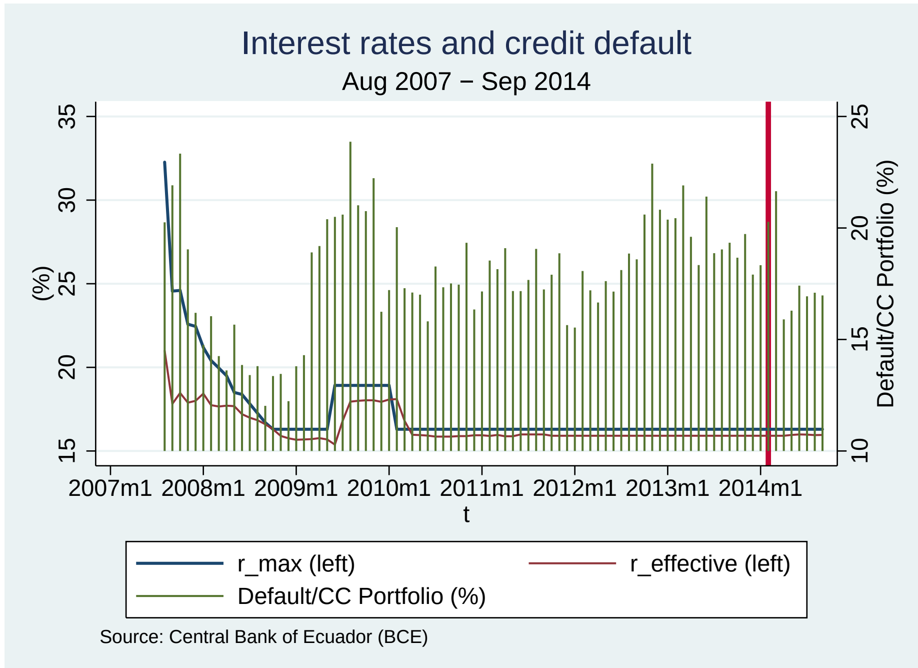
Finally, one common adjustment to ensure the probability of the types belongs to the $(0,1)$ interval, and $\sum_j p^j = 1$, is to use the multinomial logit form as follows (equivalent to the $F(.)$ distribution in the default equation):

$$p^j(Z_i\gamma_j) = \frac{\exp(Z_i\gamma_j)}{1 + \sum_{l=1}^{J-1} \exp(Z_i\gamma_j)} \tag{4.22}$$

4.4.3 Identification problem revisited

The estimation procedure proposed offers a suitable alternative for more consistent estimation of structural parameters in the credit market analysis, however, such advantages are not free of criticism. This approach is more flexible to traditional methods, particularly compared to those in Industrial Organization. Although I still rely on a parametric distribution for the choice equation, there is no a priori assumption about the distributional form of the unobserved heterogeneity but a more parsimonious definition over the number of support points of an unknown distribution, i.e. the number of types (Heckman and Singer 1984, and Hess et al. 2011).

Figure 4.2: Evolution of market interest rates (annual) and credit card default rate in Ecuador.



This flexibility, however, has some caveats, in particular, the linear approximations of the types and default probability could suffer from endogeneity problems from omitted variables (OVB) or the classical measurement error (CME) problem.

In the default approximation, if θ^j is unobserved then we will be under OVB if $Corr(\theta^j, X_i) \neq 0$ and $Corr(\theta^j, L_i) \neq 0$. There is plenty of evidence in the experimental literature that shows how individual preferences, in particular, risk and time preferences are correlated with financial decisions, educational choice and human capital accumulation, job search behavior, cognitive skills, among others. Furthermore, there is also evidence of *biological and neurological* determinants of individual preferences that may influence actual behavior. Some examples of this latter issue include Apicella et al. (2008) that find some correlation levels between testosterone levels and risk taking behavior, as well as Ramaswami et al. (1993) that relate partial suppression of brain areas and its relationship to behavior and decision making process.

A way to overcome this informational limitation is to use measures that identify specific aspects of individual preferences and use them as *proxies* in the basic estimation procedure. There are several elicitation mechanisms for risk and time preferences readily available for such endeavor and the possibility of using such information would help identify with more certainty the nature of the type identification, as oppose to a general *quality-type*. Nevertheless, such applications, specially on the field, are not cost/effective, specially taking into account that in experimental economics these mechanisms involve actual payments.[35]

A second best alternative are *survey-type* measures. Although they are not necessarily monetarily incentivized (see Dohmen et al. 2011) they can offer some revealing information over the nature of the preference phenomena and individual types.

To see this more clearly, assume the types approximation include (linearly) unobserved factors in W_i:

$$\theta_i^j = Z_i\gamma_j + W_i\lambda_j + \upsilon_i$$

θ_i^j is unobserved by the econometrician, from the information in the linear approximation for types and in the default probability, a fully characterized estimation would result from:

$$G(\bar{R}L - \theta^j) \approx G(\varphi + X_i\beta_1 + L_i\beta_2 + Z_i\gamma_j + W_i\lambda_j)$$

[35]In experimental economics, the common practice is to apply monetarily incentivized mechanisms since there is a need to argue confidently over the salience of the payment to reveal actual economics decisions. In psychology this is not necessarily the case and hypothetical measures, although noisier, might be used instead (Kahneman and Tversky 1979).

Ignorance of W_i affects estimation in two ways. First, if $Corr(Z_i, W_i) \neq 0$, then we would incur in the CME problem and estimation of γ_j should suffer from attenuation bias. Second, if $Corr(W_i, X_i) \neq 0 \quad \wedge \quad Corr(W_i, L_i) \neq 0$, then again, we fall into the persistent OVB. Including *proxies* of individual preferences would not solve entirely the problem since we can only include a limited and usually incomplete information set (Gan et al. 2011). However, the types identification *"solves"* the problem by letting *all unobserved factors (unobserved heterogeneity)* be captured by the specification of the number of types included in the model, i.e. determine the level J for the parameter θ. I call this *identification beyond functional form* in this paper.

Identification assumption: Conditional on the types, factors (covariates) that affect the default probability are independent; in other words, factors that characterize the types affect the default probability (and credit demand) only through the type probability (Gan et al.2011).[36]

The identification strategy is intuitive. On one hand, the types determination captures the unobserved variation as a sufficient estimator of potential confounders (individual preferences). At the same time, when additional information (W_i) is available, similar to the excluding restriction on a 2SLS identification, we require a strong correlation of W_i with respect to the types probability (first stage), while not being an explicit determinant of the default probability.

[36] Henry, Kitamura and Salani (2014) discuss the partial identification problem of mixture models in a similar fashion. The "exclusion restriction" for them discussed can be translated to the orthogonality condition of W_i with respect to the outcome variables (in our case: default probability D_i, and credit demand.), conditional on the types and other regresors.

4.5 Results

4.5.1 The probability of default

The preferred specification (summarized in Table 3) reports significant differences in the parameters estimated for the two *quality-types* of clients in the credit market. Most of all, there are significant differences from the baseline models estimated (OLS, Probit or Logit in Table 2). I interpret this as evidence of presence of private information in the credit market and as a consequence, different behavioral responses (strategic behavior) in credit card holders. Furthermore, the baseline model estimates show very significant coefficients while the *type consistent* model shows that the influence of the covariates is heterogeneous conditional on the types, hence statistical significance in most cases is driven by only one of the types and not both at the same time.

Lets first look at the probability of type identification. The type identification comes from the subset of variables included in the Type equation. The model specification relies on the fact that if a type-identifying variable affects the default rates, it does so only through the type probability, hence the estimated coefficients are not directly traceable back to the baseline coefficients. We use a fairly wide set of variables for the type identification that include education and labor information, not previously analyzed. I find evidence that the type heterogeneity is weekly related to the type of job of the credit holders, in particular credit card holders with high or medium responsibility jobs are significantly more likely to belong to the *high-quality-type*. Other job descriptions available, like military or police jobs, are not significantly informative.

Interestingly, age is a very informative factor. The older an individual is, the more likely it falls into the high-quality-type category. At the same time, significance of

Table 4.2: Default probability and marginal effects, baseline models

	Default = 1 [age of portfolio >=30 days (implicit default)]				
	(1) OLS	(2) Probit		(3) Logit	
	β / SE	β / SE	Mfx	β / SE	Mfx
Demographics					
Age(in years)	−0.0001	−0.0015	−0.0002	−0.0042	−0.0002
	(0.0003)	(0.0025)		(0.0047)	
Age squared	0.0000***	0.0001***	0.0000***	0.0001***	0.0000***
	(0.0000)	(0.0000)		(0.0000)	
Woman=1	0.0043**	0.0481**	0.0050**	0.0925**	0.0049**
	(0.0019)	(0.0212)		(0.0430)	
Married=1	−0.0025*	−0.0090	−0.0009	−0.0305	−0.0016
	(0.0014)	(0.0120)		(0.0230)	
Widower=1	0.0039	0.0376	0.0039	0.0719	0.0038
	(0.0051)	(0.0498)		(0.1083)	
Divorced=1	−0.0042**	−0.0125	−0.0013	−0.0376	−0.0020
	(0.0021)	(0.0183)		(0.0374)	
College(pf)=1	−0.0019	−0.0090	−0.0009	−0.0290	−0.0015
	(0.0020)	(0.0154)		(0.0345)	
Science degree	0.0069***	0.0565***	0.0059***	0.1207***	0.0065***
	(0.0019)	(0.0152)		(0.0324)	
H/M Responsibility(wd)=1	−0.0045**	−0.0282*	−0.0030*	−0.0656*	−0.0035*
	(0.0019)	(0.0163)		(0.0377)	
Mltry/Police(pf)=1	−0.0035	−0.0179	−0.0019	−0.0316	−0.0017
	(0.0033)	(0.0317)		(0.0619)	
Family Members (#)	0.0016**	0.0161***	0.0017***	0.0351***	0.0019***
	(0.0007)	(0.0061)		(0.0133)	
Main cities					
Quito	−0.0078***	−0.0670***	−0.0070***	−0.1294***	−0.0069***
	(0.0021)	(0.0191)		(0.0391)	
Guayaquil	−0.0010	0.0027	0.0003	0.0202	0.0011
	(0.0021)	(0.0187)		(0.0379)	
Cuenca	0.0023	0.0171	0.0018	0.0158	0.0008
	(0.0019)	(0.0167)		(0.0342)	
Ambato	0.0028	0.0052	0.0005	0.0209	0.0011
	(0.0018)	(0.0162)		(0.0334)	
Economic variables					
Income(USD)	−0.0000	−0.0000***	−0.0000***	−0.0000***	−0.0000***
	(0.0000)	(0.0000)		(0.0000)	
Tot. Net Worth (USD)	−0.0000**	−0.0000**	−0.0000**	−0.0000**	−0.0000**
	(0.0000)	(0.0000)		(0.0000)	
Tot.CC Debt/Income	−0.0084***	−0.0353***	−0.0037***	−0.0704***	−0.0038***
	(0.0014)	(0.0098)		(0.0197)	
% Debt at Risk on CRB	0.2927***	1.1233***	0.1177***	2.0968***	0.1122***
	(0.0132)	(0.0523)		(0.0921)	
Ownerwhip Type	−0.0048***	−0.0302**	−0.0032**	−0.0680**	−0.0036**
	(0.0016)	(0.0153)		(0.0301)	
Properties (#)	−0.0023	−0.0237	−0.0025	−0.0470	−0.0025
	(0.0027)	(0.0264)		(0.0572)	
Houses(#)	0.0029**	0.0223*	0.0023*	0.0426*	0.0023*
	(0.0014)	(0.0114)		(0.0244)	
Cars(#)	0.0002	0.0060	0.0006	0.0108	0.0006
	(0.0011)	(0.0107)		(0.0229)	
Debts on CRB(#)	−0.0028***	−0.0079***	−0.0008***	−0.0121**	−0.0006**
	(0.0004)	(0.0026)		(0.0060)	
Returned Checks=1	−0.0028	0.0537	0.0056	0.0756	0.0040
	(0.0087)	(0.0466)		(0.1048)	
Ytd.Avrg.Payments/Cr.Balance	−1.3268***	−5.8248***	−0.6103***	−10.6474***	−0.5696***
	(0.0224)	(0.1325)		(0.2713)	
Microcredit=1	0.0048	0.0218	0.0023	0.0472	0.0025
	(0.0037)	(0.0264)		(0.0543)	
Portfolio/Tot.Debt	−0.0497***	−0.4345***	−0.0455***	−0.9617***	−0.0515***
	(0.0057)	(0.0350)		(0.0654)	
Mth.Bal.Payment/Income	0.1099***	0.6545***	0.0686***	1.2792***	0.0684***
	(0.0093)	(0.0515)		(0.0917)	
Time of membership (months)	−0.0000***	−0.0003***	−0.0000***	−0.0005***	−0.0000***
	(0.0000)	(0.0001)		(0.0001)	
Constant	1.3326***	3.7486***	***	6.9263***	***
	(0.0249)	(0.1400)		(0.2813)	
Observations	102657	102657		102657	
R^2	0.384				
Pseudo R^2		0.401		0.397	
AIC	−3249.499	40723.659		41023.645	
BIC	−2991.942	40981.216		41290.741	

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors in parenthesis, clustered at city levels.
Estimations are based on the control sample for the study.
Marginal effects (Mfx) calculated at the median.

the nonlinear age function (age squared) suggest evidence of behavior that follows the life cycle hypothesis where the quality relationship (related to the credit demand) has a saturation point. Figure 3 offers more evidence of the life cycle patterns. While average income keeps a more stable relationship in older ages (3a), credit card demand decreases significantly around retirement age of 65 (3b).

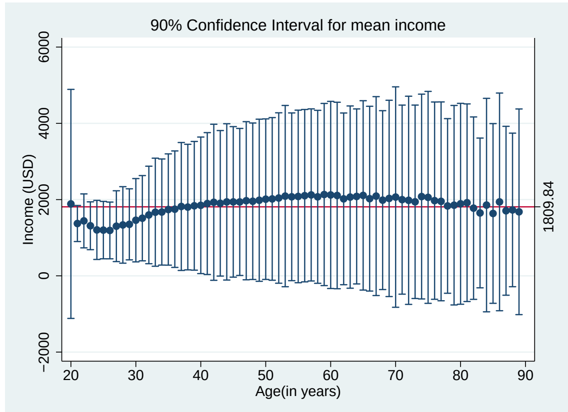
Another factor that significantly explains the type identification is the number of family members of the credit card holder. Whether an individual is the type that builds or belongs to large families, the more likely he falls into the *low-quality-type*. Causality in such statement is troublesome, evidently the higher the household needs the higher the debt burden, or the higher the bankruptcy threshold (defined in the model), hence the higher the incentives to default. Nevertheless, it is a defining characteristic that significantly helps explain the types, thus improves the class sorting. Something similar occurs with the client's gender. In our estimation results, although with lower significance (10%), women are less likely to belong to the *high-quality-type*. In our sample, the median women's income is slightly lower than their male counterparts; around USD 200.00 and USD 280.00 difference (delinquent clients included). However, differences are steeper when it comes to net-worth, between USD 3500.00 and USD 7000.00 difference. Interestingly, although men tend to hold higher credit card balances overall, when looking only the delinquent accounts, the ratio of balance/income is practically the same by gender (see Table 2). Income per se is not a significant factor on the type equation, however, it is still possible that gender differences on quality-type are driven by lags on the effects of the gender income-gap. In the U.S. the Equal Credit Opportunity and Fair Housing Act (1976) forbids the use of gender information for credit scoring. After its enactment it is claimed that gender differences disappeared (Bertola et al.2006). In our sample, although we can not say much towards the source of these differences, it is clear that they are persistent and help explain partially the types differentiation.

Finally, in the search for instruments that allow for the exclusion restriction to work, I included a dummy variable for whether the subject enter the credit market during a period of interest rate stability. Figure 2 shows changes in interest rates in the market (effective and legal maximum). Interest rate controls were imposed in Ecuador during years 2009 and 2010. They act as an exogenous shock in the market. Beyond account age (and possibly an adaptive institutional process), there are no significant differences between clients separated in both periods (on observables). The hypothesis is that attitudes towards credit are different for clients that faced periods of interest rate uncertainty. A complete theoretical treatment of such argument is out of the scope of this study; nevertheless, I let explicit the estimation results for future exploration. Clients that enter the credit market during a period of interest rate certainty are more likely to be of high-quality-type, however the statistical significance is weak (10%).

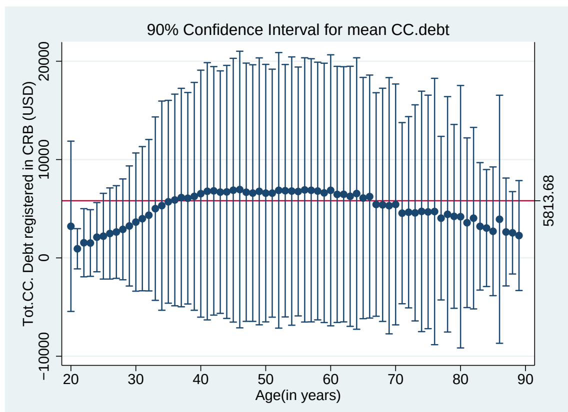
I do not find significant evidence of type differentiation due to any other factors. Without further exploration it is hard to interpret what *quality-types* represent, however; accounting from insights over experimental evidence, I argue that the relationship found about individual preferences in experimental studies holds on this results. In experimental evidence, the age is negatively correlated with present bias (instant gratification), higher time discounting (impatience) and higher risk aversion; accounting for age into the type identification would imply that a *low-quality-type* of client falls better into that description. Something similar can be said about measures of financial responsibility (Eckel, Johnson and Montmarquette, 2005) or financial literacy. Individuals that score higher on those measures tend to have better foresight over consumption planing (consumption smoothing). I also argue that the identification of responsibility jobs suits that argument (although not cleanly)

Figure 4.3: Life cycle patterns for income and debt

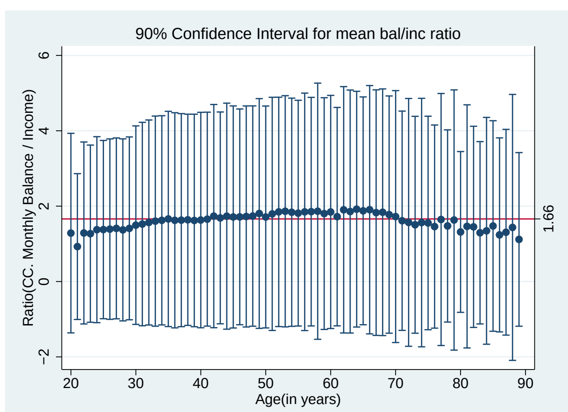
(a)



(b)



(c)



if we are willing to believe that such responsibility is a signal of some financial or managerial skills. In particular, higher job-responsibility clients are less likely to be a *low-quality-type*. Further analysis over these preferences should come from specific measures that are certainly orthogonal to the credit behavior, such information is not currently available and efforts are being developed to include more specific identification.

When looking at the Choice equation, I observe that, other things equal, *low-quality-type* clients are more likely to default. This relationship is clear from the constant in the model which captures the conditional outcome differentiation. In general, there are important behavioral differences between types when it comes to equity versus liquidity sensitivity. Although negligible in magnitude, high-quality-types reduce the probability of default when current income increases. This can be evidence of strategic behavior; while at the same time, changes in net-worth are more significant for low-quality-types. Furthermore, the higher the current monthly balance the client holds, increases the probability of default for both types but it is relatively higher for the high-quality-type, while the opposite happens with the average monthly payments, higher payments reduce the probability of default greatly for low-quality-types.

There is also evidence of presence of moral hazard when types are identified in the market. For example, if a low-quality-type becomes a business owner, his probability of default increases significantly, as evidence of riskier decisions involved. Business ownership has the opposite sign for high-quality-types, however its effect is not statistically significant. Something similar occurs when it comes to credit access. The higher the number of debts a client holds on the credit risk bureau the higher the probability of default for low-quality-types, while the opposite holds for the high-quality-types. In the same line, the higher the total debt/income ratio, the lower the probability of default of high-quality-types. Both pieces of evidence push the same idea: the availability of credit resources affect heterogeneously the probability of default of credit card users conditional on their types, something not captured in traditional default analysis techniques.

A more intuitive approach can be captured through the marginal effects, i.e. how much the default probability is affected by a change of one unit on the relevant regresor. Table 4 summarizes the marginal effects for two versions of the Type Consistent model, abusing previous notation we described them as follows:

- Type-consistent "unconditional" approach:

$$Pr(D_i = 1) = \underbrace{p^1(Z_i\gamma_1)}_{\text{Type equation}} \overbrace{G(\varphi^1 + X_i\beta^1)}^{\text{Choice equation}} + p^2(Z_i\gamma_2)G(\varphi^2 + X_i\beta^2)$$

- Type-consistent "conditional" approach:

$$Pr(D_i = 1) = \begin{cases} G(\varphi^1 + X_i\beta^1) & , \text{if } \hat{\pi}_{\tilde{\theta}^1} > \hat{\pi}_{\tilde{\theta}^2} \\ G(\varphi^2 + X_i\beta^2) & , \text{if } \hat{\pi}_{\tilde{\theta}^1} < \hat{\pi}_{\tilde{\theta}^2} \end{cases}$$

  where the posterior probabilities are calculated as,

$$\hat{\pi}_{\tilde{\theta}^j} = \frac{p^j(Z_i\gamma_j)G(\varphi^j + X_i\beta^j)}{\sum_{j=1}^{J} p^j(Z_i\gamma_j)G(\varphi^j + X_i\beta^j)}$$

Evaluated at the median levels for a "typical" client, and considering those regresors that are statistically significant; interestingly, the economic variable that captures a great deal of the default probability, and that shows heterogeneous effects by type of client, is the ratio of the Year-to-date Average Payment over the pending credit balance. The higher the average ratio, the lower the default probability by around 22%. High quality types reduce their probability greater by around 16 percentage points over lower types. In the same line, an increase in the ratio of pending debt (Portfolio) over total debt reduces the probability of default by around 4.4% for high quality types, a reduction over 4 percentage point higher when compared to the average effect of lower types.

I describe in this section results for a two-types model. A natural extension, as suggested on the theoretical section, would be to empirically identify the presence of higher types. There is no definitive answer as to define a particular number of types, in general, there is a trade off between a finer classification and the model predictability. Henry et al. (2014) suggest the empirical limitation of partial identification that comes from the outcome variable support, hence for a binary outcome model, the feasible number of types is limited to two.[37] To explore more in depth a possible finer classification I rely on the credit demand side of the signaling process.

4.5.2 Default analysis: out-of-sample performance

From a practical perspective, the most important application of any credit scoring technique is to achieve a useful stratification of clients as to allow the implementation of different price strategies and access, restrictions, advertisement policies, depending on the goals of the principal (lender). Great efforts in the credit market industry are destined to this goal, always taking into account the trade-off between the lender's credit risk profile and its profit maximization in any transaction.

Out-of-sample predictions for either credit access or behavior analysis, i.e. when the subject is already a client, greatly improve in the approach presented of a finite mixture density estimation. When compared to traditional techniques (baseline mod-

[37]See Henry et al. (2014), lemma 2.

Table 4.3: Default probability: type consistent model

	Baseline models		Type Consistent model	
	Logit	**Probit**	**Low Type**	**Hight Type**
Type equation (demo.)				
Age (years)	-0.0115	-0.00526	-	0.0819***
	(0.00770)	(0.00385)	-	(0.0130)
Age squared	0.000210***	0.000102***	-	-0.000856***
	(0.0000782)	(0.0000389)	-	(0.000128)
Woman=1	0.0916***	0.0474***	-	-0.122**
	(0.0284)	(0.0143)	-	(0.0491)
Married=1	-0.0338	-0.0105	-	0.0606
	(0.0348)	(0.0176)	-	(0.0601)
Widower=1	0.0653	0.0336	-	-0.327*
	(0.115)	(0.0571)	-	(0.196)
Divorced=1	-0.0457	-0.0153	-	0.0836
	(0.0580)	(0.0291)	-	(0.0985)
College (pf)=1	-0.0265	-0.00739	-	-0.0809
	(0.0377)	(0.0190)	-	(0.0624)
Science degree=1	0.120**	0.0558**	-	0.0201
	(0.0483)	(0.0243)	-	(0.0816)
High/Medium Job Responsibility	-0.0563	-0.0235	-	0.237***
	(0.0361)	(0.0182)	-	(0.0594)
Mltry/Police force	-0.0302	-0.0177	-	-0.149
	(0.0773)	(0.0384)	-	(0.128)
Family Members (\#)	0.0432***	0.0198**	-	-0.157***
	(0.0163)	(0.00822)	-	(0.0248)
Stability 'r'-period	-0.103***	-0.0514***	-	0.136**
	(0.0395)	(0.0200)	-	(0.0587)
Income(USD)			-	-0.0000283
			-	(0.0000219)
Constant			-	-0.815***
			-	(0.314)
Choice equation (econ.)				
Income(USD)	-0.0000356**	-0.0000191***	-0.0000752	-0.0000242**
	(0.0000140)	(0.00000698)	(0.000110)	(0.0000105)
Tot. Net Worth (USD)	-0.000000813**	-0.000000412**	-0.0000712***	-0.000000372
	(0.000000365)	(0.000000180)	(0.0000189)	(0.000000266)
Tot.CC Debt/Income	-0.110***	-0.0548***	-0.176*	-0.112***
	(0.0113)	(0.00563)	(0.0970)	(0.00878)
Debt at Risk on CRB (%)	2.109***	1.129***	0	0.276***
	(0.0797)	(0.0431)	(.)	(0.0698)
Ownership Type	-0.0615	-0.0270	0.456***	-0.00359
	(0.0375)	(0.0188)	(0.153)	(0.0245)
Properties (\#)	-0.0514	-0.0262	0	0.00723
	(0.0449)	(0.0225)	(.)	(0.0332)
Houses(\#)	0.0311	0.0161	0.689	0.0371*
	(0.0310)	(0.0155)	(0.453)	(0.0225)
Cars(\#)	0.00337	0.00219	-11.76	0.0242
	(0.0224)	(0.0112)	(1165.0)	(0.0167)
Debts on CRB(\#)	-0.0203**	-0.0119**	0.194***	-0.0561***
	(0.00905)	(0.00462)	(0.0436)	(0.00751)
Returned Checks ($=1$)	0.0647	0.0481	0.959**	0.0313
	(0.0791)	(0.0406)	(0.483)	(0.0608)
Ytd.Avrg.Payments/Cr.Balance	-10.74***	-5.869***	-12.66***	-6.241***
	(0.106)	(0.0554)	(1.546)	(0.0819)
Microcredit($=$1)	0.0461	0.0209	-0.0498	-0.00660
	(0.0597)	(0.0311)	(0.271)	(0.0482)
Portfolio/Tot.Debt	-0.839***	-0.372***	-0.0324	-1.058***
	(0.0610)	(0.0303)	(0.214)	(0.0505)
Hist.Wgtd.Diferred/Tot.Debt	0.479***	0.234***	0.372	0.374***
	(0.0662)	(0.0332)	(0.275)	(0.0500)
Debt Refin. ($=$1)	-0.816	-0.375	0	-0.298
	(0.520)	(0.272)	(.)	(0.343)
Mth.Bal.Payment/Income	1.331***	0.679***	0.339***	1.021***
	(0.0289)	(0.0146)	(0.115)	(0.0250)
Time of membership (months)	-0.000931***	-0.000503***	-0.0491***	0.000276*
	(0.000250)	(0.000124)	(0.0112)	(0.000142)
Constant	7.066***	3.822***	10.44***	3.919***
	(0.207)	(0.105)	(1.541)	(0.0831)
Predicted type probability	1	1	0.15	0.85
Observations	102657	102657	102657	
AIC	40976.2	40681.2	35177.4	
BIC	41291.0	40996.0	35711.5	
Log Like.	-20455.1	-20307.6	-17532.7	
LR stat-test (pv) / wrt. TC	5844.8 (0.0)	5549.2 (0.0)	Reference	

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors in parenthesis (not-clustered).
Some coefficients are set to zero for estimation purposes (convergence).
Cities fixed effects (not included in the table) for Quito, Guayaquil, Cuenca, Ambato.

Table 4.4: Default probability: marginal effects (at median)

	Conditional on types				Unconditional	
	L-type (median)	**Mfx(L)**	**H-Type (median)**	**Mfx(H)**	**Overall (median)**	**Mfx**
Choice equation (econ.)						
Ownership Type	0	1.67%	0	-0.03%	0	7.18%
Returned Checks (=1)	0	6.23%	0	0.28%	0	7.21%
Microcredit (=1)	0	-0.10%	0	-0.06%	0	7.16%
Debt Refinanced (=1)	0	0.00%	0	-2.04%	0	7.21%
Properties (#)	0	0.00%	0	0.03%	0	0.03%
Houses(#)	0	0.51%	1	0.15%	1	0.13%
Cars (#)	0	-8.62%	0	0.10%	0	0.09%
Debts on CRB(#)	3	0.14%	2	-0.23%	2	-0.20%
Time of membership (months)	33	-0.04%	80	0.00%	76	0.00%
Ytd.Avrg.Payments/Cr.Balance	0.96	-9.27%	1.00	-25.90%	1.00	-22.49%
Tot.CC Debt/Income	1.12	-0.13%	1.19	-0.46%	1.18	-0.40%
Mth.Bal.Payment/Income	0.39	0.25%	0.42	4.24%	0.41	3.68%
Income (USD)	1500.00	0.00%	1500.00	0.00%	1500.00	0.00%
Total Net Worth (USD)	0.00	0.00%	25000.00	0.00%	22000.00	0.00%
Debt at Risk on CRB (%)	0.009	0.00%	0.00	1.15%	0.00	1.14%
Hist.Wgtd.Deferred/Tot.Debt	0.439	0.27%	0.47	1.55%	0.47	1.35%
Portfolio/Tot.Debt	0.231	-0.02%	0.17	-4.39%	0.18	-3.81%

Mfx: Marginal effects for regresor x.
Conditional results assigns the corresponding outcome based on the class assigned through the posterior probabilities.

els) the Type Consistent estimation, offers better overall predictions for the "treatment sample." The observed delinquency rate (default) in this sample is 10.4%. The model that better approximates it is the Type Consistent model in its unconditional version. The conditional approach, i.e. considering the parameters of each type once the client is classified by its posterior probabilities, is in general more rigorous, and although it offers a better selection of clients in terms of their "quality" it can overly restrict the profits that come from borrowers that even though default, end up paying the whole credit. These clients, as mentioned, are a dual source of profit, one through the use in a commercial establishment and also through the interest rates due to the payment delay.

Table 5 compares the predictive performance of the Type Consistent model at a

Table 4.5: Out-of-sample prediction performance

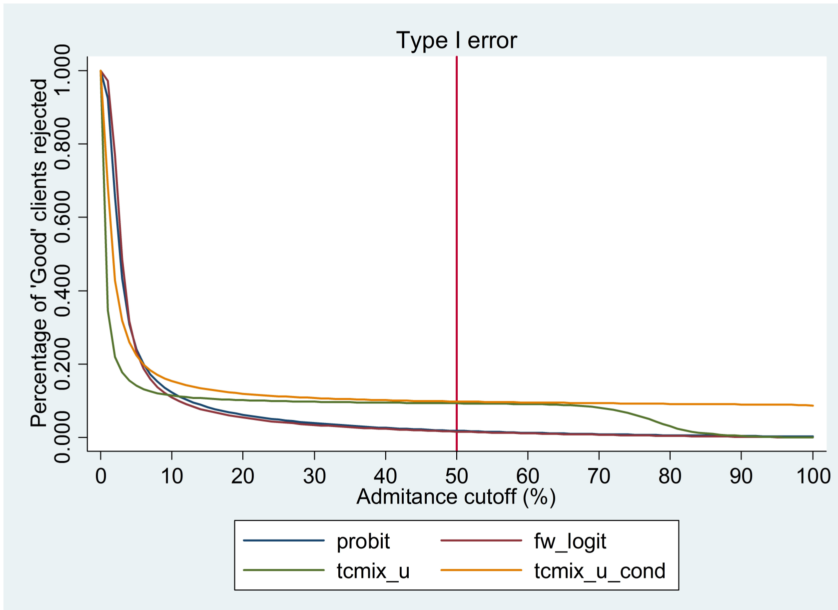
	OLS	**Forward Logit**	**Probit**	**TC-Unconditional**	**TC-Conditional**
Avg. pred. default probability (obs: 10.4%)	10.16%	10.2%	10.0%	10.3%	17.8%
Type I Error: Reject 'good' clients	1.3%	1.6%	1.9%	9.3%	9.8%
Type II Error: Accept 'bad' clients	61.4%	55.2%	55.8%	21.8%	19.3%
Predictive performance	92.5%	92.8%	92.5%	89.4%	89.2%
Ineficacy ratio	8.12	7.74	8.06	11.88	12.12

TC (Type Consistent)-Conditional assigns the corresponding probability based on the class assigned through the posterior probabilities.
Default defined over a probability higher than 50%.
Predictive performance: percentage of individuals correctly classified on a 2x2 ordering table.
Inefficacy ratio: wrong / right classifications.

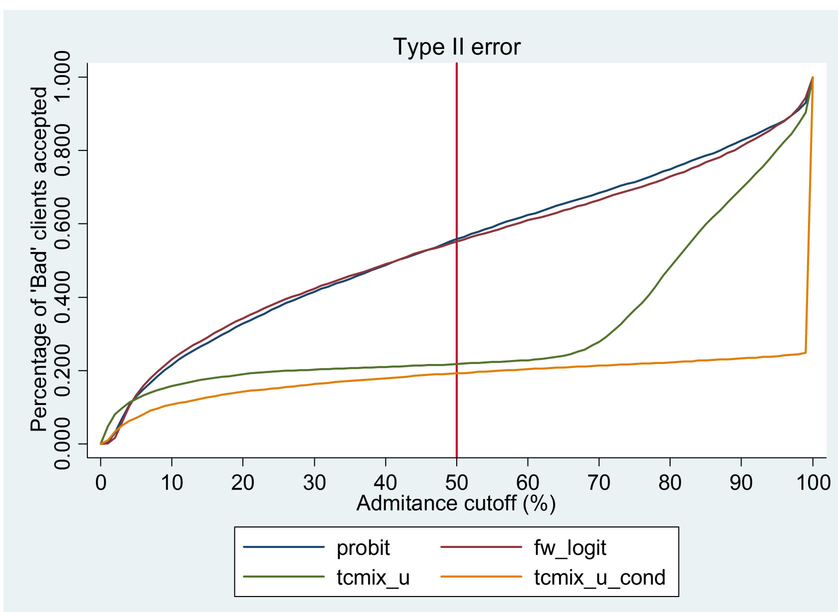
50% default rate cutoff ($Reject = \mathbb{1}[DefProb. > 0.5]$). The more rigorous the selection mechanism, the higher the credit rationing the market faces since more clients are excluded regardless of their actual quality or willingness to pay. The most rigorous specification is the conditional Type Consistent model, hence the probability to reject good clients, defined here as Type I error, is higher; nevertheless, the unconditional version remains on similar levels of the baseline models. More importantly, a relative measure of the adverse selection problem each model incurs is the Type II error, i.e. failing to reject clients that are qualitatively less creditworthy. The Type Consistent model greatly outperforms other models, regardless of the version used. The unconditional version of the Type Consistent model shows a probability of this error of 21.8%, while the conditional version reduces this probability to 19.3%, both less than half of the rate in the best baseline model (Stepwise Forward Logit). The trade off between predictive accuracy and selection rigurosity is evident in the results by looking at the predictive performance and the inefficacy ratio, although, their differences are small.

Figure 4.4: Lorenz curves for Type I and Type II errors: out-of-sample prediction performance.

(a)



(b)



A global way to observe the differences in performance of the models in the selection process is to simulate the percentage of excluded clients base on their default probability and over the whole range of potential acceptance cutoffs that a lender can implement. Similar to Gan and Mosquera (2008) and Gan et al. (2013), the Type Consistent estimation, in both variations, outperforms the baseline models and offers a better selection criteria for the relevant ranges of any scoring mechanism. Figure 4 plots the Lorenz curves for the probability of both errors and the baseline models used as reference. While on the Type I error, for admittance cutoffs higher than 8%-10%, the TC model performs worst, overall stochastic dominance can be observed for the Type Consistent model for the Type II error for most of the cutoff ranges.

4.5.3 Credit demand

When it comes to credit demand, the types classification is easily extended to higher level. In general, statistical fit improves with the number of types up to a saturation point.[38] The results support the theoretical argument of the presence of private information and different strategic behavior, classifiable into different *quality-types* of credit card holders in the market.

Based on an iterative process of model selection (Henry et al. 2014) and using traditional information criteria for statistical fit [39] the preferred specification reaches a model of 6 types, that is to say; the proposed model that accounts for unobserved heterogeneity, identifies the presence of at least 6 *quality-type* of clients (See Table 6).

[38]Keep in mind that the number of regresors also increases.

[39]Information criteria typically used are the AIC, BIC. Combined with the actual likelihood level, they offer sufficient means for statistical model selection. Additionally, fmmlc in Stata (Luedicke 2011) offers an "entropy" measure (Ramaswamy 1993) that based on the posterior probabilities captures the power of the components' classification in the model. I report this measures for completeness. Incidentally, the highest entropy is reached on the preferred specification chosen.

Table 4.6: Percentage types classification and model fit comparison

Component	Baseline (1 Type)	2 Types	3 Types	4 Types	5 Types	6 Types
1	100	88.44	53.91	15.4	14.62	14.48
2		11.56	6.35	4.5	3.94	3.93
3			39.74	49.33	23.84	23.82
4				30.77	15.71	15.69
5					41.89	41.81
6						0.26
Entropy	0	0.520	0.408	0.367	0.479	0.532
LogLik (lf.)	-131569.56	-123964.62	-122050.13	-121284.19	-120504.93	-120415.33
AIC (r.)	263199.13	248049.23	244294.25	242836.38	241351.86	241246.67
BIC (r.)	263480.96	248612.88	245205.48	244095.18	242958.25	243200.64

Estimations performed using the *fmm* Stata module from Deb (2007) and post estimation command *fmmcl* from Luedicke (2011).
Percentage classification based on most likely class from the posterior probabilities.
Entropy measure based on Ramaswamy (1993).

There are important differences in the magnitude and significance of the determinant coefficients for the demand equation (ln(Total Credit Card debt)). Furthermore, conditional on having any debt (100% of our sample) ordinary estimation (such as OLS) of a demand equation involves inevitable endogeneity problems due primarily to OVB. Under the identification assumption imposed, once controlled for unobserved heterogeneity, estimation results are consistent and offer interesting insights over heterogeneous demand motivations.

Table 7 presents the baseline models and Table 8 summarizes the main results of this section. Similar to the default analysis, coefficients for the "Type equation" are not directly comparable to those on the baseline model (OLS) however they offer a first glance at the differences in significance of the factors involved on credit demand, when controlling for unobserved heterogeneity. While most of the coefficients in the baseline model are significant, only a few appear to offer valuable information for the types classification. Furthermore, some types of clients are clearly identifiable based on this demographic factors while others do not. Take "age" as an example, it statistically explains the credit demand in the baseline model, however it fails to offer any differentiable information over the types classification for the demand model. Other regresors, such as becoming a widow, having a college degree, having a high or medium responsibility job description and entering the market over a period of interest rate stability (less market uncertainty), all are factors that show statistical significance heterogeneously on types.

The choice equation shows how powerful the estimation strategy is when it comes to extract behavioral differences of the elasticities (or semi-elasticities) of economic variables with respect to credit demand. All regresors except one are significant in the baseline model while the Type consistent model shows identifiable behavioral differences of each regresor conditional on the clients' type. Take the two heaviest types in the classification (Type 3 and 5, Table 10) based on the posterior probabilities (see section of empirical model for details); regresors for both are different in magnitude and statistical significance, implying that clients classified into these classes have identifiable heterogeneous underlying preferences (strategic behavior). Failing to account for this would result on inconsistent estimation and biased inference. Clients classified as Type 3 (23.82%) in average are more responsive to increase their demand than those classified as Type 5 (41.81%) when they increase their time of membership, their ratio of monthly balance payment over income and they become business owners. The opposite effect is described when looking at other relevant regresors in the model. Interestingly, other things equal, clients classified as Type 2 (3.93%) show a lower unconditional credit demand when compared to the other Types[40]. At the same time, their demand sensitivity is significantly higher for the ratio of monthly balance payment over income, returned checks and number of debts on the credit risk bureau. Different to the rest, equity composition (houses, cars,

[40]Note, by looking at the "Sigma" coefficient that their distribution dispersion is higher

Table 4.7: Credit demand [ln(Total CC. Debt)]: baseline models

	(1) OLS		(2) OLS_cfe	
	β	SE	β	SE
Demographics				
Age(in years)	0.0437***	(0.0049)	0.0437***	(0.0050)
Age squared	−0.0005***	(0.0000)	−0.0005***	(0.0000)
Woman=1	0.0055	(0.0109)	0.0003	(0.0123)
Married=1	0.0740***	(0.0147)	0.0856***	(0.0096)
Widower=1	0.1027***	(0.0203)	0.1113***	(0.0186)
Divorced=1	0.0888***	(0.0191)	0.0980***	(0.0207)
College(pf)=1	0.0109	(0.0102)	0.0133	(0.0093)
Science degree	−0.0138	(0.0087)	−0.0151*	(0.0089)
H/M Responsibility(wd)=1	0.0796***	(0.0092)	0.0792***	(0.0089)
Mltry/Police(pf)=1	0.0179	(0.0124)	0.0078	(0.0096)
Family Members (#)	−0.0068**	(0.0029)	−0.0065**	(0.0027)
Stable 'r'-period	−0.2332***	(0.0175)	−0.2325***	(0.0168)
Main cities				
Quito	0.1886***	(0.0329)	0.2467***	(0.0029)
Guayaquil	0.3074***	(0.0320)	0.3670***	(0.0039)
Cuenca	−0.1556***	(0.0329)	−0.1009***	(0.0063)
Ambato	−0.1865***	(0.0326)	−0.1317***	(0.0061)
Economic variables				
Income(USD)	0.0002***	(0.0000)	0.0002***	(0.0000)
Tot. Net Worth (USD)	0.0000***	(0.0000)	0.0000***	(0.0000)
Tot.CC Debt/Income	0.0642***	(0.0033)	0.0633***	(0.0037)
% Debt at Risk on CRB	−0.1014**	(0.0409)	−0.1206***	(0.0353)
Ownership Type	0.0578***	(0.0095)	0.0530***	(0.0077)
Properties (#)	−0.0648***	(0.0127)	−0.0626***	(0.0120)
Houses(#)	−0.0560***	(0.0079)	−0.0568***	(0.0078)
Cars(#)	0.0424***	(0.0066)	0.0426***	(0.0063)
Debts on CRB(#)	0.3653***	(0.0072)	0.3644***	(0.0076)
Returned Checks=1	0.2614***	(0.0234)	0.2501***	(0.0191)
Ytd.Avrg.Payments/Cr.Balance	−0.2815***	(0.0706)	−0.2939***	(0.0671)
Microcredit=1	−0.2654***	(0.0228)	−0.2550***	(0.0254)
Portfolio/Tot.Debt	0.1042***	(0.0383)	0.1109***	(0.0369)
Hist.Wgtd.Deferred/Tot.Debt	1.1233***	(0.0776)	1.1209***	(0.0784)
Times Debt Refin. [#:1 to 5]	0.1102	(0.0763)	0.0796	(0.0788)
Mth.Bal.Payment/Income	0.3726***	(0.0165)	0.3650***	(0.0153)
Time of membership (months)	0.0012***	(0.0001)	0.0012***	(0.0001)
Constant	4.7746***	(0.1035)	4.7447***	(0.0894)
Observations	88794		88794	
AIC	263199.134		261975.021	
BIC	263480.956		262247.450	
LogLik	−131569.567		−130958.511	

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parenthesis (not-clustered). Cities fixed effects (not included in the table).

properties) and becoming a business owner are not important determinant factors for credit demand for these clients.

I will not extend on the individual interpretation of the coefficients as they are extensively exposed. The punchline of this section is that when controlling for heterogeneous preferences (unobserved) allows for the identification of behavioral differences that are valuable information for types stratification. Such knowledge allows lenders to improve over risk pricing strategies and credit risk management.

Different to the default probability, the focus of the credit demand analysis is on the identification of heterogeneous determinants rather than the out-of-sample predictions. Extensions of this work can include tests of forecast performance, however I consider this of second order interest for this research. Other extensions include a cost/benefit analysis of the implementation of the type consistent model by evaluating the "value at risk" in each strategy.

4.6 Conclusions

Individual preferences are a fundamental aspect of intertemporal choice problems; however, capturing this heterogeneity and, in particular, relating it to relevant preferences such as those over risk and time, is a theoretical and empirical challenge. An important aspect of the intertemporal problem is related to the access to credit markets. Subjects have different intrinsic motivations to decide over the size of a loan, the actual use of the credit and whether to payback the loan or default. Building on a basic model for a two periods intertemporal problem of loan demand, following Gan and Mosquera (2008) and Gan et al. (2013), I derive a basic statistical structure to identify client's unobserved heterogeneity, classified into *quality-types*, and offer a first interpretation that fosters to build a bridge over the empirical evidence of our statistical approach and that from other experimental measures.

Table 4.8: Credit demand [ln(Total CC. Debt)]: type consistent model

	Baseline model			**Type Consistent model**			
	OLS (cl.se)	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**
Type equation (demo.)							
Age(in years)	0.0437***	-	-0.0349	-0.1920	-0.1685	0.1669	-0.1092
	(0.0049)	-	(0.1450)	(0.1437)	(0.1452)	(0.1452)	(0.1445)
Age squared	-0.0005***	-	0.0006	0.0020	0.0016	-0.0018	0.0011
	(0.0000)	-	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)
Woman=1	0.0055	-	0.4719	0.4603	0.5093	0.5124	0.5827*
	(0.0109)	-	(0.3420)	(0.3439)	(0.3452)	(0.3424)	(0.3418)
Married=1	0.0740***	-	-0.6085	-0.5134	-0.5193	-0.1954	-0.3121
	(0.0147)	-	(0.4034)	(0.4057)	(0.4063)	(0.4048)	(0.4039)
Widower=1	0.1027***	-	-2.0129***	-1.8388**	-1.1779	-1.4952**	-1.5705**
	(0.0203)	-	(0.7459)	(0.7591)	(0.8278)	(0.7568)	(0.7420)
Divorced=1	0.0888***	-	-0.5951	-0.4393	-0.1592	-0.0663	-0.2355
	(0.0191)	-	(0.6315)	(0.6368)	(0.6429)	(0.6333)	(0.6328)
Family Members (#)	-0.0068**	-	-0.1742	-0.0258	0.0353	-0.1233	-0.0838
	(0.0029)	-	(0.1155)	(0.1165)	(0.1228)	(0.1151)	(0.1147)
College(pf)=1	0.0109	-	1.0291*	1.0059*	0.7330	1.0775**	0.8848
	(0.0102)	-	(0.5435)	(0.5457)	(0.5469)	(0.5475)	(0.5461)
Science degree	-0.0138	-	-0.7743	-0.7852	-0.4763	-0.8337	-0.7966
	(0.0087)	-	(0.5445)	(0.5479)	(0.5493)	(0.5473)	(0.5464)
H/M Responsibility(wd)=1	0.0796***	-	0.3696	0.4788	0.5210	0.9471***	0.4975
	(0.0092)	-	(0.3506)	(0.3528)	(0.3532)	(0.3552)	(0.3505)
Mltry/Police(pf)=1	0.0179	-	0.2428	0.0454	-0.2749	0.3670	-0.1054
	(0.0124)	-	(0.5863)	(0.5945)	(0.6249)	(0.5946)	(0.5916)
Stable 'r'-period	-0.2332***	-	-0.0043	0.4244	20.8720***	-0.0974	-20.7047***
	(0.0175)	-	(0.3215)	(0.3236)	(0.3584)	(0.3218)	(0.5183)
Income(USD)	-	-	0.0001	-0.0004*	-0.0001	0.0001	-0.0002
	-	-	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Constant	-	-	4.9186	8.2002**	-11.1287***	0.0423	8.0466**
	-	-	(3.3647)	(3.3030)	(3.3072)	(3.3473)	(3.3215)
Choice equation (econ.)							
Ownership Type	0.0578***	-0.0708**	-0.1255	0.1589***	0.1611***	0.0692***	0.0570***
	(0.0095)	(0.0331)	(0.0786)	(0.0271)	(0.0167)	(0.0163)	(0.0005)
Properties (#)	-0.0648***	-0.0695**	-0.1713*	-0.0471	0.0002	-0.0634***	0.1458***
	(0.0127)	(0.0342)	(0.1023)	(0.0435)	(0.0225)	(0.0207)	(0.0006)
Houses(#)	-0.0560***	-0.0240	0.0550	-0.0734**	-0.0431***	-0.0451***	-0.0991***
	(0.0079)	(0.0215)	(0.0565)	(0.0351)	(0.0141)	(0.0136)	(0.0005)
Cars(#)	0.0424***	0.0492***	0.0854*	-0.0164	0.0293***	0.0251**	-0.0138***
	(0.0066)	(0.0171)	(0.0485)	(0.0190)	(0.0106)	(0.0102)	(0.0004)
Debts on CRB(#)	0.3653***	0.4261***	0.7003***	0.3073***	0.1790***	0.3184***	0.2396***
	(0.0072)	(0.0131)	(0.0223)	(0.0076)	(0.0087)	(0.0093)	(0.0002)
% Debt at Risk on CRB	-0.1014**	-0.2512**	0.1081	0.0601	-0.1799*	-0.0450	-0.9281***
	(0.0409)	(0.1143)	(0.2443)	(0.0622)	(0.0962)	(0.0801)	(0.0014)
Returned Checks=1	0.2614***	0.1942***	0.8181***	0.1640***	0.1055***	0.2400***	0.6428***
	(0.0234)	(0.0612)	(0.1872)	(0.0536)	(0.0397)	(0.0385)	(0.0009)
Ytd.Avrg.Payments/Cr.Balance	-0.2815***	-0.0796	1.2824***	-0.3482***	-0.1970**	-0.5987***	-0.5325***
	(0.0706)	(0.1295)	(0.2659)	(0.0651)	(0.0912)	(0.0781)	(0.0027)
Microcredit=1	-0.2654***	-0.2465***	-0.8589***	-0.1995***	-0.1807***	-0.2220***	-0.1451***
	(0.0228)	(0.0628)	(0.1732)	(0.0365)	(0.0318)	(0.0323)	(0.0010)
Portfolio/Tot.Debt	0.1042***	0.3952***	0.1920**	-0.1148***	-0.1646***	0.2279***	-0.1089***
	(0.0383)	(0.0495)	(0.0924)	(0.0370)	(0.0360)	(0.0382)	(0.0011)
Hist.Wgtd.Deferred/Tot.Debt	1.1233***	1.8387***	1.8764***	0.5619***	-0.0265	1.1737***	0.6955***
	(0.0776)	(0.0746)	(0.1132)	(0.0558)	(0.0531)	(0.0681)	(0.0015)
Tot.CC Debt/Income	0.0642***	0.0105	0.0141	0.0558***	0.0881***	0.0982***	0.1106***
	(0.0033)	(0.0125)	(0.0153)	(0.0080)	(0.0099)	(0.0089)	(0.0002)
Times Debt Refin. [#:1 to 5]	0.1102	0.3794*	1.7184***	0.0000	0.0139	-0.1311	-3.4668***
	(0.0763)	(0.2212)	(0.2326)	(.)	(0.3180)	(0.2044)	(0.0009)
Mth.Bal.Payment/Income	0.3726***	1.2208***	0.6049***	0.1275***	0.0756***	0.0882***	0.2424***
	(0.0165)	(0.1114)	(0.0729)	(0.0257)	(0.0262)	(0.0308)	(0.0005)
Income(USD)	0.0002***	0.0002***	-0.0001	0.0002***	0.0001***	0.0003***	0.0002***
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tot. Net Worth (USD)	0.0000***	0.0000**	-0.0000	0.0000	0.0000	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Time of membership (months)	0.0012***	0.0017***	0.0036***	0.0187***	0.0009***	0.0004***	0.0011***
	(0.0001)	(0.0002)	(0.0003)	(0.0007)	(0.0001)	(0.0001)	(0.0000)
Constant	4.7746***	4.1422***	1.1098***	5.9957***	8.0141***	6.3223***	7.1962***
	(0.1035)	(0.1776)	(0.3967)	(0.0881)	(0.1023)	(0.1289)	(0.0022)
Sigma		0.8293***	1.4715***	0.6467***	0.5381***	0.6370***	0.0012***
		(0.0189)	(0.0186)	(0.0200)	(0.0167)	(0.0239)	(0.2698)
Percentage per type	100	14.48	3.93	23.82	15.69	41.81	0.26
Observations	88794			88794			
AIC	263199.134			241230.667			
BIC	263480.956			243109.482			
LogLik	-131569.567			-120415.334			

Estimations performed using the *fmm* Stata module from Deb (2007) and post estimation command *fmmcl* from Luedicke (2011).
Percentage classification based on most likely class from the posterior probabilities.

Credit analysis relies traditionally on the observed historical information of their clients to define credit policies such as: access, credit limits, interest rates; besides other non-pricing alternatives. In this context, knowledge of the individual heterogeneity of clients reduces the informational asymmetries in the market that potentially degenerate into credit rationing. When observing default behavior, we find evidence for two *quality-types* of clients that are very different from each other, supporting previous findings (Gan and Mosquera 2008), although relying on a slightly different identification strategy and a more detailed data set. I find no evidence of education being an important factor for types classification; yet, I am cautious about this evidence since there is no big heterogeneity in this covariate among clients in the data set at hand. The most informative variables for type identification are job description, family composition, age and gender. More importantly, economic determinants of credit default behavior are statistically different conditional on client's types. Out-of-sample predictions show important improvements in potential clients' selection mechanism when using the Type Consistent estimation (conditional or unconditional on types) versus any other regresion alternative (OLS, Probit, Logit, Stepwise forward/backwards logit) Higher rigurosity involves a trade off in prediction accuracy, as a result, the probability to reject "good" clients (Type I error) mildly deteriorates, nevertheless, the main potential error, from an adverse selection perspective, in any scoring process, i.e. failing to reject "bad" clients (Type II error)is substantially reduced when considering the unobserved heterogeneity through the finite mixture estimation proposed.

Finally, when it comes to credit card demand, I successfully identify 6 *quality-types* with distinguishable behavioral differences over the economic determinants in the model. Predictably, the demand function fit increases precision and offers better statistical performance over traditional estimation strategies. Risk pricing and credit risk management strategies can be highly improved when considering types stratification derived from individual preferences.

The policy implication of the *type-consistent* estimation supported in this paper leans towards a more efficient credit market allocation, potentially reducing *credit rationing*. There are important welfare gains (Pareto improvements) by accounting for heterogeneous types selection since a better identification of *quality-clients* allows lenders to offer multiple (finer) contract conditions and risk pricing strategies, incorporating into the market reliable (credit worthy) clients that were crowded out due to inconsistent credit score selection process. At the same time, better contract conditions or other non-pricing techniques of lenders can improve their risk management decision process, a fundamental topic for systematic financial health in a modern economy.

5. CONCLUSIONS

This thesis compiles three essays on experimental and applied economics that study unexpected influences of heterogeneous institutional environments.

The first essay documents exact replications of four classic experiments in social dilemma games. The most relevant finding echoes the call on other experimental disciplines for the need of replication studies to confirm and advance in the gathering of robust scientific knowledge. Selective reporting, publication bias and even scientific dishonesty, particularly in social sciences, are important problems of scientific dissemination that can bias not only the knowledge within the discipline but negatively influence policy implementation. We offer compelling evidence of an attenuation effect in replication studies also pervasive in experimental economics. Most outcomes in the original studies do not replicate and, when they do the effects are much smaller. Also, we show the presence of unexpected context influences; experimental subjects in Texas, mostly undergrad students, contribute consistently more and free ride less than in the original studies. Beyond predictable sample differences, there are other factors that filter to results in the lab and which interdependence is not well understood. Cultural differences, political dominance, social cohesion, economic institutions, formal or informal markets, among others, all play a role in economic decisions, hence the only way to permeate findings in the discipline is to promote more robust replications.

The second essay studies economic efficiency on two different institutional arrangements of sanctioning power in social dilemmas: endogenous versus exogenous power delegation. The introduction of a democratic participation mechanism into sanctioning power institutions for a public good provision, reduces the free-riding problem and enhances collective action, without sacrificing economic efficiency (cost/effective). Although a weak effect, this result offers valuable insights to explain differences in governance outcomes of common pool resources and provision of public goods whether in the lab or in the field. Further research in social dilemma environments should consider the fact that the manager selection mechanism is not innocuous, and unobserved incentives, such as the perception of authority legitimacy, and other beliefs might be at play.

The final essay, chapter IV, offers an empirical alternative to unravel unobserved heterogeneity in the credit market as a means to reduce the potential credit rationing problems (lack of credit supply). Specifically, the adverse selection problem can be greatly reduced by the implementation of a finite mixture estimation to account for different quality-types of clients for default decisions and credit demand. This essay successfully shows that behavior predictability greatly improves over traditional scoring techniques. The probability of accepting a non-creditworthy client reduces to less than half when compared to baseline models. Acknowledging client's unobserved preferences supports the transition towards behavioral scoring techniques that favor the willingness to repay the loan over strictly the ability to pay, where most current practice lies. Extensions require more experimental work that reconcile dominant statistical techniques.

REFERENCES

Acemoglu, D., S. Johnson, and J. A. Robinson (2005) “Institutions as the fundamental cause of long-run growth,” in P. Aghion ed. *Handbook of Economic Growth*: Elsevier, pp. 385–472.

Anderson, L. R., J. M Mellor, and J. Milyo (2005) *Do liberals play nice? The effects of party and political ideology in public goods and trust games*, Vol. 13: Elsevier, JAI Press, Edited by John Morgan.

Andreoni, J. (1990) “Impure altruism and donations to public goods: A theory of warm-glow giving,” *Economic Journal*, Vol. 100, No. 401, pp. 464–477.

——— (1995a) “Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments,” *The Quarterly Journal of Economics*, Vol. 110, No. 1, pp. 1–21.

——— (1995b) “Cooperation in Public-Goods Experiments: Kindness or Confusion?,” *American Economic Review*, Vol. 85, No. 4, pp. 891–904.

Andreoni, J. and L. Gee (2012) “Gun For Hire: Delegated Enforcement and Peer Punishment in Public Goods Provision,” *Journal of Public Economics*, Vol. 96, No. 11-12, pp. 1036–1046.

Andreoni, J. and C. Sprenger (2012) “Estimating Time Preferences from Convex Budgets,” *American Economic Review*, Vol. 102, No. 7, pp. 3333–56.

Arbak, E. and M. C. Villeval (2011) “Endogenous leadership selection and influence,” Documents De Travail - Working Paper 07-07, Institute of Economic Theory and

Analysis.

Arya, S., C. Eckel, and C. Wichman (2013) “Anatomy of the credit score,” *Journal of Economic Behavior & Organization*, Vol. 95, pp. 175–185.

Ashley, R., S. Ball, and C. Eckel (2010) “Motives for Giving: A Reanalysis of Two Classic Public Goods Experiments,” *Southern Economic Journal*, Vol. 77, No. 1, pp. 15–26.

Baldassarri, D. and G. Grossman (2011) “Centralized sanctioning and legitimate authority promote cooperation in humans,” *Proceedings of the National Academy of Sciences*, Vol. 108, No. 27, pp. 11023–11027.

——— (2013) “The effect of group attachment and social position on prosocial behavior. Evidence from lab-in-the-field experiments,” *PloS one*, Vol. 8, No. 3, p. e58750.

Ball, S., C. Eckel, P. J. Grossman, and W. Zame (2001) “Status In Markets,” *The Quarterly Journal of Economics*, Vol. 116, No. 1, pp. 161–188.

Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Börsch-Supan, D. Brownstone, D. S. Bunch et al. (2002) “Hybrid choice models: Progress and challenges,” *Marketing Letters*, Vol. 13, No. 3, pp. 163–175.

Bertaut, C. C. and M. Haliassos (2005) “Credit cards: facts and theories,” *Available at SSRN 931179*.

Bertola, G., R. Disney, and C. B. Grant (2006) *The economics of consumer credit*: Mit Press.

Bester, H. (1985) “Screening vs. rationing in credit markets with imperfect information," *The American Economic Review*, pp. 850–855.

Bolton, G. E. and A. Ockenfels (2000) "ERC: a theory of equity, reciprocity, and competition," *American Economic Review*, Vol. 90, No. 1, pp. 166–193.

Botelho, A., G. W. Harrison, L. Pinto, and E. E. Rutström (2005) "Social norms and social choice," Working Papers 30, Universidade do Minho.

Bougherara, D., L. Denant-Boemont, and D. Masclet (2011) "Cooperation and framing effects in provision point mechanisms: Experimental evidence," *Ecological Economics*, Vol. 70, No. 6, pp. 1200–1210.

Boyes, W. J., D. L. Hoffman, and S. A. Low (1989) "An econometric analysis of the bank credit scoring problem," *Journal of Econometrics*, Vol. 40, No. 1, pp. 3–14.

Brandts, J. and D. J. Cooper (2007) "It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure," *Journal of the European Economic Association*, Vol. 5, No. 6, pp. 1223–1268.

Brandts, J., D. J. Cooper, and R. A. Weber (2014) "Legitimacy, Communication and Leadership in the Turnaround Game," Working Papers 755, Barcelona Graduate School of Economics.

Brooks, A. C. (2007) *Who Really Cares: The Surprising Truth About Compassionate Conservatism–America's Charity Divide–Who Gives*: Basic Books.

Carpenter, J., S. Kariv, and A. Schotter (2012) "Network architecture, cooperation and punishment in public good experiments," *Review of Economic Design*, Vol. 16, No. 2-3, pp. 93–118.

Carpenter, J. and P. H. Matthews (2009) “What norms trigger punishment?” *Experimental Economics*, Vol. 12, No. 3, pp. 272–288.

Carpenter, J. P. (2007) “The demand for punishment,” *Journal of Economic Behavior & Organization*, Vol. 62, No. 4, pp. 522–542.

Carroll, C. (1997) “Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis,” *The Quarterly Journal of Economics*, Vol. 112, No. 1, pp. 1–55.

Cartwright, E., J. Gillet, and M. Van Vugt (2013) “Leadership by Example in the Weak-Link Game,” *Economic Inquiry*, Vol. 51, No. 4, pp. 2028–2043.

Casari, M. (2005) “On the design of peer punishment experiments,” *Experimental Economics*, Vol. 8, No. 2, pp. 107–115.

Casari, M. and L. Luini (2009) “Cooperation under alternative punishment institutions: An experiment,” *Journal of Economic Behavior & Organization*, Vol. 71, No. 2, pp. 273–282.

——— (2012) “Peer punishment in teams: expressive or instrumental choice?” *Experimental Economics*, Vol. 15, No. 2, pp. 241–259.

Cassar, A., L. Crowley, and B. Wydick (2007) “The effect of social capital on group loan repayment: Evidence from field experiments,” *The Economic Journal*, Vol. 117, No. 517, pp. F85–F106.

Charness, G., R. Cobo-Reyes, and N. Jiménez (2008) “An investment game with third-party intervention,” *Journal of Economic Behavior & Organization*, Vol. 68, No. 1, pp. 18–28.

Charness, G. and M. Rabin (2002) “Understanding Social Preferences with Simple

Tests," *Quarterly Journal of Economics*, Vol. 117, No. 3, pp. 817–869.

Chaudhuri, A. (2011) "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics*, Vol. 14, No. 1, pp. 48–83.

Chaudhuri, A. and T. Paichayontvijit (2010) "Recommended play and performance bonuses in the minimum effort coordination game," *Experimental Economics*, Vol. 13, No. 3, pp. 346–363.

Crook, J. N., D. B. Edelman, and L. C. Thomas (2007) "Recent developments in consumer credit risk assessment," *European Journal of Operational Research*, Vol. 183, No. 3, pp. 1447–1465.

Croson, R. and U. Gneezy (2009) "Gender differences in preferences," *Journal of Economic Literature*, pp. 448–474.

Croson, R. T. A. (2007) "Theories Of Commitment, Altruism And Reciprocity: Evidence From Linear Public Goods Games," *Economic Inquiry*, Vol. 45, No. 2, pp. 199–216.

Dal Bó, P. (2014) "Experimental evidence on the workings of democratic institutions," *Institutions, Property Rights, and Economic Growth: The Legacy of Douglass North*, p. 266.

Dal Bó, P., A. Foster, and L. Putterman (2008) "Institutions and behavior: Experimental evidence on the effects of democracy,"Technical report, National Bureau of Economic Research.

Deb, P. (2007) "FMM: Stata module to estimate finite mixture models," Statistical

Software Components, Boston College Department of Economics, December.

Deb, P. and P. K. Trivedi (2002) "The structure of demand for health care: latent class versus two-part models," *Journal of Health Economics*, Vol. 21, No. 4, pp. 601–625.

Decker, T., A. Stiehler, and M. Strobel (2003) "A Comparison of Punishment Rules in Repeated Public Good Games: An Experimental Study," *Journal of Conflict Resolution*, Vol. 47, No. 6, pp. 751–772.

Dufwenberg, M. and G. Kirchsteiger (2004) "A theory of sequential reciprocity," *Games and Economic Behavior*, Vol. 47, No. 2, pp. 268–298.

Eckel, C. C. (2007) "Economic Games for Social Scientists," *In Laboratory Experiments in the Social Sciences*, pp. 497–515, edited by Murray Webster and Jane Sell.

Eckel, C. C., E. Fatas, and R. Wilson (2010) "Cooperation and status in organizations," *Journal of Public Economic Theory*, Vol. 12, No. 4, pp. 737–762.

Eckel, C. C. and P. J. Grossman (2004) "Giving to Secular Causes by the Religious and Nonreligious: An Experimental Test of the Responsiveness of Giving to Subsidies,"Technical Report 2, Saint Cloud State University.

——— (2008a) "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices," *Journal of Economic Behavior & Organization*, Vol. 68, No. 1, pp. 1–17.

——— (2008b) "Men, women and risk aversion: Experimental evidence," *Handbook of Experimental Economics Results*, Vol. 1, pp. 1061–1073.

Eckel, C. C., H. Harwell, and J. G. Castillo (2015) "Four Classic Public Goods Experiments: A replication study," working paper, Texas A&M University, Forthcoming in Research in Experimental Economics, Vol.18.

Eckel, C. C., C. Johnson, C. Montmarquette, and C. Rojas (2007) "Debt aversion and the demand for loans for postsecondary education," *Public Finance Review*, Vol. 35, No. 2, pp. 233–262.

Eckel, C. C. and R. K. Wilson (2007) "Social learning in coordination games: does status matter?" *Experimental Economics*, Vol. 10, No. 3, pp. 317–329.

Eckel, C., C. Montmarquette, and C. Johnson (2004) *Saving decisions of the working poor: Short-and long-term horizons*: CIRANO.

Ertan, A., T. Page, and L. Putterman (2005) "Can endogenously chosen institutions mitigate the free-rider problem and reduce perverse punishment?"Technical report, Working Paper, Brown University, Department of Economics.

——— (2009) "Who to punish? Individual decisions and majority rule in mitigating the free rider problem," *European Economic Review*, Vol. 53, No. 5, pp. 495–511.

Eymann, A., A. Boersch-Supan, and R. Euwals (2001) "Risk Attitude, Impatience, and Portfolio Choice,"Technical report, Working Paper, University of Mannheim, Germany.

Falk, A., E. Fehr, and U. Fischbacher (2005) "Driving Forces behind Informal Sanctions," *Econometrica*, Vol. 73, No. 6, pp. 2017–2030.

Fehr, E. and S. Gächter (2000a) "Cooperation and punishment in public goods experiments," *American Economic Review*, Vol. 90, No. 4, pp. 980–994.

Fehr, E. and S. Gächter (2000b) "Fairness and retaliation: The economics of reciprocity," *The Journal of Economic Perspectives*, pp. 159–181.

Fehr, E., H. Herz, and T. Wilkening (2013) "The lure of authority: Motivation and incentive effects of power," *American Economic Review*, Vol. 103, No. 4.

Fehr, E. and K. M. Schmidt (1999) "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 817–868.

Feinstein, J. S. (1990) "Detection controlled estimation," *Journal of Law and Economics*, pp. 233–276.

Fischbacher, U. (2007) "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, Vol. 10, No. 2, pp. 171–178.

Fisher, J. and A. Lyons (2010) "Information and credit access: using bankruptcy as a signal," *Applied Economics*, Vol. 42, No. 25, pp. 3175–3193.

Fowler, J. H. and C. D. Kam (2007) "Beyond the self: Social identity, altruism, and political participation," *Journal of Politics*, Vol. 69, No. 3, pp. 813–827.

Frederick, S., G. Loewenstein, and T. O'donoghue (2002) "Time discounting and time preference: A critical review," *Journal of Economic Literature*, pp. 351–401.

Friedman, M. (1957) *A Theory of the Consumption Function*: National Bureau of Economic Research, Inc.

Gan, L., M. A. Hernández, and Y. Liu (2013) "Group lending with heterogeneous types,"Technical report, National Bureau of Economic Research.

Gan, L., F. Huang, and A. Mayer (2011) "A simple test of private information in the insurance markets with heterogeneous insurance demand,"Technical report, National Bureau of Economic Research.

Gan, L. and R. Mosquera (2008) "An Empirical Study of the Credit Market with Unobserved Consumer Typers,"Technical report, National Bureau of Economic Research.

Grant, C. (2007) "Estimating credit constraints among US households," *Oxford Economic Papers*, Vol. 59, No. 4, pp. 583–605.

Greiner, B. (2004) "An online recruitment system for economic experiments," *Forschung und wissenschaftliches Rechnen (Research in scientific computing)*, pp. 79–93, edited by Kurt Kremer and Volker Macho.

Gross, D. B. and N. S. Souleles (2002) "Do Liquidity Constraints And Interest Rates Matter For Consumer Behavior? Evidence From Credit Card Data," *The Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 149–185.

Grossman, P. and C. Eckel (2012) "Giving versus taking: a "real donation" comparison of warm glow and cold prickle in a context-rich environment,"Technical report, Monash University, Department of Economics.

Guillen, P., C. Schwieren, and G. Staffiero (2007) "Why feed the Leviathan?" *Public Choice*, Vol. 130, No. 1-2, pp. 115–128.

Gupta, S. and P. K. Chintagunta (1994) "On using demographic variables to determine segment membership in logit mixture models," *Journal of Marketing Research*, pp. 128–136.

Gürerk, O., B. Irlenbusch, and B. Rockenbach (2006) "The competitive advantage of sanctioning institutions," *Science*, Vol. 7, No. 312, pp. 108–111.

Hamman, J. R., R. A. Weber, and J. Woon (2011) "An experimental investigation of electoral delegation and the provision of public goods," *American Journal of Political Science*, Vol. 55, No. 4, pp. 738–752.

Hand, D. J. and W. E. Henley (1997) "Statistical classification methods in consumer credit scoring: a review," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 160, No. 3, pp. 523–541.

Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström (2005) "Risk aversion and incentive effects: Comment," *American Economic Review*, pp. 897–901.

Harrison, G. W., M. I. Lau, E. E. Rutström, and M. B. Sullivan (2005) "Eliciting risk and time preferences using field experiments: Some methodological issues," *Field Experiments in Economics*, Vol. 10, pp. 125–218.

Heckman, J. and B. Singer (1984) "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica: Journal of the Econometric Society*, pp. 271–320.

Henry, M., Y. Kitamura, and B. Salanié (2014) "Partial identification of finite mixtures in econometric models," *Quantitative Economics*, Vol. 5, No. 1, pp. 123–144.

Herrmann, B., C. Thöni, and S. Gächter (2008) "Antisocial punishment across societies," *Science*, Vol. 319, No. 5868, pp. 1362–1367.

Hess, S., M. Ben-Akiva, D. Gopinath, and J. Walker (2011) "Advantages of latent class over continuous mixture of logit models," *Institute for Transport Studies,* *University of Leeds. Working Paper.*

Hobbes, T. and I. Shapiro (1990) *Leviathan, Or the Matter, Forme [Form], & Power of a Common-wealth Ecclesiasticall [Ecclesiastical] and Civill [Civil]*: Yale University Press.

Holt, C. A. and S. K. Laury (2002) "Risk aversion and incentive effects," *American Economic Review*, Vol. 92, No. 5, pp. 1644–1655.

——— (2005) "Risk aversion and incentive effects: New data without order effects," *American Economic Review*, pp. 902–904.

Houser, D. and R. Kurzban (2002) "Revisiting kindness and confusion in public goods experiments," *American Economic Review*, pp. 1062–1069.

Hung, A. A. and C. R. Plott (2001) "Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions," *American Economic Review*, Vol. 91, No. 5, pp. 1508–1520.

Ioannidis, J. P. (2005) "Why most published research findings are false," *PLoS medicine*, Vol. 2, No. 8, p. e124.

Isaac, R. M. and J. M. Walker (1988) "Group size effects in public goods provision: The voluntary contributions mechanism," *The Quarterly Journal of Economics*, pp. 179–199.

J, F. P. and V. C. A (2010) "The Source and Significance of Confusion in Public Goods Experiments," *The B.E. Journal of Economic Analysis & Policy*, Vol. 10, No. 1, pp. 1–42.

Jaffee, D. M. and T. Russell (1976) "Imperfect information, uncertainty, and credit rationing," *The Quarterly Journal of Economics*, pp. 651–666.

Keane, M. P. and K. I. Wolpin (1997) "The career decisions of young men," *Journal of Political Economy*, Vol. 105, No. 3, pp. 473–522.

Klein, R. A., K. A. Ratliff, M. Vianello, R. B. Adams Jr, Š. Bahník, M. J. Bernstein, K. Bocian, M. J. Brandt, B. Brooks, C. C. Brumbaugh et al. (2014) "Investigating variation in replicability: A "many labs" replication project.," *Social Psychology*, Vol. 45, No. 3, p. 142.

Kumru, C. S. and L. Vesterlund (2010) "The effect of status on charitable giving," *Journal of Public Economic Theory*, Vol. 12, No. 4, pp. 709–735.

Laibson, D., A. Repetto, and J. Tobacman (2007) "Estimating Discount Functions with Consumption Choices over the Lifecycle," Working Paper 13314, National Bureau of Economic Research.

LaPorta, R., F. L. D. Silanes, A. Shleifer, and R. W. Vishny (1998) "Law and Finance," *Journal of Political Economy*, Vol. 106, No. 6, pp. 1113–1155.

Ledyard, J. O. (1995) "Public Goods: A Survey of Experimental Research," in J. Kagel and A. Roth eds. *Handbook of Experimental Economics*: Princeton University Press.

Lehrer, J. (2010) "The truth wears off," *The New Yorker*, Vol. 13, p. 52.

Levy, D. M., K. Padgitt, S. J. Peart, D. Houser, and E. Xiao (2011) "Leadership, cheap talk and really cheap talk," *Journal of Economic Behavior & Organization*, Vol. 77, No. 1, pp. 40–52.

Lipset, S. M. (1959) "Some social requisites of democracy: Economic development and political legitimacy," *American Political Science Review*, Vol. 53, No. 01, pp. 69–105.

Luedicke, J. (2011) "FMMLC: Stata module for postestimation with finite mixture models," Statistical Software Components, Boston College Department of Economics, June.

Margolis, M. and M. Sances (2013) "Who Really Gives? Partisanship and Charitable Giving in the United States," *Partisanship and Charitable Giving in the United States*.

Meier, S. and C. Sprenger (2010) "Present-biased preferences and credit card borrowing," *American Economic Journal: Applied Economics*, pp. 193–210.

Milde, H. and J. G. Riley (1988) "Signaling in credit markets," *The Quarterly Journal of Economics*, pp. 101–129.

Nikiforakis, N. (2008) "Punishment and counter-punishment in public good games: Can we really govern ourselves?" *Journal of Public Economics*, Vol. 92, No. 1-2, pp. 91–112.

North, D. C. (1981) *Structure and change in Economic History*, New York: Norton and Company.

O'Gorman, R., J. Henrich, and M. Van Vugt (2009) "Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation," *Proceedings of the Royal Society B: Biological Sciences*, Vol. 276, No. 1655, pp. 323–329.

Ostrom, E. (1992) *Governing the Commons: The Evolution of Institutions for Col-*

lective Action, New York: Cambridge University Press.

Park, E.-S. (2000) "Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding," *Journal of Economic Behavior & Organization*, Vol. 43, No. 4, pp. 405–421.

Plott, C. R. (1982) "Industrial organization theory and experimental economics," *Journal of Economic Literature*, pp. 1485–1527.

Prelec, D. and D. Simester (2001) "Always leave home without it: A further investigation of the credit-card effect on willingness to pay," *Marketing Letters*, Vol. 12, No. 1, pp. 5–12.

Putterman, L., J.-R. Tyran, and K. Kamei (2011) "Public goods and voting on formal sanction schemes," *Journal of Public Economics*, Vol. 95, No. 9, pp. 1213–1222.

Ramaswamy, V., W. S. Desarbo, D. J. Reibstein, and W. T. Robinson (1993) "An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data," *Marketing Science*, Vol. 12, No. 1, pp. 103–124.

Schooler, J. (2011) "Unpublished results hide the decline effect.," *Nature*, Vol. 470, No. 7335, p. 437.

Sefton, M., R. Shupp, and J. Walker (2007) "The Effect of Rewards and Sanctions in Provision of Public Goods," *Economic Inquiry*, Vol. 45, No. 4, pp. 671–690.

Shanley, J. and P. J. Grossman (2007) "Paradise to parking lots: Creation versus maintenance of a public good," *The Journal of Socio-Economics*, Vol. 36, No. 4, pp. 523–536.

Shefrin, H. M. and R. H. Thaler (1988) "The behavioral life-cycle hypothesis," *Eco-*

nomic Inquiry, Vol. 26, No. 4, pp. 609–643.

Sonnemans, J., A. Schram, and T. Offerman (1998) "Public good provision and public bad prevention: The effect of framing," *Journal of Economic Behavior & Organization*, Vol. 34, No. 1, pp. 143–161.

Sterling, T. D. (1959) "Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa," *Journal of the American Statistical Association*, Vol. 54, No. 285, pp. 30–34.

Stiglitz, J. E. and A. Weiss (1981) "Credit rationing in markets with imperfect information," *The American Economic Review*, pp. 393–410.

Sutter, M., S. Haigner, and M. G. Kocher (2010) "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations," *Review of Economic Studies*, Vol. 77, No. 4, pp. 1540–1566.

Thomas, L. C. (2000) "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers," *International Journal of Forecasting*, Vol. 16, No. 2, pp. 149–172.

Willinger, M. and A. Ziegelmeyer (1999) "Framing and cooperation in public good games: an experiment with an interior solution," *Economics letters*, Vol. 65, No. 3, pp. 323–328.

Yong, E. (2012) "Nobel laureate challenges psychologists to clean up their act," *Nat. News*, Vol. 490.

APPENDIX A

EXPERIMENTAL INSTRUCTIONS: LEVIATHAN VS. DEMOCRACY

A.1 General instructions

You are now taking part in an economic experiment. Depending on your decisions and the decisions of other participants, you will be able to earn money. How you can earn money is described in these instructions. **Please read them carefully**. You will have to answer some questions to check that you understand the instructions.

During the experiment you are not allowed to communicate with other participants. If you have a question, raise your hand. We will come to answer your questions. Sometimes you may have to wait a short while before the experiment continues. Please be patient. Thanks for your patience and cooperation.

Your earnings in the experiment will be calculated in points. Points will be converted to RMB at the following exchange rate:

1 point = 0.05 RMB

Upon the completion of the experiment, you will also receive a participation fee of RMB 10. At the end of the experiment your total earnings will be paid out to you in cash.

The experiment has two phases (10 periods for each, 20 periods in total). The following instructions explain the details of phase 1. The instructions of the subsequent phase will be handed out later.

A.2 Instructions for the first phase

Please read these instructions carefully. Again, you will have to answer some questions to check that you understand the instructions.

In the experiment, all participants are randomly divided into groups of 5. That is, you are in a group with four other participants. The group members will be the same throughout the entire experiment. Nobody knows which other participants are in their group, and nobody will be informed of who was in which group after the experiment.

Phase 1 includes 10 periods. In each period, each group member, including yourself, will have to make a decision on how to allocate an endowment of 20 points. In each period, you and the four other members in your group simultaneously decide how to allocate the endowment into two accounts: 1. Group account. 2. Private account. You will decide how many points to allocate to the group account. Only integers between 0 and 20 are allowed for this purpose. The remaining points will automatically be allocated to your private account.

Your earnings

Your earnings depend on the total number of points in the group account, and the number of points in your private account.

Your total earning in each period can be calculated by the following formula:

20 – (points you allocated to the group account) + 0.4 * (sum of points allocated by all group members to the group account)

For each point you allocate into your private account you get 1 point as earnings. For example, your earnings from the private account equal 10 points if you allocate 10 points to it.

Your earnings from the group account equal the sum of points allocated to the group account by all 5 group members multiplied by 0.4. For each point you allocate to the group account everyone in your group get 0.4 points as earnings. For example, if every group member, including yourself, allocates 10 points to the group account, the sum of points in the group account is 50 and then your earnings from the group account are equal to 20 points.

Note that you receive 1 point for each point you allocate to your private account. If instead you allocate 1 extra point to the group account, your earnings from the group account increase by 0.4 points and your earnings from your private account decrease by 1 point. However, allocating 1 extra point to the group account can increase the earnings of all the other 4 group members by 0.4 points. Therefore, the total group earnings increase by 2 points. Note that you also obtain earnings from points allocated to the group account by others.

Examples

Suppose you allocate 10 points to the group account, the second and third members of your group each allocate 20 points to the group account, and the remaining two members both allocate 0 point to the group account. Then the sum of points in the group account is 10 + 20 + 20 + 0 + 0 = 50 points. Each group member receives earnings of 0.4 * 50 = 20 points from the group account. Your total earnings are: 20 – 10 + (0.4 * 50) = 10 + 20 = 30 points. The second and third members' earnings are: 20 − 20 + (0.4 * 50) = 0 + 20 = 20 points. The fourth and fifth members' earnings are: 20 – 0 + (0.4 * 50) = 20 + 20 = 40 points.

Please raise your hand if you have any question.

A.3 Instructions for the second phase

A.3.1 Leviathan treatment

Please read these instructions carefully. Again, you will have to answer some questions to check that you understand the instructions.

This phase, including 10 periods in total, is like the previous one in that you continue to interact with the same four participants in your group and in each period you make a decision about allocating 20 points to either a private account or a group account. Your earnings are determined in the same way as in Phase 1 of the experiment. (For each point you allocate to your private account you get 1 point as earnings. Your earnings from the group account equal the sum of points allocated to the group account by all 5 group members multiplied by 0.4).

However, in this phase, before you begin to make decisions, **ONE** participant in your group will be selected as a manager by the experimenter. The selected manager will be unchanged throughout this phase.

How is the manager selected?

Between phase 1 and phase 2, you can observe the allocation decisions of each group member in phase 1. Then the experimenter will choose the manager based on a predefined distribution, unknown to you. You will observe who was chosen among the information from phase 1. Note that the Manager will be the same during the remaining periods (i.e. until the end of the experiment).

Your decisions in this phase

There are two stages in each period. In the first stage, you make your allocation decision and then observe the decisions of the other group members along with your earnings. In the second stage, 4 points are automatically collected from each group member. Hence each group has an account of 20 points in each period. The manager has an opportunity to use the points in this account to reduce the earnings of one participant in your group.

Suppose you are selected as a manager. After the first stage of each period, you will observe the amount allocated to the group account by each member in your group. Meanwhile, you will receive 4 points from each group member including yourself. Then you will be asked to choose up to one group member and how much you want to reduce his/her earnings by using the 20 points. Each point you want to reduce that participant's earnings costs 1 point. You are free to leave all group members' earnings unchanged. Note that the points that are not used to reduce others' earnings will be returned to every group member equally.

Earnings in each period are calculated as follows:

20 -4 points – (points you allocate to group account) + 0.4*(sum of points allocated by all in group to group account) – (points the manager uses to reduce your earnings if you are selected) + (the remaining points not used by the manager),

Examples

Suppose you are the manager in your group. In this period, you allocate 20 points to the group account, the second and third members of your group each allocate 10 points to the group account, and the remaining two members both allocate 5 point to the group account. Then the sum of points in the group account is 20 + 10 + 10 + 5 + 5 = 50 points. Each group member receives earnings of 0.4 * 50 = 20 points from the group account. Your gross earnings are: 20 – 20 + (0.4 * 50) = 0 + 20 = 20 points. The second and third members' gross earnings are: 20 – 10 + (0.4 * 50) = 10 + 20 = 30 points. The fourth and fifth members' gross earnings are: 20 – 5 + (0.4 * 50) = 15 + 20 = 35 points. In the second stage, you receive 4 points from each group member, 20 points in total. Now you decide to reduce the fifth member's earning by 10 points. Then the remaining 10 points will be returned to all group members, 2 point for each.

Your net earnings are 20 - 4 + 2 = 18 points. The second and third members' net earnings are: 30 – 4 + 2 = 28 points. The four member's net earnings are: 35 – 4 + 2 = 33 points. The fifth member's net earnings are: 35 – 4 – 10 + 2 = 23 points.

Please raise your hand if you have any questions.

A.3.2 Democracy treatment

Please read these instructions carefully. Again, you will have to answer some questions to check that you understand the instructions.

This phase, including 10 periods in total, is like the previous one in that you continue to interact with the same four participants in your group and in each period you make a decision about allocating 20 points to either a private account or a group account. Your earnings are determined in the same way as in Phase 1 of the experiment. (For each point you allocate to your private account you get 1 point as earnings. Your earnings from the group account equal the sum of points allocated to the group account by all 5 group members multiplied by 0.4).

However, in this phase, before you begin to make decisions, ONE participant in your group will be elected as a manager by voting. The elected manager will be unchanged throughout this phase.

How is the manager selected?

Between phase 1 and phase 2, you can observe the allocation decisions of each group member in phase 1. Then you will vote for any member in your group (except yourself) to be the Manager. The one who receives the most votes will be elected. In case of a tie, the Manager will be randomly selected from those tied ones. Note that the Manager will be the same during the remaining periods (i.e. until the end of the experiment).

Your decisions in this phase

There are two stages in each period. In the first stage, you make your allocation decision and then observe the decisions of the other group members along with your earnings. In the second stage, 4 points are automatically collected from each group member. Hence each group has an account of 20 points in each period. The manager has an opportunity to use the points in this account to reduce the earnings of one participant in your group.

Suppose you are selected as a manager. After the first stage of each period, you will observe the amount allocated to the group account by each member in your group. At the meanwhile, you will receive 4 points from each group member including yourself. Then you will be asked to choose up to one group member and how much you want to reduce his/her earnings by using the 20 points. Each point you want to reduce that participant's earnings costs 1 point. You are free to leave all group members' earnings unchanged. Note that the points that are not used to reduce others' earnings will be returned to every individual equally.

Earnings in each period are calculated as follows:

20 -4 points – (points you allocate to group account) + 0.4*(sum of points allocated by all in group to group account) – (points the manager uses to reduce your earnings if you are selected) + (the remaining points not used by the manager),

Examples

Suppose you are the manager in your group. In this period, you allocate 20 points to the group account, the second and third members of your group each allocate 10 points to the group account, and the remaining two members both allocate 5 point to the group account. Then the sum of points in the group account is 20 + 10 + 10 + 5 + 5 = 50 points. Each group member receives earnings of 0.4 * 50 = 20 points from the group account. Your gross earnings are: 20 – 20 + (0.4 * 50) = 0 + 20 = 20 points. The second and third members' gross earnings are: 20 – 10 + (0.4 * 50) = 10 + 20 = 30 points. The fourth and fifth members' gross earnings are: 20 – 5 + (0.4 * 50) = 15 + 20 = 35 points. In the second stage, you receive 4 points from each group member, 20 points in total. Now you decide to reduce the fifth member's earning by 10 points. Then the remaining 10 points will be returned to all group members, 2 point for each.

Your net earnings are 20 - 4 + 2 = 18 points. The second and third members' net earnings are: 30 – 4 + 2 = 28 points. The four member's net earnings are: 35 – 4 + 2 = 33 points. The fifth member's net earnings are: 35 – 4 – 10 + 2 = 23 points.

Please raise your hand if you have any questions.

APPENDIX B

CHAPTER 3: ADDITIONAL INFORMATION

B.1 Summary

Table B.1: Demographic summary

	Leviathan			Democracy			p-value
	mean	st dev	obs.	mean	st dev	obs.	Mann-Whitney U test
Female	0.53	0.50	75	0.53	0.50	80	0.642
Econ Major	0.51	0.50	75	0.39	0.49	80	0.000
Math	119.20	13.43	59	120.29	9.85	56	0.637
Freshmen	0.43	0.49	75	0.64	0.48	80	0.000
Sophomore	0.20	0.40	75	0.18	0.38	80	0.075
Junior	0.16	0.37	75	0.05	0.22	80	0.000
Senior	0.11	0.31	75	0.05	0.22	80	0.000
Graduate	0.11	0.31	75	0.09	0.28	80	0.071
Scholarship	0.35	0.48	60	0.40	0.49	65	0.010
Party member	0.12	0.32	60	0.12	0.33	65	0.622
Urban	0.58	0.49	60	0.63	0.48	65	0.015
Major Ethnicity	0.98	0.13	60	0.98	0.12	65	0.799
Income: below 50000	0.22	0.41	60	0.31	0.46	65	0.000
Income: 50000-100000	0.42	0.49	60	0.38	0.49	65	0.102
Income: 100000-200000	0.25	0.43	60	0.18	0.39	65	0.000
Income: above 200000	0.12	0.32	60	0.12	0.33	65	0.622
Single Child	0.38	0.49	60	0.31	0.46	65	0.000

B.2 Robustness checks

Table B.2: Determinants of contributions: robustness checks

	Dependent variable: Contributions (ECU)			
	Fixed Effects: different specifications			
	(1)	(2)	(3)	(4)
Democracy vs.Leviathan (PxD)	1.135***	0.642*	0.753*	0.681*
	(0.390)	(0.346)	(0.404)	(0.401)
Punishment (P)	4.684***	2.280***	2.746***	4.015***
	(0.280)	(0.272)	(0.317)	(0.626)
Democracy (D)	0†	0†	-0.792	-0.733
	(.)	(.)	(0.820)	(0.814)
Other's Avg.Cont. (t-1)		0.647***	0.527***	0.559***
		(0.0218)	(0.0251)	(0.0279)
Punishment received (t-1)		-0.0718*	-0.190***	-0.194***
		(0.0405)	(0.0457)	(0.0454)
Punishment in the group (t-1)		0.00417	0.0266	0.0689***
		(0.0219)	(0.0255)	(0.0266)
Trend (within phase)		-0.230***	-0.193***	-0.466***
		(0.0322)	(0.0375)	(0.0746)
Constant	7.080***	3.560***	6.439***	6.817***
	(0.138)	(0.256)	(0.668)	(0.775)
R2 (overall)	0.136	0.411	0.451	0.463
F	370.2	311.8	66.46	48.02
N	3100	2945	2945	2945
Individual FE	Yes	Yes	No	No
Group FE	No	No	Yes	Yes
Time FE	No	No	No	Yes

Notes: Standard errors in parenthesis. Sample differences are due to the use of lags.
† Variable dropped. Time invariant once controlled for individual FE.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

APPENDIX C

CHAPTER 4: ADDITIONAL INFORMATION

C.1 Summary

Table C.1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Age(in years)	45.168	12.849	20	89	171044
Woman=1	0.413	0.492	0	1	171044
Married=1	0.629	0.483	0	1	171044
Widower=1	0.016	0.124	0	1	171044
Divorced=1	0.073	0.26	0	1	171044
College(pf)=1	0.335	0.472	0	1	171044
Science degree	0.14	0.347	0	1	171044
H/M Responsibility(wd)=1	0.586	0.493	0	1	171044
Mltry/Police(pf)=1	0.041	0.198	0	1	171044
Family Members (#)	0.381	0.89	0	7	171044
Quito	0.416	0.493	0	1	171044
Guayaquil	0.206	0.404	0	1	171044
Cuenca	0.066	0.248	0	1	171044
Ambato	0.034	0.18	0	1	171044
Income(USD)	1840.834	1186.304	700	15000	171044
Tot. Net Worth (USD)	39326.294	57865.434	0	2400000	171044
Tot.CC Debt in CRB	5880.353	7240.986	1	208370.64	147817
Tot.CC Debt/Income	1.676	1.766	-29.989	33.42	171044
% Debt at Risk on CRB	0.03	0.123	0	1	171044
Ownerwhip Type	0.251	0.434	0	1	171044
Properties (#)	0.1	0.321	0	4	171044
Houses(#)	0.59	0.643	0	4	171044
Cars(#)	0.552	0.701	0	4	171044
Debts on CRB(#)	2.729	1.579	0	17	171044
Returned Checks=1	0.023	0.148	0	1	171044
Ytd.Avrg.Payments/Cr.Balance	0.949	0.113	0	1	171044
Microcredit=1	0.044	0.206	0	1	171044
Portfolio/Tot.Debt	0.257	0.285	0	1	171044
Mth.Bal.Payment/Income	0.524	0.535	0	10.212	171044
Time of membership (months)	98.460	95.571	0	552	171044

Table C.2: Average default probability by model and sample

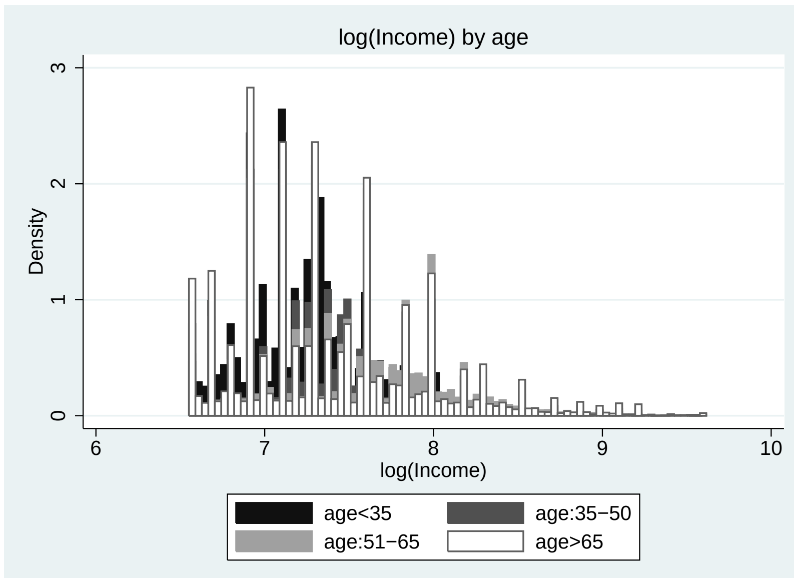
Model/Sample	Overall (%)	Control Sample (%)	Out-of-Sample (%)
Observed (whole sample)	10.31	10.26	10.38
One type			
Predicted default Probit	6.31	6.36	6.25
Predicted default Forward Logit	6.20	6.26	6.11
Type consistent (H/L types)			
Predicted default (unconditional)	*6.47*	*6.52*	*6.40*
Predicted default if H-Type	6.61	6.67	6.53
Predicted default if L-Type	6.08	6.08	6.06
Predicted default (conditional on types)	*15.66*	*15.69*	*15.62*
Predicted default if H-Type	5.71	5.73	5.68
Predicted default if L-Type	73.17	73.13	73.22
N	171,044	102,657	68,387

Conditional results assigns the corresponding outcome based on the class assigned through the posterior probabilities.

C.2 Life cycle patterns and model selection

Figure C.1: Income and debt distribution by age

(a)



(b)

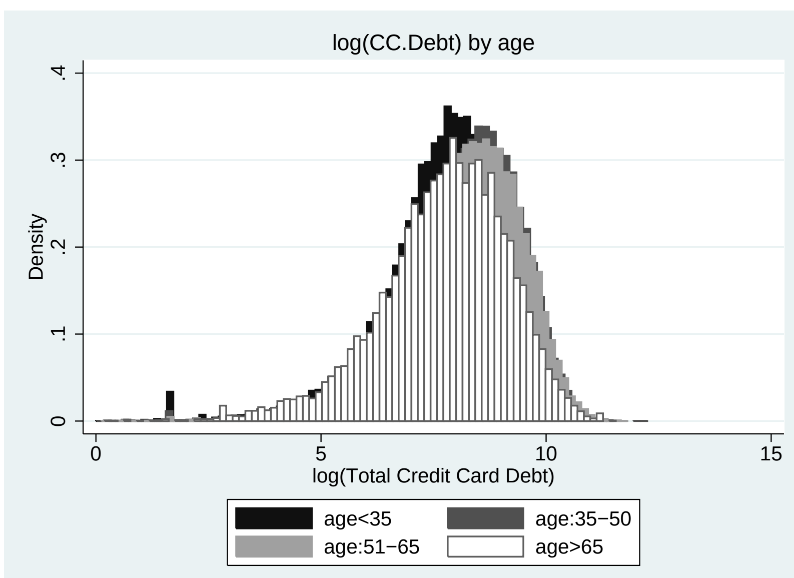
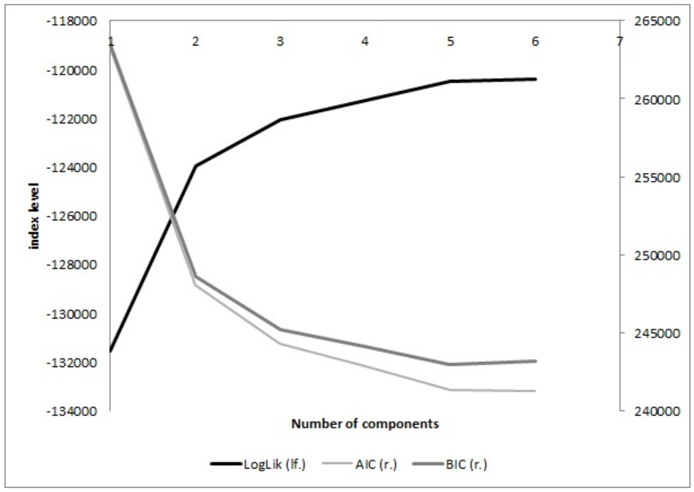


Figure C.2: Model selection by number of components: model fit marginal changes

(a)



(b)