

THREE ESSAYS ON POVERTY ANALYSIS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Juan Carlos Chavez-Martin del Campo

January 2006

© 2006 Juan Carlos Chavez-Martin del Campo

ALL RIGHTS RESERVED

THREE ESSAYS ON POVERTY ANALYSIS

Juan Carlos Chavez-Martin del Campo, Ph.D.

Cornell University 2006

This dissertation is a collection of three essays that cover issues in poverty analysis. The first essay (Partial Identification of Poverty Measures with Contaminated and Corrupted Data) applies a partial identification approach to poverty measurement when data errors are non-classical in the sense that it is not assumed that the error is statistically independent of the outcome of interest, and the error distribution has a mass point at zero. This paper shows that it is possible to find non-parametric bounds for the class of additively separable poverty measures. A methodology to draw statistical inference on partially identified parameters is extended and applied to the setting of poverty measurement. The methodology developed in this essay is applied to the estimation of poverty treatment effects of an anti-poverty program in the presence of contaminated data.

The second essay (On the Design of an Optimal Transfer Schedule with Time Inconsistent Preferences) addresses a very recent literature that studies public policy and its connection to behavioral economics. It incorporates the phenomenon of time inconsistency into the problem of designing an optimal transfer schedule. It is shown that if program beneficiaries are time inconsistent and receive all of the resources in just one payment, then the equilibrium allocation is always inefficient. In the spirit of the second welfare theorem, I also show that any efficient allocation can be obtained in equilibrium when the policymaker has full information. This

assumption is relaxed by introducing uncertainty and asymmetric information into the model. The optimal solution reflects the dilemma that a policymaker has to face when playing the roles of commitment enforcer and insurance provider simultaneously.

The third essay (Does Conditionality Generate Heterogeneity and Regressivity in Program Impacts? The Progresa Experience) studies both empirically and theoretically the consequences of introducing a conditional cash transfer scheme for the distribution of program impacts. Intuitively, if the conditioned-on good is normal, then better-off households tend to receive a larger positive impact. I formalize this insight by means of a simple model of child labor, applying the Nash-Bargaining approach as the solution concept. A series of tests for heterogeneity in program impacts are developed and applied to Progresa, an anti-poverty program in Mexico. It can be concluded that this program exhibits a lot of heterogeneity in treatment effects. Consistent with the model, and under the assumption of rank preservation, program impacts are distributionally regressive, although positive, within the treated population.

BIOGRAPHICAL SKETCH

Juan Carlos Chavez-Martin del Campo was born on May 2, 1973 in Leon, Mexico. He grew up primarily in the small town of Encarnacion de Diaz (La Chona), Jalisco. He was raised by his grand father Ruben Chavez-Aguilera, who gave him the first, non-technical, and most valuable lectures in social justice, particularly on the importance of becoming involved with the poor in order to understand their needs, behavior, and possible ways out of poverty.

In 1997, Juan Carlos received a B.A. in Economics from the Universidad Autonoma de Aguascalientes, Mexico. During his undergraduate studies, he also worked as a high school teacher, specializing in the subjects of mathematics and social sciences, and in small rural development projects in Aguascalientes state. Juan Carlos was president of the Economics Undergraduate Student Association and was actively involved in politics.

Juan Carlos received his M.A. in Economics from El Colegio de Mexico. During his time at El Colegio, Juan Carlos worked on topics related with the Mexican agriculture and rural poverty. After finishing the masters program, he became a lecturer in the Department of Economics at Universidad Autonoma de Aguascalientes.

During his graduate studies at Cornell University, Juan Carlos worked with professors David Just, Ravi Kanbur, Francesca Molinari, and Ted O'Donoghue on a broad set of topics related to the analysis of poverty. In 2002, Juan Carlos married Alejandra Leyva-Parra. Juan Carlos completed his Ph.D. in November 2005 and has accepted a position as an Assistant Professor of Economics at the Guanajuato School of Economics, Mexico. Two weeks after his defense, his first child was born: Alejandra del Carmen.

*A mis abuelos y padres,
por los valores que me han inculcado,
su confianza y sacrificios a lo largo de mi vida.*

*A mis Alejandras:
Mi esposa, por su lealtad y fuerza.
Mi hija, para que nunca olvide de donde viene.*

ACKNOWLEDGEMENTS

It is a cliché to say that this dissertation would not have been possible but for...and in this case this cliché is all too true. To write this dissertation has been the result of countless acts of generosity, support, and guidance of all types.

At Cornell University, I have had the honor of working with an excellent and diverse academic committee: David Just, Ravi Kanbur, Francesca Molinari, and Ted O'Donoghue. They taught me the importance of developing a rigorous argument when making a scientific claim and made me think deeper about my findings, their consistency, and their pertinence. Ravi, my chair, always displayed a great ability to connect fine-tune details with the relevant issues of the problem at hand, which helped me keep my research in perspective. More importantly, I have learned from him the relevance of being consistent with one's personal philosophy.

Throughout my life, my grand parents Lucia and Ruben, and my parents Rosa Elena and Carlos have been the cornerstone where I have found support, values, faith, and counseling. My wife Alejandra has been the unconditional friend, the loyal partner, and the smile and the strength that make me continue in this journey. Most of my successes cannot be explained, and many of my failures could not be overcome without them.

I am very grateful for financial support provided by the Consejo de Ciencia y Tecnologia del Estado de Aguascalientes (CONCYTEA), the Consejo Nacional de Ciencia y Tecnologia (CONACYT), and the Ford-MacArthur-Hewlett Foundation Graduate Fellowship in the Social Sciences.

TABLE OF CONTENTS

1	Partial Identification of Poverty Measures with Contaminated and Corrupted Data	1
1.1	Introduction	2
1.2	Poverty Measurement: Conceptual Framework	4
1.2.1	Specific Poverty Measures	6
1.3	Statement of the Problem	6
1.4	Partial Identification of Poverty Measures	8
1.5	Characterizing Identification Regions	11
1.5.1	Identification Breakdown Points for Poverty Measures	11
1.5.2	Length	13
1.6	Statistical Inference for Partially Identified Poverty Measures	15
1.6.1	Confidence Intervals	16
1.6.2	Hypothesis Testing	18
1.7	Poverty Comparisons	20
1.7.1	Statistical Inference	21
1.8	Application: Evaluation of an Anti-Poverty Program with Missing Treatments	22
1.8.1	Progresa	22
1.8.2	Poverty Treatment Effects	23
1.8.3	Monotone Treatment Response, Data Contamination, and Missing Treatments	26
1.9	Application: Measurement of Rural Poverty in Mexico	27
1.10	Conclusions	28
1.11	Appendix: Proofs	32
2	On the Design of an Optimal Transfer Schedule with Time Inconsistent Preferences	40
2.1	Introduction	41
2.2	The Model	44
2.3	First-Best Consumption Maintenance Programs	47
2.3.1	Transfer Schedules without Commitment: The One-Payment CMP	48
2.3.2	Reestablishing Efficiency through Transfer Schemes	51
2.4	Second-Best Consumption Maintenance Programs	52
2.5	Conclusions	57
2.6	Appendix: Proofs	60
3	Does Conditionality Generate Heterogeneity and Regressivity in Program Impacts?	
	The Progresa Experience	68
3.1	Introduction	69

3.2	Progesa	72
3.2.1	Data: A Quasi-Experimental Design	73
3.2.2	Progesa’s selection of localities and beneficiary households	74
3.3	A Simple Model of Human Capital Investment and CCTS	75
3.3.1	One-Sided Altruism	75
3.3.2	Two-Sided Altruism	77
3.3.3	Conditional Cash Transfers: Efficiency vs Equity	79
3.4	The Evaluation Problem	80
3.4.1	Average Treatment Effects	81
3.4.2	Fréchet Space	83
3.4.3	Partial Identification of Mobility Treatment Effects	84
3.5	Testing for Homogeneity in Program Impacts	87
3.6	Identification of Program Impacts under Monotonicity Assumptions	95
3.7	Conclusions	102
3.8	Appendix A: Proofs and Derivations	104
3.9	Appendix B: Estimation and Bootstrap Algorithm using the Empirical Quantiles	111
3.10	Appendix C: Quantile Treatment Effects	114

LIST OF TABLES

1.1	Identification regions and confidence intervals for treatments effects on poverty: PROGRESA 1999	25
1.2	Identification regions under monotonicity assumptions: PROGRESA 1999	27
1.3	Identification regions and confidence intervals for FGT poverty measures under contamination model: Rural Mexico, 2002	29
1.4	Identification regions and confidence intervals for FGT poverty measures under corruption model: Rural Mexico, 2002	30
3.1	Per Capita Mean Outcomes and Treatment Effects	82
3.2	Mobility Treatment Effects	86
3.3	Average Loss	87
3.4	F test for $H_0 : \frac{\sigma_1}{\sigma_0} = 1$	89
3.5	Levene's statistics	90
3.6	Summary statistics for the bootstrap distribution of $\hat{\Phi}_2^*$ using empirical quantiles of F_0 and F_1	93
3.7	Summary statistics for the bootstrap distribution of $\hat{\Phi}_2^*$ using a $k/\min(n, m)$ resampling scheme.	94
3.8	Quantile Treatment Effects: Total Expenditure	99
3.9	Quantile Treatment Effects: Food Purchase	100

LIST OF FIGURES

3.1	Quantile Treatment Effects: Total Expenditure	101
3.2	Quantile Treatment Effects: Food Purchase	102

Chapter 1
Partial Identification of Poverty
Measures with Contaminated and
Corrupted Data

1.1 Introduction

Much of the statistical analysis of poverty measurement regards the data employed to estimate a poverty measure as error-free. However, it is amply recognized that measurement error is a very common phenomenon for most data sets used in the estimation of poverty. This problem is particularly relevant for developing countries, where the majority of the poor are concentrated, since financial, technological, and logistical constraints are more likely to affect the quality of the data.

Measurement error can affect the estimation of poverty in different ways. For example, the poverty line may be set for heterogeneous groups of people without considering idiosyncratic differences in the cost of basic needs, arbitrary imputations may be made when missing and zero outcomes appear in the sample, and the variable of interest may be misreported by an important subset of survey respondents. Often the methodologies applied to solve these problems are arbitrary; at the same time, the results are highly sensitive to such adjustments. For instance, Szekely, Lustig, Cumpa and Mejia (2000) applied several techniques to adjust for misreporting in Latin America. In the case of Mexico they found that, depending on the method for performing the adjustment, either 14 percent or 76.6 percent of the population is below the poverty line (in absolute terms it implies a difference of 57 million individuals). This has important policy implications since, depending on which of these numbers is used as a reference, the amount of resources directed to social programs can be considered either appropriate or totally insufficient.

Several approaches have been developed in order to analyze the effects of measurement error on poverty measurement. For instance, Chesher and Schluter (2002) study multiplicative measurement error distributed continuously and indepen-

dently of true income to investigate the sensitivity of welfare measures to alternative amounts of measurement error. Ravallion (1994) considers additive random errors when estimating individual-specific poverty lines, finding that heterogeneity in error distributions generates ambiguous poverty rankings. An alternative approach, robust estimation, aims at developing point estimators that are not highly sensitive to errors in the data.¹ The objective is to guard against worse-case scenarios that errors in the data could conceivably produce. In that sense it takes an ex-ante perspective of the problem. Cowell and Victoria-Feser (1996) apply this approach to poverty measurement by using the concept of the influence function to assess the influence of an infinitesimal amount of contamination upon the value of a poverty statistic (Hampel 1974). They find that poverty measures that take as their primitive concept poverty gaps rather incomes of the poor are in general robust under this criterion.

In the present study, we do not consider classical measurement error, that is to say, we do not assume the existence of chronic errors affecting every observation, neither do we assume that the outcome of interest is statistically independent of the error. Instead of assuming that the error distributions have no mass point at zero, we consider the impact of intermittent errors by setting an upper bound to the proportion of gross errors within the data. Since a poverty measure is not point identified under the assumptions of the model of errors under consideration, we follow Horowitz and Manski (1995) and apply a partial identification approach.² By using a fully non-parametric method, we show that for the family of additively

¹See Hampel et al (1986) and Huber (1981) for a comprehensive treatment of robust inference.

²Examples of applications of this approach in other settings are Molinari (2005a) and Dominitz and Sherman (2005). See Manski (2003) for an overview of this literature

separable poverty measures it is possible to find identification regions under very mild assumptions.

The paper is organized as follows. Section 2 introduces some important concepts for poverty measurement. Section 3 states the problem formally, presenting both the contaminated and corrupted sampling models within the context of poverty measurement. Section 4 investigates the identification region for additively separable poverty measures (ASP). It is shown that we can find upper and lower bounds for this class of poverty measures with both contaminated and corrupted data. Section 5 characterizes the identification regions of ASP measures through their length and breakdown points. Section 6 applies two conceptually different types of confidence intervals for partially identified poverty measures. The implications for hypothesis testing when a poverty measure is not point identified are also discussed. Section 7 provides some insight on the effect of both data contamination and data corruption for poverty comparisons. Sections 8 and 9 give two empirical illustrations of the methodology developed in the paper. Most of the mathematical details are in the Appendix.

1.2 Poverty Measurement: Conceptual Framework

Let \mathcal{A} denote the σ -algebra of Lebesgue measurable sets on \mathbb{R} . Let \mathcal{P} denote the set of all probability distributions on $(\mathbb{R}, \mathcal{A})$. Thus for any $P \in \mathcal{P}$ the triple $(\mathbb{R}, \mathcal{A}, P)$ is a probability space. Let $z \in \mathbb{R}_{++}$ be the poverty line.

A person is said to be in poverty if her income, $y \in \mathbb{R}$, or any other measure of her economic status is strictly below z . An aggregate poverty index is defined as a functional of the distribution $P \in \mathcal{P}$. Formally:

Definition 1.1 *A Poverty Index is a functional $\Pi(P; z) : \mathcal{P} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ that*

indicates the degree of poverty when a particular variable has distribution P and the poverty line is z .

An important type of poverty measures is the *Additively Separable Poverty*(ASP) class which is defined as follows:

$$\Pi(P; z) = \int \pi(y; z)dP \quad (1.1)$$

where $\pi(y; z) : \mathbb{R} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ is the poverty evaluation function, an indicator of the severity of poverty for a person with income y when the poverty line is fixed at z .

Since the axiomatic approach to poverty measurement proposed by Sen (1976), most economists interested in the phenomenon of poverty have quantified poverty in a manner consistent with those principles. One of those principles, the *focus axiom*, requires a poverty measure to be independent of the income distribution of the non poor. The *monotonicity axiom* says that, everything else equal, a reduction in the income of a poor individual must increase the poverty measure; the *transfer axiom* emphasizes the positive effect of a regressive transfer on the poverty measure, that is to say, given other things, a pure transfer of income from a poor individual to any other individual that is richer must increase the poverty measure. Finally, Kakwani (1980) has proposed a 4th property that prioritizes transfers taking place down in the distribution, other things being equal. These distributional concerns are made operational through the characteristics of the poverty evaluation function $\pi(y; z)$. It is usually assumed that $\pi(y; z)$ is continuous for $y < z$, non increasing in its first argument and non decreasing in its second argument. It is also assumed that $\pi(y; z)$ is convex in its first argument and $\pi(y; z) = 0$ for $y \geq z$.

1.2.1 Specific Poverty Measures

Watts (1968) proposed a poverty measure which is defined as follows:

$$\Pi_W = \int \mathbf{1}(y < z) \ln\left(\frac{y}{z}\right) dP \quad (1.2)$$

This poverty measure satisfies Sen's monotonicity and transfer axioms as well as Kakwani's transfer-sensitivity axiom.

Foster, Greer and Thorbecke (1984) proposed an α -class of poverty measures, Π_α , which can be obtained by:

$$\Pi_\alpha = \int \mathbf{1}(y < z) \left(1 - \frac{y}{z}\right)^\alpha dP \quad (1.3)$$

Π_α satisfies monotonicity axiom for $\alpha > 0$, transfer axiom for $\alpha > 1$, and transfer sensitivity axiom for $\alpha > 2$.

Hagenaars (1987) provided a poverty measure that satisfies all three axioms. The specific poverty measure he gave is

$$\Pi_H = \int \mathbf{1}(y < z) \left(1 - \frac{\ln y}{\ln z}\right) dP \quad (1.4)$$

Finally, we consider the Clark, Hemming and Ulph (1981) poverty measure:

$$\Pi_\beta = \frac{1}{\beta} \int \mathbf{1}(y < z) \left(1 - \left(\frac{y}{z}\right)^\beta\right) dP \quad (1.5)$$

which satisfies the monotonicity axiom for all $\beta > 0$, and both transfer axioms for $\beta < 1$. Finally, Chakravarty (1983) derived a poverty measure which is equal to $\Pi_{Ch} = \beta\Pi_\beta$. This measure also satisfies all three axioms for $\beta \in (0, 1)$.

1.3 Statement of the Problem

Let each member j of population J be characterized by the tuple (y_1^j, y_0^j) in the space $\mathbb{R} \times \mathbb{R}$, where y_1^j is the outcome of interest denoting the "true" equivalent income (or expenditure) for a given poverty line z . Let the random variable

$(y_1, y_0) : J \longrightarrow \mathbb{R} \times \mathbb{R}$ have distribution $P(y_1, y_0)$. Let a random sample be drawn from $P(y_1, y_0)$. Let's assume that instead of observing y_1 , one observes a random variable y defined by:

$$y \equiv wy_1 + (1 - w)y_0 \tag{1.6}$$

Realizations of y with $w = 0$ are said to be data errors, those with $w = 1$ are error-free, and y itself is a contaminated version of y_1 . Let $Q(y)$ denote the distribution of the observable y . Let $P_i = P_i(y_i)$ denote the marginal distribution of y_i . Let $P_{ij} = P_{ij}(y_i | w = j)$ denote the distribution of y conditional on the event $w = j$ for $i, j \in \{0, 1\}$. Let $p = P(w = 0)$ be the marginal probability of a data error. With data errors, the sampling process does not identify P_1 (the object of interest) but only $Q(y)$, the distribution of the observable y . By the law of total probability, these two distributions can be decomposed as follows:

$$P_1 = (1 - p)P_{11} + pP_{10} \tag{1.7}$$

$$Q(y) = (1 - p)P_{11} + pP_{00} \tag{1.8}$$

This problem can be approached from different perspectives. In robust estimation P_1 is held fixed and $Q(y)$ is allowed to range over all distributions consistent with both equations. In the context of poverty measurement, the objective would be to estimate the maximum possible distance between $\Pi(Q; z)$ and $\Pi(P_1; z)$. In contrast, the present analysis holds $Q(y)$ fixed because it is identified by the data, and P_1 is allowed to range over all distributions consistent with (3) and (4). This approach recognizes that the parameter of interest might not be point identified, but it can often be bounded.

The sampling process reveals only the distribution $Q(y)$. However, informative identification regions emerge if knowledge of the empirical distribution is combined

with a non-trivial upper bound, λ , on p .

This investigations analyzes two different cases of data errors. In the first case, we will assume that the occurrence of data errors is independent of the sample realizations from the population of interest. Formally

$$P_1 = P_{11} \tag{1.9}$$

This particular model of data errors is known as "contaminated data" or "contaminated sampling" model (Huber 1981). In the other case, (9) does not hold and it is only assumed that there exists a non-trivial upper bound on the error probability. Horowitz and Manski (1995) refer to this case as "corrupted sampling".

Define the sets

$$\mathcal{P}_1(p) \equiv \mathcal{P} \cap \{(1-p)\phi_{11} + p\phi_{10} : (\phi_{11}, \phi_{10}) \in \mathcal{P}_{11}(p) \times \mathcal{P}\} \tag{1.10}$$

$$\mathcal{P}_{11}(p) \equiv \mathcal{P} \cap \left\{ \frac{Q - p\phi_{00}}{1-p} : \phi_{00} \in \mathcal{P} \right\} \tag{1.11}$$

If there exists a non-trivial upper bound, λ , on the probability of data errors, then it can be proved that P_{11} and P_1 belong to the sets $\mathcal{P}_{11}(\lambda)$ and $\mathcal{P}_1(\lambda)$ respectively, where $\mathcal{P}_{11}(\lambda) \subset \mathcal{P}_1(\lambda)$. These restrictions are sharp in the sense that they exhaust all the available information, given the maintained assumptions (Horowitz and Manski 1995).

1.4 Partial Identification of Poverty Measures

Suppose that a proportion $p < 1$ of the data is erroneous. Furthermore, assume there exists a non-trivial upper bound, λ , on p : $p \leq \lambda < 1$.³ From the analysis

³In practice, upper bounds on the probability of data errors can be estimated from a validation data set or by the proportion of imputed data in the sample. See Kreider and Pepper (2004) for an application of a validation model.

above, we know that the distribution of interest P_1 is not identified: i.e. $\mathcal{P}_1(\lambda)$ is not a singleton.

Even though P_1 is not point identified, it is partially identified in the sense that it belongs to the identification region $\mathcal{P}_1(\lambda)$. There is a mapping from this set into the domain in \mathbb{R} of a given poverty measure Π . Therefore, the question arises whether there is a way to characterize the identification region of Π . As we will see below, it is possible to do so for the class of ASP poverty measures by ordering the distributions in \mathcal{P}_λ according to a stochastic dominance criterion. Such criterion is defined as follows:

Definition 1.2 *Let $F, G \in \mathcal{P}$. Distribution F First Order Stochastically dominates (FOD) distribution G if*

$$F((-\infty, x]) \leq G((-\infty, x])$$

for all $x \in \mathbb{R}$.

In the case of monotone functions, there is a well-known result that will be helpful to obtain identification regions for the ASP measures:

Lemma 1.1 *The Distribution F first-order stochastically dominates the distribution G if and only if, for every non decreasing function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, we have*

$$\int \varphi(x) dF(x) \geq \int \varphi(x) dG(x) \tag{1.12}$$

Finally, let me introduce a basic concept that is a building block for identification regions.

Definition 1.3 *For $\alpha \in (0, 1]$, the α -quantile of $Q(y)$ is given by*

$$r(\alpha) = \inf\{t : Q((-\infty, t]) \geq \alpha\}.$$

Now we can state the main result of this section. Following the approach of Horowitz and Manski (1995) to find sharp bounds on parameters that respect stochastic dominance ⁴ we can construct identification regions for ASP measures.

Proposition 1.1 *Let it be known that $p \leq \lambda < 1$. Define probability distributions L_λ and U_λ on \mathbb{R} as follows:*

$$L_\lambda = \begin{cases} \frac{Q(y \leq t)}{1-\lambda} & \text{for } t < r(1-\lambda) \\ 1 & \text{otherwise} \end{cases}$$

$$U_\lambda = \begin{cases} 0 & \text{for } t < r(\lambda) \\ \frac{Q(y \leq t) - \lambda}{1-\lambda} & \text{otherwise} \end{cases}$$

If $\Pi(P; z)$ belongs to the family of Additively Separable Poverty Measures and the poverty evaluation function is non-increasing in y , then identification regions for $\Pi(P_{11}; z)$ and $\Pi(P_1; z)$ are given by:

$$\mathbf{H}[\Pi(P_{11}; z)] = [\Pi_l(U_\lambda; z), \Pi_u(L_\lambda; z)] \quad (1.13)$$

and

$$\mathbf{H}[\Pi(P_1; z)] = [(1-\lambda)\Pi_l(U_\lambda; z) + \lambda\psi_0, (1-\lambda)\Pi_u(L_\lambda; z) + \lambda\psi_1] \quad (1.14)$$

where $\psi_0 = \inf_{y \in \mathbb{R}_+} \pi(y; z)$ and $\psi_1 = \sup_{y \in \mathbb{R}_+} \pi(y; z)$.

PROOF: See Appendix.

These results are quite intuitive. In the case of contaminated data, the smallest feasible value of $\Pi(P_{11}; z)$ occurs when we place all of the erroneous data as far out as possible in the left-hand tail of the observed distribution Q . Similarly, to

⁴A parameter $\delta(\cdot)$ respects stochastic dominance if $\delta(F) \geq \delta(G)$ whenever F FOD G.

obtain the largest feasible value of $\Pi(P_{11}; z)$, L_λ places all of the erroneous data as far out as possible in the right-hand tail of the observed income distribution. If the data is corrupted, we follow a similar procedure, placing all of the erroneous data at $\inf_y \pi(y; z)$ and $\sup_y \pi(y; z)$ instead.

Example 1.1 *Assume $P_1 = P_{11}$. Let $Q(y) = U[0, 1]$, $0 < p < \lambda < z < 1 - \lambda$. Let the poverty measure be given by $\varphi = \int_0^\infty 1(y < z) d\phi$. Then, $\varphi(P_1; z) \in [\frac{z-\lambda}{1-\lambda}, \frac{z}{1-\lambda}]$. If $P_1 \neq P_{11}$ then $\varphi(P_1; z) \in [z - \lambda, z + \lambda]$. Notice that $\varphi(Q; z)$ belongs to both intervals.*

1.5 Characterizing Identification Regions

The objective of this section is to describe the properties of the identification region for ASP measures. Our approach is not normative in that we are not arguing that one poverty measure is better than another based on our findings. We analyze identification regions through two concepts: identification breakdown points and length of the identification region, with the hope of shedding some light on the identification properties of poverty measures.

1.5.1 Identification Breakdown Points for Poverty Measures

We denote by \mathcal{D} the family of ASP measures indexed by j with poverty evaluation function $\pi_j(y; z)$ satisfying $\pi_j(y; z) = 0$ for all $y \geq z$, increasing in its second argument, decreasing in its first argument, and continuous and convex for all $y < z$. Moreover, we assume the existence of a constant $c_j \in \mathbb{R}_+$ such that $\pi_j(0; z) = c_j$. We denote by $\mathfrak{R}_j = \{\Pi_j(P; z); P \in \mathcal{P}\}$ the range of a poverty measure $\Pi_j(P; z)$ in

\mathcal{D} . More precisely, the range of a poverty measure in \mathcal{D} is given by $\mathfrak{R}_j = [0, c_j]$.⁵

From the literature on robust estimation we borrow the concept of breakdown point which in the present setting can be interpreted as *the largest fraction of erroneous data that can be in a sample without driving a poverty measure to either boundary of its range*. However, as noticed by Horowitz and Manski (1995), there are some conceptual differences between the breakdown point in robust estimation and its counterpart in identification analysis. While in the partial identification approach λ is evaluated at the empirical distribution Q , in robust estimation it is evaluated at the distribution of interest P_1 . More formally, *the identification breakdown point of a poverty measure* $\Pi(P; z)$ when data are contaminated can be constructed as follows: for some ASP measure in \mathcal{D} define

$$\phi(\lambda_j) = \Pi_j(L_\lambda; z) - c_j \quad (1.15)$$

$$\psi(\lambda_j) = \Pi_j(U_\lambda; z) \quad (1.16)$$

and let $\lambda_j^\phi = \sup\{\lambda : \phi_j(\lambda) < 0\}$, and $\lambda_j^\psi = \sup\{\lambda : \psi_j(\lambda) > 0\}$. The identification breakdown point for an ASP measure is given by:

$$\lambda_j^* = \min\{\lambda_j^\phi, \lambda_j^\psi\} \quad (1.17)$$

Let $H_Q = \int 1(y < z)dQ$ be the head-count ratio or proportion of the poor for the observed distribution Q , and let λ^H be its breakdown point. Clearly the head-count ratio is an element of \mathcal{D} . We have the following proposition:

Proposition 1.2 *For all $Q \in \mathcal{P}$, we have*

$$\lambda^H = \inf\{\lambda_j^* : j \in \mathcal{D}\}$$

⁵Most of the ASP measures used in empirical work belong to this class. For example, the Foster, Greer and Thorbecke (1984), and the Clark, Hemming and Ulph (1981) families of poverty measures are two elements of \mathcal{D} .

PROOF: See Appendix.

Therefore, the breakdown point for the head-count ratio is a lower-bound of the set \mathcal{D} .

1.5.2 Length

Another way to "compare" the different poverty measures is through the length of their identification regions. Although we are not arguing here that one can choose one poverty measure over another based on this criterion, the results obtained in this section provide some initial insights about the behavior of the different poverty measures for the model of errors under consideration. To formalize the analysis, let $m : \mathcal{B} \rightarrow \mathbb{R}_+$ be the Lebesgue measure on the Borel sets, \mathcal{B} , of \mathbb{R}_+ . Here is the main result of this section:

Proposition 1.3 *Let $\Pi_1(P; z) : \mathcal{P} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ and $\Pi_2(P; z) : \mathcal{P} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ be two additively separable poverty measures with non-increasing evaluation functions $\pi_1(y; z) : \mathbb{R}_{++} \times \mathbb{R} \rightarrow \mathbb{R}$ and $\pi_2(y; z) : \mathbb{R}_{++} \times \mathbb{R} \rightarrow \mathbb{R}$, respectively. Suppose that $\pi_2(y; z) \geq \pi_1(y; z)$ for all $y < z$ and $\pi_2(y; z) = \pi_1(y; z)$, otherwise. Let $z \leq \max\{r(\lambda), r(1 - \lambda)\}$ If the data is either corrupted or contaminated, then*

$$m(\mathbf{H}[\Pi_2]) \geq m(\mathbf{H}[\Pi_1])$$

PROOF: See Appendix.

We can get a similar result by imposing more assumptions on the "shape" of the poverty evaluation function. In particular, we can use the fact that some families of poverty measures are generated by "convexifying" a poverty evaluation function in order to show the existence of length orderings within families of poverty measures. The following two corollaries state this result more formally:

Corollary 1.1 *Let $\pi_1(y; z) : \mathbb{R}_+ \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ be a non increasing, and continuous on $(0, z)$ poverty evaluation function, with $z \leq \max\{r(\lambda), r(1 - \lambda)\}$, and f be a convex function on $\pi_1(y; z)$ such that*

$$\mathbf{A1.} \quad \pi_2(y; z) = f \circ \pi_1(y; z)$$

$$\mathbf{A2.} \quad f(\pi_1(0; z)) \leq \pi_1(0; z)$$

$$\mathbf{A3.} \quad f(\pi_1(z; z)) = \pi_1(z; z)$$

$$\text{Then } m(\mathbf{H}[\Pi_2]) \leq m(\mathbf{H}[\Pi_1]).$$

PROOF: See Appendix.

Corollary 1.2 *Given two continuous poverty evaluation functions $\pi_1(y; z) : \mathbb{R}_+ \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ and $\pi_2(y; z) : \mathbb{R}_+ \times \mathbb{R}_{++} \rightarrow \mathbb{R}$, $z \leq \max\{r(\lambda), r(1 - \lambda)\}$ such that*

$$\mathbf{A4.} \quad \pi_1(0; z) = \pi_2(0; z)$$

$$\mathbf{A5.} \quad \pi_1(y; z) = \pi_2(y; z), \text{ for all } y \geq z$$

$$\mathbf{A6.} \quad \pi_i' < 0, \pi_i'' > 0 \text{ on } (0, z), i = 1, 2$$

$$\mathbf{A7.} \quad -\frac{\pi_1(y; z)''}{\pi_1(y; z)'} \geq -\frac{\pi_2(y; z)''}{\pi_2(y; z)'} \text{ uniformly on } (0, z)$$

$$\text{Then } m(\mathbf{H}[\Pi_2]) \leq m(\mathbf{H}[\Pi_1]).$$

PROOF: See Appendix.

Example 1.2 *An α -ordering.*

Let $\pi_\alpha(y; z) : \mathbb{R}_+ \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ be defined as follows:

$$\pi_\alpha = \begin{cases} (1 - \frac{y}{z})^\alpha & \text{if } y \in [0, z) \\ 0 & \text{if } y > z \end{cases}$$

Define $f_c(x) = x^c$. Clearly this is a convex function for all $x > 0$ and $c \geq 1$.

Take any positive integers α_1 and α_2 such that $\alpha_1 = k\alpha_2 > 0$. Hence,

$$\pi_{\alpha_1}(y; z) = f_k \circ \pi_{\alpha_2}$$

By Corollary 1.1, $m_{\alpha_2} \geq m_{\alpha_1}$.

Example 1.3 *An e-ordering*

Let $\pi_{Ch}(y; z) : \mathbb{R}_+ \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ be defined as follows:

$$\pi_{Ch} = \begin{cases} 1 - \left(\frac{y}{z}\right)^e & \text{if } y \in [0, z) \\ 0 & \text{if } y > z \end{cases}$$

for $e \in (0, 1)$. After some algebraic manipulations we have:

$$-\frac{\pi''_{Ch}}{\pi'_{Ch}} = \frac{(1-e)}{y}$$

on $(0, z)$. Therefore, $m_e(\mathbf{H}[\Pi_{Ch}])$ is decreasing on e by Corollary 1.2.

Example 1.4 *Length rankings.*

Let $\alpha \geq 1$, $\beta \in (0, 1)$, $e \in (0, 1)$. Then it is easy to show, applying Proposition 1.3, that the following length rankings hold: $m(\mathbf{H}[\Pi_W]) \geq m(\mathbf{H}[\Pi_C]) \geq m(\mathbf{H}[\Pi_\alpha])$ and $m(\mathbf{H}[\Pi_W]) \geq m(\mathbf{H}[\Pi_C]) \geq m(\mathbf{H}[\Pi_{Ch}])$.

1.6 Statistical Inference for Partially Identified Poverty Measures

In this section, we obtain two conceptually different types of confidence sets for the identification regions of poverty measures. The first type of confidence set uses the Bonferroni's inequality to develop confidence intervals that asymptotically cover the entire identification region with at least probability γ . For the second type of confidence set, we follow Imbens and Manski (2004) by applying confidence intervals that asymptotically cover the true value of the poverty measure with at least this probability. We also discuss some implications of this methodology for hypothesis testing in the context of partially identified poverty measures.

1.6.1 Confidence Intervals

Let $(\mathbb{R}, \mathcal{A}, Q)$ be a probability space, and let \mathcal{P} be a space of probability distributions. The distribution Q is not known, but a random sample y_1, y_2, \dots, y_n is available.

In the point identified case ($\lambda = 0$), a consistent estimator of the class of ASP measures is given by

$$\hat{\Pi} = \frac{1}{n} \sum_{i=1}^n \pi(y_i; z) \quad (1.18)$$

where $\pi(y; z)$ is a measurable poverty evaluation function. By applying The Central Limit Theorem, the standard $100 \cdot \gamma\%$ confidence interval for $\Pi(P; z)$ is given by:

$$CI_{\gamma}^{\Pi} = \left[\hat{\Pi} - z_{\frac{\gamma+1}{2}} \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\Pi} + z_{\frac{\gamma+1}{2}} \frac{\hat{\sigma}}{\sqrt{n}} \right] \quad (1.19)$$

where $\hat{\sigma} = \sigma + o_p(1)$ and z_{τ} is the τ quantile of the standard normal distribution.⁶

To derive the asymptotic properties for the Bonferroni confidence set, we will make use of a result on L-statistics due to Stigler (1973), who explores the asymptotic behavior of trimmed means. Define the confidence interval $CI_{\gamma}^{[\Pi_l, \Pi_u]}$ as

$$CI_{\gamma}^{[\Pi_l, \Pi_u]} = \left[\hat{\Pi}_l - z_{\frac{\gamma+1}{2}} \frac{\hat{\sigma}_l}{\sqrt{n}}, \hat{\Pi}_u + z_{\frac{\gamma+1}{2}} \frac{\hat{\sigma}_u}{\sqrt{n}} \right] \quad (1.20)$$

Where $\hat{\Pi}_l$, $\hat{\Pi}_u$, $\hat{\sigma}_l^2$, and $\hat{\sigma}_u^2$ are estimators satisfying, respectively

A8. $\hat{\Pi}_l = \Pi_l + o_p(1)$

A9. $\hat{\Pi}_u = \Pi_u + o_p(1)$

A10. $\hat{\sigma}_l^2 = \frac{Var_{U_{\lambda}}(\pi(y; z)) + (\pi(r(1-\lambda)) - \Pi_l)\lambda}{1-\lambda} + o_p(1)$

A11. $\hat{\sigma}_u^2 = \frac{Var_{L_{\lambda}}(\pi(y; z)) + (\pi(r(\lambda)) - \Pi_u)\lambda}{1-\lambda} + o_p(1)$

We have the following result

⁶Kakwani (1993) describes this methodology for ASP measures.

Proposition 1.4 *Let $\lambda < 1$ be known. Assume $\int \pi(y; z)^2 dQ < \infty$. Let $r(1 - \lambda)$ and $r(\lambda)$ be continuity points of $Q(y)$. Let the poverty evaluation function, $\pi(y; z)$, be a non-increasing function that is continuous at $r(1 - \lambda)$ and $r(\lambda)$. Then*

$$\lim_{n \rightarrow \infty} \mathbb{P}([\Pi_l, \Pi_u] \subset CI_\gamma^{[\Pi_l, \Pi_u]}) \geq \gamma \quad (1.21)$$

PROOF: See Appendix.

For the second type of confidence interval, define $\Delta = \Pi_U - \Pi_L$ and $\hat{\Delta} = \hat{\Pi}_U - \hat{\Pi}_L$ and consider the following set of regularity conditions, which are equivalent to the assumptions imposed by Imbens and Manski (2004).⁷

A13. $Q(y) \in \mathcal{F}$, where \mathcal{F} is the set of distribution functions for which $\int |\pi(y; z)|^3 dQ < \infty$, Q'' is bounded in the neighborhoods of $r(\lambda)$ and $r(1 - \lambda)$ while $Q'(r(\lambda)) > 0$ and $Q'(r(1 - \lambda)) > 0$.

A14. $\underline{\sigma}^2 \leq \sigma_l^2, \sigma_u^2 \leq \bar{\sigma}^2$ for some positive and finite $\underline{\sigma}^2$ and $\bar{\sigma}^2$.

A15. $\Pi_u - \Pi_l \leq \bar{\Delta} < \infty$

A16. For all $\epsilon > 0$ there are $\nu > 0$, K and n_0 such that $n \geq n_0$ implies $Pr\left(\sqrt{n} |\hat{\Delta} - \Delta| > K\Delta^\nu\right) < \epsilon$, uniformly in $Q \in \mathcal{F}$.

Define the confidence interval \overline{CI}_γ^Π as:

$$\overline{CI}_\gamma^\Pi = \left[\hat{\Pi}_l - \frac{\overline{C}_n \hat{\sigma}_l}{\sqrt{n}}, \hat{\Pi}_u + \frac{\overline{C}_n \hat{\sigma}_u}{\sqrt{n}} \right] \quad (1.22)$$

where \overline{C}_n satisfies

$$\Phi\left(\overline{C}_n + \sqrt{n} \frac{\hat{\Delta}}{\max(\hat{\sigma}_l, \hat{\sigma}_u)}\right) - \Phi(-\overline{C}_n) = \gamma \quad (1.23)$$

⁷More precisely, we have made use of the results on uniform convergence of trimmed means developed by De Wett (1976) to develop a set of regularity conditions equivalent to those required by Imbens and Manski (2004) to obtain their asymptotic result.

Proposition 1.5 *Let $\lambda < 1$. Let $r(1 - \lambda)$ and $r(\lambda)$ be continuity points of $Q(y)$. Let the poverty evaluation function, $\pi(y; z)$, be a non-increasing function that is continuous at $r(1 - \lambda)$ and $r(\lambda)$. Suppose **A13-A16** hold. Then*

$$\lim_{n \rightarrow \infty} \inf_{P \in \mathcal{P}} \mathbb{P} \left(\Pi \in \overline{CI}_{\gamma}^{\Pi} \right) \geq \gamma \quad (1.24)$$

PROOF: See Appendix.

1.6.2 Hypothesis Testing

Consider the implications of testing hypothesis of the form:

$$H_0 : \Pi = \Pi_0$$

versus

$$H_1 : \Pi \neq \Pi_0$$

When a parameter is not point identified, the power of a test is not a straightforward extension of the point identified case. For instance, consider the test

$$\text{reject } H_0 \text{ if } \frac{\sqrt{n}(\hat{\Pi}_l - \Pi_0)}{\sigma_l} > z_{\frac{\gamma+1}{2}} \text{ or } \frac{\sqrt{n}(\hat{\Pi}_u - \Pi_0)}{\sigma_u} < -z_{\frac{\gamma+1}{2}}$$

The rejection region is

$$R = \left\{ (y_1, \dots, y_n) : \frac{\sqrt{n}(\hat{\Pi}_l - \Pi_0)}{\sigma_l} < z_{\frac{\gamma+1}{2}} \text{ or } \frac{\sqrt{n}(\hat{\Pi}_u - \Pi_0)}{\sigma_u} < -z_{\frac{\gamma+1}{2}} \right\}$$

and the power function is defined by

$$\beta_n(\Pi) = \mathbb{P}_{\Pi}((y_1, \dots, y_n) \in R)$$

Define the events

$$A_n = \left\{ Y_n + \frac{\sqrt{n}(\Pi_l - \Pi_0)}{\sigma_l} > z_{\frac{\gamma+1}{2}} \right\}$$

$$B_n = \left\{ Z_n + \frac{\sqrt{n}(\Pi_u - \Pi_0)}{\hat{\sigma}_u} < -z_{\frac{\gamma+1}{2}} \right\}$$

where $Y_n = \frac{\sqrt{n}(\hat{\Pi}_l - \Pi_l)}{\hat{\sigma}_l}$ and $Z_n = \frac{\sqrt{n}(\hat{\Pi}_u - \Pi_u)}{\hat{\sigma}_u}$. From proposition 1.4, we can deduce that this test has a level $1 - \gamma$ since

$$\begin{aligned} \lim_{n \rightarrow \infty} \beta_n(\Pi_0) &= \lim_{n \rightarrow \infty} \mathbb{P}(A_n \cup B_n) \\ &\leq 1 - \lim_{n \rightarrow \infty} \mathbb{P}([\Pi_l, \Pi_u] \subset CI_\gamma^{[\Pi_l, \Pi_u]}) \\ &\leq 1 - \gamma \end{aligned}$$

Next, suppose the true value of Π is $\Pi^* \neq \Pi_0$. If Π is point identified, the probability of correctly rejecting the null hypothesis, H_0 , tends to 1 asymptotically. On the other hand, if the parameter is not point identified, the power of the test for values other than Π_0 is not longer equal to one in general. To verify that this is the case, it will be helpful to divide the analysis in several cases:

$$i) \Pi_0 \in [\Pi_l, \Pi_u]$$

In this case $\lim_{n \rightarrow \infty} \beta_n(\Pi^*) = \lim_{n \rightarrow \infty} \beta_n(\Pi_0) \leq 1 - \gamma$. Hence, a type II error is more likely to arise whenever Π_0 belongs to the identification region.

$$ii) \Pi_0 < \Pi_l$$

Notice that

$$\begin{aligned} \lim_{n \rightarrow \infty} \beta_n(\Pi^*) &\geq \lim_{n \rightarrow \infty} \mathbb{P}(A_n \cup B_n) \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}(\hat{\Pi}_l - \Pi_l)}{\hat{\sigma}_l} + \frac{\sqrt{n}(\Pi_l - \Pi_0)}{\hat{\sigma}_l}\right) \\ &= 1 \end{aligned}$$

where I have used the fact that $\frac{\sqrt{n}(\hat{\Pi}_l - \Pi_l)}{\hat{\sigma}_l} + \frac{\sqrt{n}(\Pi_l - \Pi_0)}{\hat{\sigma}_l}$ will converge to $+\infty$ in probability. Since $\beta(\Pi^*)$ is a probability measure, we have $\lim_{n \rightarrow \infty} \beta_n(\Pi^*) = 1$.

iii) $\Pi_0 > \Pi_u$

By a similar argument to the one applied in *ii*), we have $\lim_{n \rightarrow \infty} \beta_n(\Pi^*) = 1$.

Interestingly, the power $\beta(\Pi^*)$ is a decreasing function of λ because the size of the identification region is positively related to it: the larger the value of the upper bound λ , the more likely it is that Π_0 belongs to the identification region, implying a higher a probability that a type II error will occur.

1.7 Poverty Comparisons

This section addresses both identification and inference problems when comparing some poverty measure between two populations and data errors are generated by the models under consideration. The problem is formulated as follows: there are two populations, A and B, characterized by distributions F and G , respectively. Moreover, we assume the existence of upper bounds λ_A and λ_B on the proportion of data errors. We are interested in comparing, in terms of some ASP measure, the two populations.

Define the difference between the poverty measures corresponding to distribution F and G as $D = \Pi(F; z) - \Pi(G; z)$. Proposition 1.1 can be used to obtain informative, although not necessarily sharp, outer bounds on D given λ_A and λ_B .⁸⁹

Proposition 1.6 *Let it be known that $p_A \leq \lambda_A < 1$ and $p_B \leq \lambda_B < 1$. If $\Pi(P; z)$*

⁸In principle, it is not necessary to restrict both distributions to have same type of data errors. For instance, distribution A could be characterized by contaminated data while distribution B by corrupted data. The analysis and conclusions would not change by including that level of detail.

⁹As noticed by Manski (2003), outerbounds on differences between parameters that respect stochastic dominance are generally non-sharp. In the present case, for these to be sharp, there would have to exist two distributions of errors that jointly make $\Pi(F; z)$ and $\Pi(G; z)$ attain their sharp bounds.

belongs to the family of additively separable poverty measures and the poverty evaluation function is non-increasing in y , then identification regions for $D(F_{11}, G_{11}; z)$ and $D(F_1; G_1; z)$ are given by

$$\mathbf{H}[D(F_{11}, G_{11}; z)] = [\Pi_{\lambda_A}^l(F; z) - \Pi_{\lambda_B}^u(G; z), \Pi_{\lambda_A}^u(F; z) - \Pi_{\lambda_B}^l(G; z)] \quad (1.25)$$

and

$$\mathbf{H}[D(F_1, G_1; z)] = [D_1^l, D_1^u] \quad (1.26)$$

where

$$D_1^u = (1 - \lambda_A)\Pi_{\lambda_A}^u(F; z) - (1 - \lambda_B)\Pi_{\lambda_B}^l(G; z) + \lambda_A\psi_1 - \lambda_B\psi_0$$

$$D_1^l = (1 - \lambda_A)\Pi_{\lambda_A}^l(F; z) - (1 - \lambda_B)\Pi_{\lambda_B}^u(G; z) + \lambda_A\psi_0 - \lambda_B\psi_1$$

1.7.1 Statistical Inference

Let y_1, \dots, y_n and y_1, \dots, y_m be two independent random samples drawn from F and G , respectively. We will construct confidence intervals for the identification region of the poverty difference $\Pi_A - \Pi_B$.

Define the confidence interval $CI_\gamma^{D_l, D_u}$ as follows

$$CI_\gamma^{[D_l, D_u]} = \left[\hat{\Pi}_l(F) - \hat{\Pi}_u(G) - z_{\frac{\gamma+1}{2}} \hat{\sigma}^*, \hat{\Pi}_u(F) - \hat{\Pi}_l(G) + z_{\frac{\gamma+1}{2}} \hat{\sigma}^{**} \right] \quad (1.27)$$

where

$$\hat{\sigma}^* = \sqrt{\frac{\hat{\sigma}_{lF}^2}{n} + \frac{\hat{\sigma}_{uG}^2}{m}}$$

$$\hat{\sigma}^{**} = \sqrt{\frac{\hat{\sigma}_{uF}^2}{n} + \frac{\hat{\sigma}_{lG}^2}{m}}$$

Proposition 1.7 *Let $\lambda_i < 1$, $i = A, B$ be known,. Assume $E_i(\pi(y; z)^2) < \infty$. Let $r_i(1 - \lambda_i)$ and $r_i(\lambda_i)$ be continuity points and let $m, n \rightarrow \infty$ such that $\frac{m}{m+n} \rightarrow \epsilon \in$*

$(0, 1)$. Let the poverty evaluation function, $\pi(y; z)$, be continuous at $r_i(1 - \lambda_i)$ and $r_i(\lambda_i)$. Then

$$\lim_{n, m \rightarrow \infty} \mathbb{P}([D_l, D_u] \subset CI_\gamma^{[D_l, D_u]}) \geq \gamma \quad (1.28)$$

PROOF: See Appendix.

1.8 Application: Evaluation of an Anti-Poverty Program with Missing Treatments

1.8.1 Progresa

In 1997, the Mexican government introduced the Programa de Educacion, Salud y Alimentacion (the Education, Health, and Nutrition Program), better known as Progresa, and recently renamed Oportunidades, as an important element of its more general strategy to eradicate poverty in Mexico. The program is characterized by a multiplicity of objectives such as improving the educational, health and nutritional status of poor families.

Progresa provides cash transfers, in-kind health benefits and nutritional supplements to beneficiary families. Moreover, the delivery of the cash transfers is exclusively through the mothers, and is linked to children's enrollment and school attendance. This conditionality works as follows: in localities where Progresa operates, those households classified as poor with children enrolled in grades 3 to 9, are eligible to receive the grant every two months. The average bi-monthly payment to a beneficiary family amounts to 20 percent of the value of bi-monthly consumption expenditures prior to the beginning of the program. Moreover, these grants are estimated taking into account the opportunity cost of sending children to school, given the characteristics of the labor market, household production, and

gender differences. By the end of 2002, nearly 4.24 million families (around 20 percent of all Mexican households) were incorporated into the program. These households constitute around 77 percent of those households considered to be in extreme poverty.

Because of logistical and financial constraints, the program was introduced in several phases. The sequentiality of the program was capitalized by randomly selecting 506 localities in the states of Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi and Veracruz. Of the 506 localities, 320 localities were assigned to the treatment group and the rest were assigned to the control group. In total 24,077 households were selected to participate in the evaluation sample. The first evaluation survey took place in March 1998, 2 months before the distribution of benefits started. 3 rounds of surveys took place afterwards: October/November 1998, June 1999 and November 1999. The localities that served as control group started receiving benefits by December 2000. However, as noticed by Buddelmeyer and Skoufias (2004), in the treatment localities 27% of the eligible population had not received any benefits by March 2000 due to some administrative error.

1.8.2 Poverty Treatment Effects

Let us introduce some basic notation that will be helpful for the rest of the section. There are two potential states of the world, (y_1, y_0) , for each individual, where y_1 and y_0 are the outcomes that an individual would obtain if she were and she were not, respectively, a beneficiary of PROGRESA. Lets denote observed outcome by y and program participation by the indicator variable d , where $d = 1$ if the individual participates in the program, and $d = 0$ otherwise. The policy-maker observes (y, d) , but he cannot observe both states (y_1, y_0) . Formally, the

policymaker observes the random variable $y = dy_1 + (1 - d)y_0$.

We are interested in the poverty treatment effect (PTE) on the treated. This effect is given by:

$$\Delta = \Pi(F(y_1 | d = 1); z) - \Pi(F(y_0 | d = 1); z)$$

Where $F(y_1 | d = 1)$ is the distribution of the outcome of interest for the treated group, and $F(y_0 | d = 1)$ is its counterfactual. Randomization guarantees the identification of PTE since we have $F(y_0 | d = 1) = F(y_0 | d = 0)$.

As it was mentioned above, in the case of PROGRESA we have a problem of measurement error for the treatment group since a *proportion* of the households selected as beneficiaries had not received the cash transfer by the year 2000. Applying the model of section 3 to the current setting, let each individual in the treatment group be characterized by the tuple (y_{11}, y_{10}) , where y_{11} and y_{10} are the outcomes that an individual randomized in the treatment group would obtain if she were and she were not, respectively, participating in PROGRESA. Instead of observing y_{11} , one observes a contaminated variable y_1 defined by

$$y_1 \equiv wy_{11} + (1 - w)y_{10} \tag{1.29}$$

From section 3 we know that $F(y_{11}) = F(y_1 | d = 1)$ cannot be point identified if $E(w) < 1$. However, it can be partially identified if we possess some information on the marginal probability of data errors $p = P(w = 0)$, in particular if there exists a non trivial upper bound on this probability.

If one assume that w is independent of y_{11} , which is equivalent to say that data from the treatment group is contaminated, then we can apply the results obtained in section 4 to find the identification region for the PTE:

$$\mathbf{H}[\Delta] = [\Pi_\lambda^l(F(y_{11}); z) - \Pi(F(y_0 | d = 1); z), \Pi_\lambda^u(F(y_{11}); z) - \Pi(F(y_0 | d = 1); z)]$$

where $\lambda < 1$ is as an upperbound on the probability of not receiving treatment when the unit of analysis has been randomized in the treatment group.

Under the assumptions of proposition 7, the following confidence interval, $CI_\gamma^{[\Delta_l, \Delta_u]}$, asymptotically covers the PTE with at least probability γ :

$$\left[\hat{\Pi}_l(F_{11}) - \hat{\Pi}(F_0) - z_{\frac{\gamma+1}{2}} \sqrt{\frac{\hat{\sigma}_{IF}^2}{n} + \frac{\hat{\sigma}_{F_0}^2}{m}}, \hat{\Pi}_u(F_{11}) - \hat{\Pi}(F_0) + z_{\frac{\gamma+1}{2}} \sqrt{\frac{\hat{\sigma}_{F_{11}}^2}{n} + \frac{\hat{\sigma}_{F_0}^2}{m}} \right]$$

Table 1.1 presents an application of the present analysis to the PROGRESA data set. Column 1 introduces a parameter measuring the severity of poverty for the FGT poverty measure described below. We use consumption as welfare indicator, and the poverty line z is set equal to the median consumption for the control group. We use an upper bound on the proportion of errors of 0.27, the proportion of households who had not received benefits from Progresa by 2000. Columns 2 and 3 presents treatment effects on poverty and 95% confidence intervals for this parameter, respectively, without taking into consideration the contamination problem, that is to say, assuming that the parameter is point identified. Finally, columns 4,5, and 6 introduce, respectively, upper and lower bounds on the PTE, and Bonferroni confidence intervals for the identification region.

Table 1.1: Identification regions and confidence intervals for treatments effects on poverty: PROGRESA 1999

	Δ	$CI_{0.95}^\Delta$	Δ_l	Δ_u	$CI_{0.95}^{[\Delta_l, \Delta_u]}$
$\alpha = 0$	-0.068	[-.083,-.053]	-0.278	0.092	[-0.296,0.105]
$\alpha = 1$	-0.039	[-.045,-.033]	-0.148	0.009	[-0.153,0.015]
$\alpha = 2$	-0.021	[-.025,-.017]	-0.076	0.000	[-0.079,0.005]

1.8.3 Monotone Treatment Response, Data Contamination, and Missing Treatments

Monotonicity assumptions have been applied in other places to exploit their identifying power. Manski (1997) investigates what may be learned about treatment response under the assumptions of monotone, semi-monotone, and concave-monotone response functions. He shows that these assumptions have identifying power, particularly when compared to the situation where no prior information exists. In a missing treatments environment, Molinari (2005b) shows that one can extract information from the observations for which treatment data are missing using monotonicity assumptions.

Given the design of PROGRESA, one should expect that the outcome of interest (in our case consumption per capita) increases with program participation. More formally, we should expect that $y_{11} \geq y_{10}$. We have the following result

Proposition 1.8 *Suppose that $y_{11} \geq y_{10}$. Let it be known that $p \leq \lambda < 1$. Then sharp bounds for $\Pi(P_{11}; z)$ and $\Pi(P_1; z)$ are given by the identification region*

$$[\Pi(U_\lambda; z), \Pi(Q(y); z)]$$

PROOF: See Appendix.

Table 1.2 introduces the effect of the monotonicity assumption on the identification region for PTE. Clearly, considering the monotonic effect of Progresa on the treated population improves the inferential analysis of PTE by considerably shrinking the identification region.

Table 1.2: Identification regions under monotonicity assumptions: PROGRESA 1999

	Δ_l	Δ_u	$CI_{0.95}^{[\Delta_l, \Delta_u]}$
$\alpha = 0$	-0.278	-.068	[-0.296, -.053]
$\alpha = 1$	-0.148	-0.039	[-0.153, -.033]
$\alpha = 2$	-0.076	-0.021	[-0.079, -.017]

1.9 Application: Measurement of Rural Poverty in Mexico

The methodology developed in this paper is applied to the data obtained from the 2002 *Encuesta Nacional de Ingreso y Gasto de los Hogares* (ENIGH) held by INEGI (2002). This household income and expenditure survey is one of a series of surveys that are carried out under the same days of each year using identical sampling techniques.

The households are divided into zones of high and low population density. Low density population zones are those areas with fewer than 2500 inhabitants. It is common to identify these areas as rural ones. The rest of the zones (those with more than 2500 inhabitants) are identified as urban areas. The sample is representative for both urban and rural areas and at the national level. For the purposes of this study, we will just concentrate on the rural sub-sample which includes 6753 observations.

We have considered the extreme poverty line for rural areas constructed by INEGI-CEPAL for the 1992 ENIGH, following the methodology applied by the Ministry of Social Development in Mexico (2002) to inflate both the poverty line and all of the data into August 2000 prices. The rural poverty line is equal to 494.77 monthly 2002 pesos. In this paper we have used per capita current disposable income as indicator of economic welfare.¹⁰ It is divided into monetary

¹⁰Due to lack of information, a final transformation of the original data was

and non-monetary income. The monetary sources include wages and salaries, entrepreneurial rents, incomes from cooperatives, transfers, and other monetary sources. Non-monetary incomes include gifts, autoconsumption, imputed rents, and payments in kind.

The identification regions and the three different 95% confidence intervals for the class of FGT poverty measures are presented for both the contamination and the corruption models in Tables 1.3 and 1.4, respectively. We have no estimate of the frequency of data errors in the sample, so we present a sensitivity analysis using different values of λ . The first confidence interval corresponds to the point identified case ($\lambda = 0$). It is based on the point estimator ± 1.96 times its standard error. The second confidence interval is equal to the estimator of the lower bound minus 1.96, and the estimator of the upper bound plus 1.96 times their standard errors. The third confidence interval is the adjusted interval for the parameter \bar{C}_N . We found that there is almost no difference between the last two types of confidence intervals, that is to say, between the confidence interval covering the entire identification region and the one that provides the appropriate coverage for the poverty measure.

1.10 Conclusions

This paper has introduced the problems of data contamination and data corruption into the context of poverty measurement. When a proportion of the data is measured with error, a poverty measure cannot be point identified. However, we have shown that for the class of additively separable poverty measures it is required: we will assume that each household member obtains the same proportion of total income as the others.

Table 1.3: Identification regions and confidence intervals for FGT poverty measures under contamination model: Rural Mexico, 2002

λ	$\Pi_{\alpha\lambda}^L$	$\Pi_{\alpha\lambda}^U$	$CI_{0.95}^{\Pi}$	$CI_{0.95}^{[\Pi_L, \Pi_U]}$	$\overline{CI}_{0.95}^{\Pi}$
$\alpha = 0$					
0.00	0.287	0.287	[0.276, 0.298]		
0.01	0.282	0.289		[0.271, 0.300]	[0.272, 0.299]
0.02	0.275	0.292		[0.265, 0.304]	[0.266, 0.302]
0.03	0.268	0.294		[0.257, 0.306]	[0.259, 0.304]
0.05	0.252	0.299		[0.241, 0.311]	[0.243, 0.309]
0.07	0.234	0.304		[0.223, 0.316]	[0.225, 0.314]
0.10	0.209	0.312		[0.198, 0.325]	[0.200, 0.323]
$\alpha = 1$					
0.00	0.093	0.093	[0.089, 0.098]		
0.01	0.088	0.094		[0.084, 0.099]	[0.085, 0.098]
0.02	0.083	0.095		[0.079, 0.100]	[0.080, 0.099]
0.03	0.077	0.096		[0.074, 0.101]	[0.074, 0.100]
0.05	0.066	0.097		[0.062, 0.103]	[0.063, 0.102]
0.07	0.055	0.099		[0.052, 0.106]	[0.053, 0.105]
0.10	0.042	0.101		[0.039, 0.109]	[0.040, 0.108]
$\alpha = 2$					
0.00	0.042	0.042	[0.040, 0.045]		
0.01	0.038	0.043		[0.036, 0.046]	[0.036, 0.045]
0.02	0.034	0.043		[0.032, 0.047]	[0.033, 0.046]
0.03	0.031	0.043		[0.029, 0.048]	[0.029, 0.047]
0.05	0.024	0.044		[0.022, 0.049]	[0.022, 0.048]
0.07	0.018	0.045		[0.016, 0.050]	[0.017, 0.050]
0.10	0.011	0.046		[0.010, 0.053]	[0.011, 0.052]

possible to find identification regions under very mild assumptions. In particular, if there is an upper bound on the proportion of errors, we can obtain identification regions that take the form of closed intervals.

We consider the problem of statistical inference when a poverty measure is not point identified. Two type of confidence intervals are applied in the present study. For the first type, we have developed Bonferroni's confidence intervals that cover the entire identification region with some fixed probability. The second type applies and extends the results of Imbens and Manski (2004) by covering the true

Table 1.4: Identification regions and confidence intervals for FGT poverty measures under corruption model: Rural Mexico, 2002

λ	$\Pi_{\alpha\lambda}^L$	$\Pi_{\alpha\lambda}^U$	$CI_{0.95}^{\Pi}$	$CI_{0.95}^{[\Pi_L, \Pi_U]}$	$\overline{CI}_{0.95}^{\Pi}$
$\alpha = 0$					
0.00	0.287	0.287	[0.276, 0.298]		
0.01	0.279	0.296		[0.268, 0.307]	[0.270, 0.306]
0.02	0.270	0.307		[0.259, 0.318]	[0.261, 0.316]
0.03	0.260	0.316		[0.250, 0.327]	[0.251, 0.325]
0.05	0.239	0.334		[0.229, 0.345]	[0.231, 0.344]
0.07	0.218	0.352		[0.208, 0.364]	[0.209, 0.362]
0.10	0.188	0.381		[0.179, 0.393]	[0.180, 0.391]
$\alpha = 1$					
0.00	0.093	0.093	[0.089, 0.098]		
0.01	0.087	0.103		[0.083, 0.108]	[0.084, 0.107]
0.02	0.081	0.113		[0.077, 0.118]	[0.078, 0.117]
0.03	0.075	0.123		[0.071, 0.128]	[0.072, 0.127]
0.05	0.063	0.142		[0.059, 0.148]	[0.060, 0.147]
0.07	0.051	0.162		[0.048, 0.168]	[0.049, 0.167]
0.10	0.038	0.191		[0.035, 0.198]	[0.036, 0.197]
$\alpha = 2$					
0.00	0.042	0.042	[0.040, 0.045]		
0.01	0.038	0.052		[0.036, 0.055]	[0.036, 0.055]
0.02	0.034	0.062		[0.032, 0.066]	[0.032, 0.065]
0.03	0.030	0.072		[0.028, 0.076]	[0.028, 0.075]
0.05	0.022	0.092		[0.021, 0.097]	[0.021, 0.096]
0.07	0.016	0.112		[0.015, 0.117]	[0.015, 0.116]
0.10	0.010	0.141		[0.009, 0.147]	[0.010, 0.146]

value of a poverty measure with at least some fixed probability. We also consider the problem of poverty comparisons, extending the methodology developed in the first part of the paper to a setting where two populations are compared in terms of poverty.

The results obtained in the paper are illustrated by means of two applications. The first application analyzes the effect of contaminated data on poverty treatment effects for an anti-poverty program in Mexico. The second application is a sensitivity analysis for the measurement of rural poverty in Mexico under different degrees of data contamination and data corruption. The two empirical applications show the importance of considering these types of data errors, when it is pertinent, to get a more accurate measurement of the phenomenon of poverty.

1.11 Appendix: Proofs

Proof of Proposition 1.1: We need to show that $\Pi(U_\lambda; z) \leq \Pi(P; z)$ and $\Pi(L_\lambda; z) \geq \Pi(P; z)$ for all $P \in P_\lambda$. Set $\psi(y; z) = -\pi(y; z)$, so $\psi(y; z)$ is a non-decreasing function. By lemma 1.1, it suffices to prove that U_λ stochastically dominates every member of P_λ and L_λ is stochastically dominated by every member of that set. The rest of the proof is identical to proposition 4 in Horowitz and Manski (1995) \square

Proof of Proposition 1.2: Take any probability distribution Q in \mathcal{P} . Clearly, the identification breakdown point for the head-count ratio is given by

$$\lambda^H = \min\{H_Q, 1 - H_Q\}$$

Since $\pi_j(y, z) = 0$ for all $y \geq z$ and $j \in \mathcal{D}$, we have that $\lambda_j^\psi = H_Q$ for all poverty measures in \mathcal{D} . Next, I claim that $\lambda_j^\phi \geq 1 - H_Q$. Assume, towards a contradiction, that $\lambda_j^\phi < 1 - H_Q$. Define $\lambda^* = \frac{\lambda_j^\phi + 1 - H_Q}{2}$ and let $\delta(c_j)$ be the Dirac measure at c_j . Clearly, we have

$$\Pi_j(L_{\lambda_j^\phi}; z) \leq \Pi_j(L_{\lambda^*}; z) \leq \Pi_j(L_{1-H_Q}; z) \leq \Pi_j(\delta(c_j); z) = c_j$$

A contradiction. Hence, $\{\lambda_j^\phi, \lambda_j^\psi\} \geq \{H_Q, 1 - H_Q\}$ for all $j \in \mathcal{D}$, and the result follows. \square

Lemma 1.2 *Let P_1 and P_2 be two probability measures on $(\mathbb{R}, \mathcal{B})$, with \mathcal{B} the Borel sets of \mathbb{R} . Define the sets $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$, where $\mathbb{R} = \mathbf{A}_1 \cup \mathbf{A}_2 \cup \mathbf{A}_3$, $\sup \mathbf{A}_2 \leq \inf \mathbf{A}_3$, $\mathbf{A}_1 \cap \mathbf{A}_i = \emptyset$, $i = 2, 3$, and*

$$\Lambda = \{(P_1, P_2) : P_1(A) = P_2(A), \forall A \in \mathcal{B} \cap \mathbf{A}_1; P_1(\mathbf{A}_3) = P_2(\mathbf{A}_2) = 0\}$$

Let $\delta(x) : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function and suppose there exists some $z \in \mathbf{A}_1 \cup \mathbf{A}_2$ with $\delta(x) = 0$ for all $x \geq z$, and $\delta(x) \geq 0$, otherwise. Then:

i) F_2 first order stochastically dominates F_1 , where F_2 and F_1 are the distribution functions implied by probability measures P_2 and P_1 , respectively.

ii) $E_{P_2}(\delta(x)) \leq E_{P_1}(\delta(x))$, $\forall (P_1, P_2) \in \Lambda$

PROOF:

i) Straightforward

ii) Because $\delta(x) = 0$ for all $x \in \mathbf{A}_3$, we have:

$$E_{P_1}(\delta(x)) = \int_{\mathbf{A}_1} \delta(x) dP_1 + \int_{\mathbf{A}_2} \delta(x) dP_1$$

$$E_{P_2}(\delta(x)) = \int_{\mathbf{A}_1} \delta(x) dP_2 + \int_{\mathbf{A}_2} \delta(x) dP_2$$

Therefore, $E_{P_2}(\delta(x)) \leq E_{P_1}(\delta(x))$ iff

$$\int_{\mathbf{A}_1} \delta(x) dP_2 + \int_{\mathbf{A}_2} \delta(x) dP_2 \leq \int_{\mathbf{A}_1} \delta(x) dP_1 + \int_{\mathbf{A}_2} \delta(x) dP_1$$

\Leftrightarrow

$$\int_{\mathbf{A}_2} \delta(x) dP_2 \leq \int_{\mathbf{A}_2} \delta(x) dP_1$$

Since $P_2(\mathbf{A}_2) = 0$ and $\delta(x) \geq 0$ the result follows. \square

Definition 1.4 A class \mathcal{G} of subsets of Ω is called a λ -system if

i) $\Omega \in \mathcal{G}$

ii) If $G_1, G_2 \in \mathcal{G}$ and $G_1 \supseteq G_2$ then $D_1 \setminus D_2 \in \mathcal{G}$.

iii) If $\{G_n\}$ is an increasing sequence of sets in \mathcal{G} , the $\bigcup_{n=1}^{\infty} G_n \in \mathcal{G}$

Lemma 1.3 (Sierpinski 1928) If \mathcal{F} is stable under finite intersections, and if \mathcal{G} is a λ -system with $\mathcal{G} \supseteq \mathcal{F}$, then $\mathcal{G} \supseteq \sigma(\mathcal{F})$

Proof of Proposition 1.3: Define $\delta(y) = \pi_2(y; z) - \pi_1(y; z)$. We have

$$\begin{aligned} m_2 &= \int \pi_2(y; z) dL_\lambda - \int \pi_2(y; z) dU_\lambda \\ &= \int \pi_1(y; z) dL_\lambda + \int \delta(y) dL_\lambda - \int \pi_1(y; z) dU_\lambda - \int \delta(y) dU_\lambda \\ &= m_1 + \int \delta(y) dL_\lambda - \int \delta(y) dU_\lambda \end{aligned}$$

By construction, there exist sets $\mathbf{A}_1, \mathbf{A}_2$ and \mathbf{A}_3 in \mathbb{R} such that $\mathbb{R} = A_1 \cup A_2 \cup A_3$, $\sup \mathbf{A}_2 \leq \inf \mathbf{A}_3$, and $P_{L_\lambda}(\mathbf{A}_3) = P_{U_\lambda}(\mathbf{A}_2) = 0$. Moreover, $\delta(y) \geq 0$ for all y . By lemma 1.2, it suffices to show that $(P_{L_\lambda}, P_{U_\lambda}) \in \Lambda$. We have four cases: $\mathbf{A}_1 = [\min\{r(\lambda), r(1-\lambda)\}, \max\{r(\lambda), r(1-\lambda)\}]$, $\mathbf{A}_1 = (\min\{r(\lambda), r(1-\lambda)\}, \max\{r(\lambda), r(1-\lambda)\}]$, $\mathbf{A}_1 = [\min\{r(\lambda), r(1-\lambda)\}, \max\{r(\lambda), r(1-\lambda)\})$, and $\mathbf{A}_1 = (\min\{r(\lambda), r(1-\lambda)\}, \max\{r(\lambda), r(1-\lambda)\})$. I will analyze just first case. A similar argument works for the other three. Define the sets $\mathbf{A}_2 = (-\infty, \min\{r(\lambda), r(1-\lambda)\})$ $\mathbf{A}_3 = (\max\{r(\lambda), r(1-\lambda)\}, \infty)$, and $A_1 = [\min\{r(\lambda), r(1-\lambda)\}, \max\{r(\lambda), r(1-\lambda)\}]$. By inspection, we have $P_{L_\lambda}(A_3) = P_{U_\lambda}(A_2) = 0$. Let $\mathcal{B}(\mathbf{A}_1)$ be the Borel sigma-field on \mathbf{A}_1 . I will show that $P_{L_\lambda}(A) = P_{U_\lambda}(A)$ for all $A \in \mathcal{B}(\mathbf{A}_1)$ by applying a generating class argument. Write \mathcal{E} for the class of all intervals $(\min\{r(\lambda), r(1-\lambda)\}, t]$, with $t \in \mathbf{A}_1$. The following series of claims proves this result:

Claim 1: $\sigma(\mathcal{E}) = \mathcal{B}(\mathbf{A}_1)$

Let \mathcal{O} stand for the class of all open subsets of \mathbf{A}_1 , so $\mathcal{B}(\mathbf{A}_1) = \sigma(\mathcal{O})$. Each interval $(\min\{r(\lambda), r(1-\lambda)\}, t]$ in \mathcal{E} has a representation $\bigcap_{n=1}^{\infty} (\min\{r(\lambda), r(1-\lambda)\}, t + \frac{1}{n})$. $\sigma(\mathcal{O})$ contains all open intervals, and it is stable under countable intersections. Hence, $\mathcal{E} \subset \mathcal{B}(\mathbf{A}_1)$. On the other hand, each open interval (a, t) on \mathbf{A}_1 has a representation $(a, t) = \bigcup_{n=1}^{\infty} (\min\{r(\lambda), r(1-\lambda)\}, t - \frac{1}{n}] \cap (\min\{r(\lambda), r(1-\lambda)\}, a]^c$, so $\mathcal{O} \subset \sigma(\mathcal{E})$ and thus $\sigma(\mathcal{E}) = \mathcal{B}(\mathbf{A}_1)$.

Claim 2: $\mathcal{D} = \{A \in \mathcal{B}(\mathbf{A}_1) : P_{U_\lambda}(A) = P_{L_\lambda}(A)\}$ is a λ -system

i) $\mathbf{A}_1 \in \mathcal{D}$ follows from the fact that $P_{U_\lambda}(\mathbf{A}_1) = P_{L_\lambda}(\mathbf{A}_1)$. ii) Let $A_1, A_2 \in \mathcal{D}$.

By the properties of a probability measure, $P_i(A_1 \cap A_2^c) = P_i(A_1) + P_i(A_2^c) - P_i(A_1 \cup A_2)$, $i = 1, 2$. $P_1(A_1 \cap A_2^c) = P_2(A_1 \cap A_2^c)$ follows after some algebraic manipulations. Finally, we need to show that \mathcal{D} is closed under increasing limits.

Let $\{A_n\}$ be an increasing sequence of sets in \mathcal{D} and $A = \bigcup_{n=1}^{\infty} A_n$. Define a sequence of indicator functions $\{\mathbf{1}_{A_n}\}$. Clearly, this is a positive and increasing sequence of functions. By the Monotone Convergence Theorem

$$\begin{aligned} \lim_{n \rightarrow \infty} E_{P_{U_\lambda}}(\mathbf{1}_{A_n}) &= E_{P_{U_\lambda}}(\mathbf{1}_A) \\ &= E_{P_{L_\lambda}}(\mathbf{1}_A) \\ &= \lim_{n \rightarrow \infty} E_{P_{U_\lambda}}(\mathbf{1}_{A_n}) \end{aligned}$$

hence $P_{U_\lambda}(A) = P_{L_\lambda}(A)$.

Claim 3: $\mathcal{D} \supseteq \mathcal{E}$

By inspection, $P_{L_\lambda}((\min\{r(\lambda), r(1-\lambda)\}, t]) = P_{U_\lambda}((\min\{r(\lambda), r(1-\lambda)\}, t])$ for all $t \in \mathbf{A}_1$.

Since \mathcal{E} is stable under finite intersections, by lemma 1.3 and claims 1, 2, and 3 we have $\mathcal{D} \supseteq \sigma(\mathcal{E}) = \mathcal{B}(\mathbf{A}_1)$. Hence $\mathcal{D} = \mathcal{B}(\mathbf{A}_1)$. \square

Proof of Corollary 1.1: By Proposition 1.2 it suffices to show that $\pi_1(y; z) \geq \pi_2(y; z)$ for all $y \in (0, z)$. By continuity and monotonicity of $\pi_1(y; z)$ on $[0, z]$ there exists $\lambda \in (0, 1)$ such that $\pi_1(y; z) = \lambda\pi_1(0; z) + (1-\lambda)\pi_1(z; z)$ for all $y \in (0, z)$. Therefore

$$\begin{aligned} f \circ \pi_1(y; z) &= f(\lambda\pi_1(0; z) + (1-\lambda)\pi_1(z; z)) \\ &\leq \lambda f \circ \pi_1(0; z) + (1-\lambda)f \circ \pi_1(z; z) \\ &\leq \lambda\pi_1(0; z) + (1-\lambda)\pi_1(z; z) \\ &= \pi_1(y; z) \end{aligned}$$

Where I have made use of the convexity of f . \square

Proof of Corollary 1.3: Condition iv) is equivalent to have $\pi_1 = f \circ \pi_2$ with $f' > 0$ and $f'' > 0$ (Pratt 1964). The result follows from corollary 1.1. \square

Before proving the rest of Lemmas and Propositions, we introduce a number of preliminary results. Let y_1, y_2, \dots, y_n be *i.i.d.* random variables with distribution function $F(y)$, and let $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ denote the order statistics of the sample. Consider the trimmed mean given by

$$S_n = \frac{1}{[(\beta - \alpha)n]} \sum_{i=[\alpha n]+1}^{[\beta n]} y_{(i)} \quad (1.30)$$

where $0 \leq \alpha < \beta \leq 1$ are any fixed numbers and $[\cdot]$ represents the greatest integer function. Let $r(\alpha)$ and $r(\beta)$ be continuity points of $F(y)$. Further, define

$$G(y) = \begin{cases} 0 & \text{if } y < r(\alpha) \\ \frac{F(y) - \alpha}{\beta - \alpha} & \text{if } r(\alpha) \leq y < r(\beta) \\ 1 & \text{otherwise} \end{cases}$$

and set

$$\mu = \int_{-\infty}^{\infty} y dG(y) \quad (1.31)$$

$$\sigma^2 = \int_{-\infty}^{\infty} y^2 dG(y) - \mu^2 \quad (1.32)$$

Lemma 1.4 (*Stigler 1973*) Assume $E(y^2) < \infty$, then

$$n^{\frac{1}{2}}(S_n - \mu) \xrightarrow{d} N(0, (1 - \alpha)^{-2}((1 - \alpha)\sigma^2 + (r(\alpha) - \mu)^2\alpha(1 - \alpha))) \text{ if } \beta = 1.$$

$$n^{\frac{1}{2}}(S_n - \mu) \xrightarrow{d} N(0, (\beta)^{-2}((\beta)\sigma^2 + (r(\beta) - \mu)^2\beta(1 - \beta))) \text{ if } \alpha = 0.$$

Lemma 1.5 (*de Wet 1976*) Assume $E(|y|^3) < \infty$, then

$$\sup \left| \mathbb{P} \left(\sqrt{N} \frac{(S_n - \mu)}{\sigma} < x \right) - \Phi(x) \right| \rightarrow 0 \text{ if } \beta = 1.$$

$$\sup \left| \mathbb{P} \left(\sqrt{N} \frac{(S_n - \mu)}{\sigma} < x \right) - \Phi(x) \right| \rightarrow 0 \text{ if } \alpha = 0.$$

Proof of Proposition 1.4: Define the events

$$A_n = \left\{ \Pi_l : \Pi_l \geq \hat{\Pi}_l - z_{\frac{\gamma+1}{2}} \frac{\hat{\sigma}_l}{\sqrt{n}} \right\}$$

$$B_n = \left\{ \Pi_u : \Pi_u \leq \hat{\Pi}_u + z_{\frac{\gamma+1}{2}} \frac{\hat{\sigma}_u}{\sqrt{n}} \right\}$$

From the definition of the confidence interval, $CI_\gamma^{[P_L, P_U]}$, and Bonferroni's inequality

$$\mathbb{P}([\Pi_l, \Pi_u] \subset CI_\gamma^{[\Pi_l, \Pi_u]}) = \mathbb{P}(A_n \cap B_n) \geq \mathbb{P}(A_n) + \mathbb{P}(B_n) - 1$$

By lemma 1.4 $\frac{\sqrt{n}(\hat{\Pi}_i - \Pi_i)}{\hat{\sigma}_i} \xrightarrow{d} \mathbf{N}(0, 1)$, $i = u, l$. Thus

$$\lim_{n \rightarrow \infty} Pr([\Pi_l, \Pi_u] \subset CI_\gamma^{[\Pi_l, P_u]}) \geq \gamma$$

and the result follows. \square

Proof of Proposition 1.5: The result is a direct consequence of lemma 1.5, and Lemma 4 in Imbens and Manski (2004). \square

Lemma 1.6 *If $X_n \xrightarrow{d} X = N(\mu_1, \sigma_1^2)$ and $Y_m \xrightarrow{d} Y = N(\mu_2, \sigma_2^2)$, and if X_n is independent of Y_m for all n and m , then $X_n + Y_m \xrightarrow{d} N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Proof: Let $Z_{n,m} = X_n + Y_m$. By independence of X_n and Y_m , its characteristic function can be written as

$$\varphi_{Z_{n,m}}(u_1, u_2) = \varphi_{X_n}(u_1)\varphi_{Y_m}(u_2)$$

By the Uniqueness Theorem we have

$$\begin{aligned} \lim_{n,m \rightarrow \infty} \varphi_{X_n}(u_1)\varphi_{Y_m}(u_2) &= \exp(iu_1\mu_1 - \frac{u_1^2\sigma_1^2}{2}) \exp(iu_2\mu_2 - \frac{u_2^2\sigma_2^2}{2}) \\ &= \exp\left(\sum_{j=1}^2 iu_j\mu_j - \frac{1}{2} \sum_{j=1}^2 u_j^2\sigma_j^2\right) \end{aligned}$$

This expression corresponds to the characteristic function of the random vector $Z = (X, Y)$, where Z is Gaussian. Moreover, X and Y are independent since $Cov(X, Y) = 0$. The result follows. \square

Proof of Proposition 1.7: Define the events

$$A_{n,m} = \left\{ \Pi_{lA} - \Pi_{uB} : z_{\frac{\gamma+1}{2}} \sqrt{\frac{m}{n+m} \hat{\sigma}_{lA} + \frac{n}{n+m} \hat{\sigma}_{uB}} \geq Y_{n,m} \right\}$$

$$B_{n,m} = \left\{ \Pi_{uA} - \Pi_{lB} : -z_{\frac{\gamma+1}{2}} \sqrt{\frac{m}{n+m} \hat{\sigma}_{lA} + \frac{n}{n+m} \hat{\sigma}_{uB}} \leq W_{n,m} \right\}$$

Where $Y_{n,m} = \sqrt{\frac{nm}{n+m}}(\hat{\Pi}_{lA} - \hat{\Pi}_{uB} - \Pi_{lA} + \Pi_{uB})$ and $W_{n,m} = \sqrt{\frac{nm}{n+m}}(\hat{\Pi}_{uA} - \hat{\Pi}_{lB} - \Pi_{uA} + \Pi_{lB})$. Notice that Lemma 1.4 implies

$$i) \lim_{n,m \rightarrow \infty} \sqrt{\frac{nm}{n+m}}(\hat{\Pi}_{iA} - \Pi_{iA}) = \lim_{n,m \rightarrow \infty} \sqrt{\frac{m}{n+m}} \sqrt{n}(\hat{\Pi}_{iA} - \Pi_{iA}) \xrightarrow{d} N(0, \epsilon \sigma_{iA}^2),$$

$i = l, u$

$$ii) \lim_{n,m \rightarrow \infty} \sqrt{\frac{nm}{n+m}}(\hat{\Pi}_{iB} - \Pi_{iB}) = \lim_{n,m \rightarrow \infty} \sqrt{\frac{n}{n+m}} \sqrt{m}(\hat{\Pi}_{iB} - \Pi_{iB}) \xrightarrow{d} N(0, (1 - \epsilon) \sigma_{iB}^2), \quad i = l, u$$

By applying Lemmas 1.4 and 1.7 together it is easy to show that $Y_{n,m} \xrightarrow{d} \mathbf{N}(0, \epsilon \sigma_{lA}^2 + (1 - \epsilon) \sigma_{uB}^2)$ and $W_{n,m} \xrightarrow{d} \mathbf{N}(0, \epsilon \sigma_{uA}^2 + (1 - \epsilon) \sigma_{lB}^2)$. By Bonferroni's inequality we have:

$$\mathbb{P}([D_l, D_u] \subset CI_\gamma^{[D_l, D_u]}) \geq \mathbb{P}(A_{n,m}) + \mathbb{P}(B_{n,m}) - 1$$

Hence $\lim_{n,m \rightarrow \infty} \mathbb{P}([D_l, D_u] \subset CI_\gamma^{[D_l, D_u]}) \geq \gamma \quad \square$

Proof of Proposition 1.8: From Proposition 1 in Horowitz and Manski (1995)

$$P_{11}(y_1) \in \mathcal{P}_{11}(\lambda) \equiv \mathcal{P} \cap \left\{ \frac{Q(y) - \lambda \phi_{00}}{1 - \lambda} : \phi_{00} \in \mathcal{P} \right\}$$

For all $x \in \mathbb{R}$, define the indicator functions $\mathbf{1}(y_{11} \leq x)$ and $\mathbf{1}(y_{10} \leq x)$. By the monotonicity assumption

$$\mathbf{1}(y_1 \leq x) \leq \mathbf{1}(y \leq x)$$

Taking expectations at both sides of this inequality, we have that

$$P(y_1 \leq x) \leq Q(y \leq x)$$

for all $x \in \mathbb{R}$. This imposes a restriction on the set $\mathcal{P}_{11}(\lambda)$ since all of the distributions in this set must stochastically dominate the observed distribution $Q(y)$.

Hence

$$\text{Max} \left\{ 0, \frac{Q(y \leq x) - \lambda \phi_{00}(y_0 \leq x)}{1 - \lambda} \right\} \leq Q(y \leq x)$$

for all $x \in \mathbb{R}$. After some algebraic manipulations, we obtain that $Q(y \leq x) \leq \phi_{00}(y_0 \leq x)$, which provides a restriction on the set of feasible distributions ϕ_{00} . Define the set of distribution functions stochastically dominated by $Q(y)$ by

$$\mathfrak{D} = \{ \phi_{00} \in \mathcal{P} : \phi_{00}(y_0 \leq x) \geq Q(y \leq x), \forall x \}$$

one can characterize the identification region for the distribution $F(y_{11})$ under the monotonicity assumption as follows:

$$P(y_{11}) \in \mathcal{P}_{11}^M(\lambda) \equiv \mathcal{P} \cap \left\{ \frac{Q(y) - \lambda \phi_{00}}{1 - \lambda} : \phi_{00} \in \mathfrak{D} \right\}$$

To prove the proposition, we just need to show that $Q(y) \in \mathcal{P}_{11}^M$ and that this distribution is stochastically dominated by all other distributions in \mathcal{P}_{11}^M . The first condition is trivially satisfied by defining $\phi_{00} = Q(y)$, and hence we have that $Q(y) \in \mathcal{P}_{11}^M(\lambda)$. Next, assume, towards a contradiction, that there exists some distribution in $\mathcal{P}_{11}^M(\lambda)$ that does not stochastically dominate $Q(y)$. Then, for some $x \in \mathbb{R}$ and some $\phi'_{00} \in \mathfrak{D}$, we have

$$\text{Min} \left\{ 1, \frac{Q(y \leq x) - \lambda \phi'_{00}(y_0 \leq x)}{1 - \lambda} \right\} > Q(y \leq x)$$

From where $Q(y \leq x) > \phi'_{00}(y_0 \leq x)$, or $1 > Q(y \leq x) > 1 + \lambda(1 - \phi_{00})$, a contradiction since $\phi'_{00} \in \mathfrak{D}$ and ϕ_{00} is a probability measure. \square

Chapter 2
On the Design of an Optimal Transfer
Schedule with Time Inconsistent
Preferences

2.1 Introduction

Public transfers constitute a very important policy tool in developed and developing societies alike. From anti-poverty programs to unemployment insurance benefits, they play a very important role as a welfare enhancing mechanism.

Mainstream public economics analyzes the problem of designing an optimal transfer schedule based on the assumption that individuals have an abundance of psychological resources: unboundedly rational, forward looking, and internally consistent.¹ Particularly, it is assumed that individuals are unbounded in their self-control and optimally follow whatever plans they set out for themselves. In this paper, we investigate the optimal design of a transfer schedule when individuals have self-control problems.

Economic theories of intertemporal choice generally assume that individuals discount the future exponentially. In other words, the choices made between today and tomorrow should be no different from the choices made between the days 200 and 201 from now, all else equal. However, experimental evidence suggests that many individuals have preferences that reverse as the date of decision making nears. Research on animal and human behavior has led scientists to conclude that preferences are roughly hyperbolic in shape, implying a high discount rate in the immediate future, and relatively lower rate over periods that are further away (Ainslie 1992; Lowenstein and Thaler 1989). Moreover, there exists field evidence of present-biased preferences and time inconsistent behavior (DellaVigna

¹An example of such approach applied in a dynamic setting is the article on unemployment insurance written by Shavell and Weiss (1979). They characterize the time sequence of benefits that maximizes the expected utility of the unemployed. Blackorby and Donaldson (1988) study second best allocations in a static model where government lacks full information about consumer types.

and Malmendier 2003; Fang and Silverman 2004). Angeletos et al (2001) calibrate the hyperbolic and exponential models using US data on savings and consumption, finding that the former model better matches actual consumers' behavior. They noticed that, in contrast to the exponential discounting model, hyperbolic households exhibit a high level of comovement between predictable changes in income and changes in consumption. This type of behavior has also been found in empirical studies that show how consumption is often very sensitive to an income transfer in the very short-run.² Similar results have been found in developing countries, particularly the development of commitment devices to face the time inconsistency problem (Rutheford 1999; Ashraf, Gons, Karlan, and Yin 2003; Ashraf, Gons, Karlan, and Yin 2006).

In this paper we present a very simple model that captures this phenomenon within the context of designing an optimal transfer schedule. We refer to this type of policy tool as a consumption maintenance program (CMP). The dynamic economic environment we study has two actors: a policymaker whose goal is to allocate an exogenous budget in order to maximize some welfare function, and an agent who takes consumption-savings decisions over time and is borrowing constrained. The policymaker is fully committed to his plan once it is established. In contrast, the beneficiary may be time-inconsistent and may not follow up his

²Stephens (2003) and Stephens (2002) study the consumption response to monthly paycheck receipt in the United States and the United Kingdom, respectively. Under the standard life-cycle/permanent income hypothesis, household consumption should not respond to paycheck arrival. Nevertheless, he finds an excessive response to paycheck receipt. In the case of the US, he shows how the sensitivity is higher for households for which Social Security represents an important proportion of their total income. In a similar study and using data on the consumption patterns of food stamp recipients in the US, Shapiro (2005) presents evidence of declining caloric intake over the 30-day period following the receipt of food stamps.

original consumption plan in the future.

Following a tradition in public economics, we begin the analysis with a first-best approach. We show that if program beneficiaries are time inconsistent and receive all the benefits in just one payment, then the equilibrium consumption allocation is *always* inefficient. In other words, it could be possible, in principle, to strictly increase the beneficiary's welfare at some point in time, without decreasing his welfare in other time periods. On the other hand, if the policymaker has total flexibility in the way he can allocate the public budget over time and can impose negative lump-sum transfers, any efficient consumption allocation can be obtained in equilibrium. Intuitively, the CMP is used as a commitment mechanism by the policymaker in order to impose time consistency for some previously chosen efficient consumption plan. We also characterize the set of feasible consumption allocations when lump-sum transfers are non-negative and the beneficiary has access to an exogenous and deterministic income flow. Therefore, we can find an analogy between the CMP and Laibson's golden eggs model (Laibson 1997a), where the commitment technology takes the form of an illiquid asset.

In a more realistic scenario, the assumption that the beneficiary's relevant information is public seems to be too strong. Income often cannot be observed by the policymaker, especially in developing countries where the informal sector is pervasive. Moreover, fully committing to some transfer schedule is not the best policy ex-ante in an uncertain environment. Therefore, not only the policymaker should consider his role as a commitment "enforcer", but also as an insurer that helps beneficiaries face the potential risk of receiving a negative income shock. We introduce this concern into our model by assuming that while the policymaker can observe the distribution of income shocks, he cannot observe their actual re-

alizations. We approach this problem from a mechanism design perspective. The solution we found represents the existent tradeoff between a more committed versus a more flexible transfer schedule.

The plan of this paper is as follows. Section 2 introduces a simple dynamic model with quasi-hyperbolic discounting into the problem of designing a transfer schedule. Section 3 studies the problem from a first-best perspective, assuming the policymaker has full information and lump-sum transfers are feasible. Section 4 characterizes the optimal transfer schedule when the policy maker only knows the distribution of income shocks. Section 5 concludes. Most of the mathematical details are in the Appendix.

2.2 The Model

Consider the following economy. Time is discrete and indexed by $t = 1, 2, \dots, T$. There is one agent who lives for $T \geq 3$ periods and one policymaker or planner. There is one consumption good x . The instantaneous utility function $u : \mathbb{R}_+ \rightarrow \mathbb{R}$ of the agent is assumed to satisfy the following conditions: $u(\cdot)$ is C^2 over $(0, \infty)$, $u'(x) > 0$, and $u''(x) < 0$.

In period t , preferences over consumption streams $x = (x_1, \dots, x_T) \in \mathbb{R}_+^T$ are representable by the utility function

$$U_t(x) = u(x_t) + \beta \sum_{\tau=t+1}^T \delta^{\tau-t} u(x_\tau)$$

where $(\beta, \delta) \in (0, 1] \times (0, 1]$. There exists a linear storage technology with gross return $R > 0$. The agent is liquidity constrained in the sense that he can save but not borrow.

The type of preferences represented by this model incorporates the so-called

quasi-geometric discounting³. The parameter δ is called the *standard discount factor* and it represents the long-run, time consistent discounting; the parameter β represents a preference for immediate gratification and is known as the *present-biased factor*. For $\beta = 1$ these preferences reduce to exponential discounting. For $\beta < 1$, the (β, δ) formulation implies discount rates that decline as the discounted event is moved further away in time.⁴

In the present analysis, we assume that the agent is sophisticated in the sense that she is fully aware of her time inconsistency problem. When preferences are dynamically inconsistent, it is standard practice to formally model the agent as a sequence of temporal selves making choices in a dynamic game. Similar to Strotz (1956), Peleg and Yaari (1973), Goldman (1980), Laibson (1997b), Laibson (1998), O'Donoghue and Rabin (2001), and O'Donoghue and Rabin (1993), we model this problem by thinking of the agent as consisting of T autonomous selves whose intertemporal utility functions are given by

$$\begin{aligned}
 U_1 &= u(x_1) + \beta\delta u(x_2) + \beta\delta^2 u(x_3) + \dots + \beta\delta^{T-2} u(x_{T-1}) + \beta\delta^{T-1} u(x_T) \\
 U_2 &= u(x_2) + \beta\delta u(x_3) + \beta\delta^2 u(x_4) + \dots + \beta\delta^{T-3} u(x_{T-1}) + \beta\delta^{T-2} u(x_T) \\
 U_3 &= u(x_3) + \beta\delta u(x_4) + \beta\delta^2 u(x_5) + \dots + \beta\delta^{T-4} u(x_{T-1}) + \beta\delta^{T-3} u(x_T) \\
 &\vdots = \vdots \\
 U_t &= u(x_t) + \beta\delta u(x_{t+1}) + \dots + \beta\delta^{T-t} u(x_T) \\
 &\vdots = \vdots \\
 U_T &= u(x_T)
 \end{aligned}$$

The government implements a consumption maintenance programme (CMP here-

³This type of preferences was originally proposed by Phelps and Pollack (1968).

⁴See Frederick, Loewenstein, and O'Donoghue (2002), for review of the (β, δ) formulation and its relation to hyperbolic discounting.

after), which consists of allocating an exogenous budget $B > 0$ to the individual through a transfer schedule $\{\tau_t\}_{t=1}^T$. The government allocates this budget over time in order to maximize the "long-run" welfare of the agent represented by the function⁵

$$W(x_1, \dots, x_t) = \sum_{t=1}^T \delta^{t-1} u(x_t) \quad (2.1)$$

Intuitively, this welfare function represents the policymaker's preference for smoother consumption paths.⁶ Following a tradition in the income maintenance program literature, we set aside the revenue-raising implications to finance this budget. We have in mind a world in which the budget B is financed by the non-target population or by some other exogenous source of funding. For the purposes of the present study, we abstract from the process of identifying the target population, focusing exclusively on the allocation of benefits.

⁵Three main approaches to evaluate welfare when preferences are time inconsistent can be found in the literature. The first approach, extensively applied in the consumption-savings literature by Goldman (1979) and Phelps and Pollak (1968), emphasizes the application of a Pareto criterion to evaluate equilibrium allocations. O'Donoghue and Rabin (1999) advocate maximizing welfare from a "long-run perspective". It involves the existence of a "... (fictitious) period 0 where the person has no decision to make and weights all future periods equally." This approach incorporates the fact that most models of present-biased preferences try to capture situations in which people pursue immediate gratification. Moreover, they consider the Pareto criterion as "too strong" because it often refuses strategies that are preferred by almost all incarnations of the agent. In that sense, ranking strategies becomes complicated since "...the Pareto criterion often refuses to rank two strategies even when one is much preferred by virtually all period selves, while the other is preferred by only one period self." Finally, there is a third approach that privileges a subset $\mathcal{C} \in 2^T$ of players. For instance, welfare may be evaluated with respect to current self's perspective. This "dictatorship of the present" approach has been applied by Cropper and Laibson (1998), and Cropper and Koszegi (2001), where the goal of the policy maker at time t is to maximize the welfare of self- t .

⁶This type of analysis, where the policymaker has an objective function that is different from that of the agent, is not new in public economics. As noticed by Kanbur, Pirttila, and Tuomala (2004) "...there is a long tradition of non-welfarist welfare economics...where the outcomes of individual behavior are evaluated using a preference function different from the one that generate the outcomes."

2.3 First-Best Consumption Maintenance Programs

In this section, we establish a benchmark case by characterizing the optimal CMP when the beneficiary's income flow $\{y_t\}_{t=1}^T$ can be observed by and lump-sum transfers are feasible to the policymaker. Formally, the set of feasible transfer schedules is given by

$$\mathcal{B}^F = \{(\tau_1, \dots, \tau_T) \in \mathbb{R}^T : \sum_{t=1}^T R^{1-t} \tau_t = B\}$$

In contrast to a time-inconsistent beneficiary, we assume that once the policymaker decides which transfer schedule will be implemented, he is fully committed to that program. We can formally model this problem as a two-stage game where the players are the policymaker and the T different incarnations of the agent. In stage 1, the policymaker announces the transfer schedule to be implemented. In stage 2, the different incarnations of the agent play a consumption-savings game.

Before proceeding with the analysis, we introduce some useful concepts and definitions as well as the equilibrium concept we will employ in this section. Let ω_t be cash on hand. This variable evolves according to

$$\omega_{t+1} = R(\omega_t - x_t) + y_{t+1} + \tau_{t+1}$$

with $\omega_1 = \tau_1 + y_1$. In the present study, we will focus on consumption strategies in which the past influences current play only through its effect on cash on hand, so the equilibrium concept for the consumption-savings game is that of *Markov perfect equilibrium*. A feasible consumption strategy for player $t \in \{1, \dots, T\}$ is given by the function $s_t : \omega_t \rightarrow [0, \omega_t]$.⁷ We say that an equilibrium allocation $x^*(\tau)$ is *induced* by a transfer schedule $\tau \in \mathcal{B}^F$ if it is supported by some Markov

⁷More formally, we could denote by S_t the set of all feasible strategies for player $t \in \{0, 1, \dots, T\}$ and by $S_1 \times S_2 \dots \times S_T$ the joint strategy space of all players.

perfect equilibrium of the consumption-savings game. A first-best CMP is derived from the solution to the two-stage game described above:

Definition 2.1 $\tau^* = (\tau_1^*, \dots, \tau_T^*) \in \mathcal{B}^F$ is a first-best CMP if $W(x(\tau^*)) \geq W(x(\tau))$ for every $\tau \in \mathcal{B}^F$, where $x(\tau^*) \in \mathbb{R}_+^T$ and $x(\tau) \in \mathbb{R}_+^T$ are equilibrium allocations induced, respectively, by the transfer schedules τ^* and τ .

2.3.1 Transfer Schedules without Commitment: The One-Payment CMP

In this section, we study the behavioral implications and welfare outcomes of *one-payment* CMP. We assume that the policy maker is constrained to transfer all of the resources in period 1, where by all resources we mean the public budget B plus the present value of the beneficiary's future income flow.⁸ In some circumstances, this is equivalent to giving access to capital markets to the beneficiary, so he would be able to borrow money against his future income stream. Besides being a benchmark case for comparisons, this seems to be the natural setup for the analysis: administrative costs, technological constraints, and other types of impediments may prevent the policymaker from distributing the budget with more flexibility.

One implication of assuming that the beneficiary is time consistent ($\beta = 1$) is that the optimal consumption path from self 1's perspective can be implemented in equilibrium: his future incarnations will consume and save the amounts he wants them to. Moreover, because the beneficiary and the policy maker share the same intertemporal preferences, an optimal CMP is to transfer the total budget

⁸This implicitly implies that negative transfers can be implemented. We will weaken this assumption later on.

in period 1. On the other hand, if the individual is time inconsistent ($\beta < 1$), this may not be an efficient policy because, as we will see below, it could be possible for the policymaker to weakly improve the welfare of the beneficiary in all periods, and to strictly increase his welfare at some period. The strategic interaction of his different incarnations might generate a coordination failure with a suboptimal outcome as a result.

In the present setting, it can be shown that for all $\beta < 1$ the equilibrium allocation $x^* \in \mathbb{R}_{++}^T$ is inefficient from a long-run perspective: we can always find a period $t < T$ such that reallocating consumption from t to some $j > t$ implies a welfare improvement. In other words, by transferring consumption from period t to period j , not only could it be possible to increase the welfare of self t , but also the welfare of their past and future incarnations.⁹ Notice that if the agent were time consistent, this behavior should not be observed in equilibrium. Having time inconsistent preferences is what opens the possibility of an inefficient equilibrium.¹⁰

Proposition 2.1 establishes that, for the one-payment CMP, if the beneficiary is time inconsistent, then the equilibrium allocation is inefficient¹¹

⁹Therefore, this result also implies that the consumption allocation is not Pareto optimal.

¹⁰In the context of a consumption-savings problem, Laibson(1996) shows how damaging in terms of welfare the type of behavior implied by quasi-hyperbolic discounting could be when the agent has a constant relative risk aversion utility function. Based on his own calibration, he argues that inadequate access to optimal savings policies translates in a welfare cost of at least $\frac{9}{10}$ of one year income. He discusses the positive effects of some policies to increase not only savings but also the welfare of each of the different selves when the agent faces a time inconsistency problem.

¹¹Although there has been some progress in the characterization of equilibria with quasi-hyperbolic discounting, the analysis of the welfare properties of those equilibria has been limited to the case of constant relative risk aversion. Intuitively, it is clear that the inefficient property of the equilibrium of the game should not be a consequence of assuming CRRA preferences. Under very general conditions, Goldman (1979) shows that an interior equilibrium consumption allocation is effi-

Proposition 2.1 *In the one-payment CMP with a time-inconsistent beneficiary, the consumption allocation, $x^*(\tau) \in \mathbb{R}_{++}^T$, arising in equilibrium is inefficient.*

PROOF: See Appendix.

The intuition behind this result is very simple: if the policymaker transfers all of the resources in just one payment, a time-inconsistent beneficiary will find himself in a situation of overconsumption. In particular, it can be shown that self T-2 will always be overconsuming in the sense that it would be possible to increase his welfare by transferring resources to the future. Since preferences are separable and monotone, the equilibrium allocation is also inefficient from a long-run perspective.

Interestingly, this result may sound counterintuitive for those who consider that providing *more* liquidity to the poor is *always* the best policy. Our conjecture is that a final answer much depends on the specific goals of a CMP. For instance, if the primary goal of a CMP is to smooth consumption over time, then a one-payment CMP may not be an optimal policy if the beneficiary is time inconsistent, other things constant. On the other hand, if the objective of the policymaker is to help beneficiaries to better face some form of risk such as income shocks, then transferring financial support in as few payments as possible seems to be a better policy, particularly if insurance markets do not work efficiently. This type of dilemmas will be analyzed more formally in section 4 where we create a second-best environment in which the policymaker must face potential tradeoffs implied by more committed, though less flexible, transfer schedules.

cient if and only if it is best for the first generation (self 1 in the current setting). Therefore, our result is a kind of corollary to the main proposition in Goldman's paper. More precisely, the stronger result we have obtained is a direct consequence of assuming intertemporal separability as well as concavity and differentiability of the instantaneous utility function.

2.3.2 Reestablishing Efficiency through Transfer Schemes

In the present setting, we have shown that any equilibrium allocation is inefficient when the policymaker transfers all of the resources in just one payment. In a first-best scenario where the policymaker has full information, it seems reasonable to expect that the best allocation from a long-run perspective can be obtained in equilibrium.¹² In fact, we will show that it is possible for the policymaker to implement *any* Pareto efficient allocation $x^* \in \mathbb{R}_+^T$ by doling out transfers such that cash on hand is equal to the optimal consumption path: i.e. $\omega_t = \tau_t + y_t = x_t^*$, for all t . Lemma 2.1 establishes this result more formally:

Lemma 2.1 *If the policymaker has full information and there is total flexibility in the way transfers can be allocated over time, then any efficient consumption allocation can be obtained in equilibrium. Moreover, there exists a unique perfect equilibrium supporting this allocation.*

PROOF: See Appendix.

We prove this proposition by applying the following line of logic. First, notice that for any efficient allocation $x^* \in \mathbb{R}_{++}^T$ the beneficiary is not overconsuming at any time period. Second, since he is not overconsuming, he has no incentive to transfer resources to the future *even if he actually could choose the point in time at which these resources will be consumed*. From here we obtain the result that the efficient allocation arises as the equilibrium allocation. Intuitively, the policymaker provides, through the lump-sum transfer scheme, a mechanism that makes the beneficiary commit to follow up an optimal consumption path.¹³

¹²Notice that the best allocation from a long-run perspective corresponds to the allocation that would be chosen by the beneficiary if he were time consistent.

¹³As a corollary to this Proposition, notice that it is always possible for the policy maker to implement a CMP that Pareto dominates the one-payment CMP.

Since the best allocation from a long-run perspective is efficient, the following result is a direct consequence of Lemma 2.1:

Proposition 2.2 *In a first-best setting where negative transfers can be implemented, the best allocation from a long-run perspective can be obtained in equilibrium.*

In a more realistic scenario, lump-sum transfers should be restricted to be non-negative. Most consumption maintenance programs do not impose any type of negative income transfer to their beneficiaries. We formally incorporate this feature by defining a new set \mathcal{B}_+^F of feasible transfers:

$$\mathcal{B}_+^F = \{(\tau_1, \dots, \tau_T) \in \mathbb{R}^T : \sum_{t=1}^T R^{1-t}\tau_t \leq B; \tau_t \geq 0 \forall t\}$$

The following corollary is a simple extension of Proposition 2 to the case with non-negative transfers.¹⁴

Corollary 2.1 *Given an efficient consumption profile $x^* \in \mathbb{R}_+^T$, if $y_t \leq x_t^* \forall t$, then x^* can be implemented with non-negative transfers.*

Intuitively, the larger the budget B is with respect to the present value of the beneficiary's income flow $\sum_{t=1}^T R^{1-t}y_t$, the more control the policymaker has over the flow of post-transfer income $\sum_{t=1}^T R^{1-t}(y_t + \tau_t)$. In consequence, the set of efficient consumption schedules that can be implemented expands as B gets larger.

2.4 Second-Best Consumption Maintenance Programs

The assumption that the policymaker has full information with respect to the beneficiary's income sequence, though helpful to establish a benchmark case to

¹⁴This result is very easily obtained as an extension of Proposition 2 by setting $\omega_t = x_t^*$ for all t . Since $y_t \leq x_t^*$, the policymaker sets $\tau_t = x_t^* - y_t$ for all t . This is a feasible choice since $\sum_t R^{1-t}(x_t^* - y_t) = \sum_t R^{1-t}\tau_t = B$

compare with, is clearly not representative of a more realistic CMP design. Incomes are far from being perfectly observable, especially in developing countries. Moreover, the assumption that the income process is deterministic does not seem to be a reasonable one since the poor are likely to face a highly uncertain economic environment. In this context, an optimal CMP should consider the existent trade-off between bringing commitment to the beneficiary with self-control problems and providing an insurance mechanism that help him overcome the ups and downs of everyday life. In other words, an optimal CMP should offer a package balancing both insurance and commitment motives.

We introduce uncertainty into the model by assuming that income is independently and identically distributed over time with probability distribution

$$y_t = \begin{cases} y_L & \text{with probability } \gamma \\ y_H & \text{with probability } 1 - \gamma \end{cases}$$

where $y_H > y_L$. We say that the beneficiary receives a negative income shock at time t if $y_t = y_L$. Analogously, we say the beneficiary receives a positive income shock at time t if $y_t = y_H$. For tractability and to keep the analysis as simple as possible we assume that the instantaneous utility function is exponential $u(x_t) = -\exp(-\alpha x_t)$. Let E_t be the expectation operator conditional on all information available at t , and let $E(-u(y_t)) = \mu < \infty$

It is assumed that, while the policymaker knows the distribution of income shocks, income realizations are not public information. Therefore, the efficient allocation of resources is impeded by the problem of incentive compatibility: if reporting a negative income shock in period t implies the reception of a higher transfer, then it is very likely that the beneficiary has an incentive to misreport his current income shock when it is positive.

Based on the revelation principle, the policymaker can restrict attention to direct revelation mechanisms with the property that the beneficiary truthfully reports her true income y_t .

For any period t , let τ_i represent the transfer at time t when the beneficiary reports income shock $i \in \{H, L\}$, and let τ'_i be the corresponding budget left at $t + 1$. In period $T - 1$, the policymaker solves the problem

$$\max_{\tau_L, \tau_H, \tau'_L, \tau'_H} \gamma[u(\tau_L + y_L) + \delta E_{T-1}u(\tau'_L + y_T)] + (1 - \gamma)[u(\tau_H + y_H) + \delta E_{T-1}u(\tau'_H + y_T)]$$

subject to the following incentive-compatibility and resource constraints

$$u(\tau_L + y_L) + \beta\delta E_{T-1}u(\tau'_L + y_T) \geq u(\tau_H + y_L) + \beta\delta E_{T-1}u(\tau'_H + y_T) \quad (2.2)$$

$$u(\tau_H + y_H) + \beta\delta E_{T-1}u(\tau'_H + y_T) \geq u(\tau_L + y_H) + \beta\delta E_{T-1}u(\tau'_L + y_T) \quad (2.3)$$

$$\tau_L + R^{-1}\tau'_L \leq B_{T-1} \quad (2.4)$$

$$\tau_H + R^{-1}\tau'_H \leq B_{T-1} \quad (2.5)$$

where B_{T-1} is the budget left at time $T - 1$. Define by $v_{T-1}(B_{T-1})$ the value function of this problem. By standard arguments, $v_{T-1}(B_{T-1})$ is strictly concave and differentiable.

Next, take any period t and suppose $v_{t+1}(B_t)$ is strictly concave and differentiable. Although the policymaker and the beneficiary disagree on the amount of discounting applied between t and $t + 1$, they both agree on the utility obtained from $t + 1$ on. By applying a standard induction argument, we have that for all t the planner solves the problem:

$$\max_{\tau_L, \tau_H, \tau'_L, \tau'_H} \gamma[u(\tau_L + y_L) + \delta v_{t+1}(\tau'_L)] + (1 - \gamma)[u(\tau_H + y_H) + \delta v_{t+1}(\tau'_H)]$$

subject to the following incentive compatible and budget constraints:

$$u(\tau_L + y_L) + \beta\delta v_{t+1}(\tau'_L) \geq u(\tau_H + y_L) + \beta\delta v_{t+1}(\tau'_H) \quad (2.6)$$

$$u(\tau_H + y_H) + \beta\delta v_{t+1}(\tau'_H) \geq u(\tau_L + y_H) + \beta\delta v_{t+1}(\tau'_L) \quad (2.7)$$

$$\tau_L + R^{-1}\tau'_L \leq B_t \quad (2.8)$$

$$\tau_H + R^{-1}\tau'_H \leq B_t \quad (2.9)$$

In what follows, we will characterize the equilibrium arising in the current setting. First, we introduce the following result that states that when the beneficiary receives a "negative" income shock he must be transferred at least the same amount than in the case where he receives a "positive" income shock in order to have an incentive-compatible equilibrium.

Lemma 2.2 $\tau_L \geq \tau_H$ in equilibrium.

PROOF: See Appendix.

The policymaker faces a tradeoff: on the one hand, he must take into account the fact that the beneficiary has a self-control problem, implying that he has an incentive to report y_L when he actually received a positive income shock. On the other hand, the policymaker plays the role of an insurer who should provide a higher transfer when the agent receives a negative income shock. In other words, the policymaker considers both benefits and costs of implementing a more "committed", though less flexible, CMP.

It was argued above that the self control problem can be parameterized by β : the lower this parameter, the stronger the preference for immediate gratification. In his role of insurer, the policymaker should consider some measure of risk that considers somehow the dispersion of income shocks. We define the following measure of risk:

$$\psi = -u(y_H - y_L) = \exp(-\alpha(y_H - y_L))$$

This measure integrates a constant $\alpha > 0$, a measure of the degree or risk aversion of the beneficiary, and a measure of the dispersion of the income shock $y_H - y_L$. This measure is based on the idea that a beneficiary's sense of well being depends on the risk he faces. We have the following result

Proposition 2.3 *If income shocks are unobservable, then the optimal CMP is designed as follows*

- i) If $\beta \leq \psi$: $\tau_H = \tau_L$ (pooling equilibrium).*
- ii) If $\beta > \psi$: $\tau_H < \tau_L$ (separating equilibrium).*

PROOF: See Appendix.

Proposition 2.3 establishes that if the beneficiary's self-control problem (parameterized by β) is relatively more serious than the vulnerability problem he faces (parameterized by ψ), then the policymaker optimally opts for a pooling equilibrium where he transfers τ^* independently of the value that the income shock takes, where τ^* satisfies $u'(\tau^*) = v'_{t+1}(B_t - \tau^*)$.

Income reports are a mechanism to extract private information that may be helpful for the design of a more efficient transfer schedule in the presence of risk. Specifically, having information on actual realizations of income shocks makes consumption smoothing an easier task for the policymaker. However, if the degree of self control is too low, the policy maker's optimal response is to offer a non-contingent transfer schedule. This is equivalent to commit to a transfer schedule at period 0, before the consumption-savings game starts. Therefore, the value of information is zero for low levels of self-control.

2.5 Conclusions

We have analyzed the problem of designing an optimal transfer schedule when the beneficiary is a dynamically-inconsistent decision maker. When he has total control over the resources from the beginning of the period under consideration, the outcome is generally inefficient. This questions the traditional view that providing more liquidity to the poor and making capital markets work more efficiently are sufficient conditions to generate efficient outcomes. In a world with imperfect individuals, perfect markets may not generate the best possible equilibrium.

If the policymaker has full information and lump-sum transfers are not restricted to be non-negative, then any efficient consumption allocation can be obtained in equilibrium. By imposing constraints on future cash-on-hand, the policymaker is able to influence the pattern of expenditure in future periods and, in consequence, to reestablish efficiency. Obviously, the set of efficient allocations that can be obtained in equilibrium is more restricted when lump-sum transfers cannot be negative: the policymaker has less influence on the final arrangement of the income flow. However, for many, if not most, CMP the budget B represents an important proportion of the total amount of resources available to the the beneficiary. This fact provides the policymaker with more degrees of freedom for reallocating resources and obtaining more efficient outcomes by means of exercising more control over the beneficiary's income flow. In this sense, a transfer schedule is a kind of commitment mechanism.

One potential drawback of this first-best approach is that, although helpful to establish a benchmark case, it does not provide an accurate description of the circumstances that a policymaker usually has to face when allocating benefits to the poor or the unemployed. Information is far from being public, and many charac-

teristics of the beneficiary, particularly income, are hidden information. Another problem is that a reasonable goal of a CMP is to help beneficiaries to face certain types of risk such as income shocks. This means that the policymaker faces a dilemma since an optimal transfer schedule explicitly designed for dealing with risky environments should be as flexible as possible. However, if the beneficiary has self-control problems, the role of the policymaker as an insurer may imply important trade-offs with its role as a commitment provider. In fact, if the self-control problem is relatively serious with respect to the degree of income uncertainty, the value of obtaining information through income reports is likely to be very low, or even negative if implementing such a mechanism implies some sort of cost such as administrative and data collection costs.

Our analysis has several limitations and possible extensions. First, we do not explicitly consider the possibility of social commitment mechanisms. This type of mechanisms are likely to arise in small communities where individuals are closer to each other and information is semi-public. In some communities, insurance mechanisms among their members naturally arise. Should we expect the same for social commitment devices such as peer pressure? Second, there may exist less interventionist commitment technologies. For instance, the policymaker could provide the beneficiary with an illiquid instrument a la Laibson (1997). He could also offer a more sophisticated mechanism where the beneficiary has the option to choose a transfer schedule from a menu. If he is aware of his self-control problem, the final consumption allocation would be the best from a current perspective, and hence efficient. Third, the second-best results of this paper could be extended to preferences outside the neighborhood of constant absolute risk aversion. Fourth, it could be assumed that income shocks are not i.i.d., following instead another

type of random process. In reality, income realizations may not be independent: a bad draw may generate a series of bad draws. In fact, the analysis of poverty traps in development economics is based on this type of dynamic mechanism. It would be very interesting to find out what the behavior of time inconsistent beneficiaries could be in such a scenario as well as to study the optimal response of the policy maker. Finally, we could introduce naivete into the model and design an optimal mechanism that takes into account the possibility of facing a mixture of sophisticated and naive individuals within the target population.

2.6 Appendix: Proofs

Lemma 2.3 *Let $x^* \in \mathbb{R}_+^T$ be some equilibrium consumption allocation. If there exist periods j and t , $j > t$, such that $u'(x_t^*) < \beta\delta^{j-t}R^{j-t}u'(x_j^*)$, then the allocation is inefficient.*

PROOF: First, we will show that there exists $\varepsilon > 0$ such that $u'(x_t^* - \varepsilon) \geq \beta\delta^{j-t}R^{j-t}u'(x_j^* + \varepsilon)$. Define the function $\phi(\varepsilon, \beta) = u'(x_t^* - \varepsilon) - \beta\delta^{j-t}R^{j-t}u'(x_j^* + \varepsilon)$. Since $u(\cdot)$ is concave and twice differentiable, we have

$$\phi'(\varepsilon, \beta) = -u''(x_t^* - \varepsilon) - \beta\delta R^{j-t}u''(x_j^* + \varepsilon)$$

which is strictly positive for all $\varepsilon \geq 0$. Since $\phi(\varepsilon, \beta)$ is continuous, there exists some $\varepsilon > 0$ such that $\phi(0, \beta) < \phi(\varepsilon, \beta) \leq 0$. Hence, by concavity of the utility function, it follows that transferring ε to period i strictly increases the welfare of selves t to j keeping the welfare of selves $j + 1$ to T constant since preferences are strictly monotone and separable. Next, we claim that the welfare of selves 1 to $t - 1$ strictly increases by transferring such amount of consumption from period t to period j . By a similar argument to the one presented above, it suffices to show that $\phi(0, \beta) < 0$ implies that $\delta^{t-\tau}u'(x_t^*) - \delta^{j-\tau}R^{j-t}u'(x_j^*) < 0$, or equivalently $\delta^{t-\tau}\phi(0, 1) < 0$, and that for any $\varepsilon > 0$ satisfying $\phi(\varepsilon, \beta) \leq 0$ we have $\delta^{t-\tau}\phi(\varepsilon, 1) \leq 0$, for all $\tau \in \{1, \dots, t - 1\}$. This follows immediately since $\delta^{t-\tau}\phi(\varepsilon, 1) \leq \phi(\varepsilon, \beta)$ for all $\varepsilon \geq 0$ and $\beta \in (0, 1]$. \square

Lemma 2.4 *An allocation $x \in \mathbb{R}_+^T$ satisfying*

$$u'(x_{T-2}) \geq \max[\beta\delta Ru'(x_{T-1}), \beta\delta^2 R^2 u'(x_T)]$$

cannot be an equilibrium allocation.

PROOF: Define the set $\Phi = \{\lambda \in \mathbb{R}^3 \mid \lambda_{T-2} = 1, \lambda_{T-1} \leq 0, \lambda_T \leq 0, \lambda_{T-1} + \lambda_T = -1\}$, and the function $\varphi(\tau) = u(x_{T-2}^* + \tau \lambda_{T-2}) + \beta \sum_{t=T-1}^T \delta^{t-T+2} u(x_t^* + R^{T-2-t} \tau \lambda_t)$.

Taking the second derivative of the function $\varphi(\cdot)$ we have

$$\varphi''(\tau) = u''(x_{T-2}^* + \tau) + \beta R^2 \lambda_{T-1}^2 u''(x_{T-1}^* + R \lambda_{T-1} \tau) + \beta R^3 \lambda_T^2 u''(x_T^* + R \lambda_T \tau)$$

which is clearly strictly negative for all $\lambda \in \Phi$, and, in consequence, strictly concave.

Hence, $\tau(\lambda) = \arg \max_{\tau \in \mathbb{R}_+} \varphi(\tau)$ is a continuous function on Φ by the Maximum theorem. Taking the first derivative of $\varphi(\tau)$ and evaluating it at $\tau = 0$, we have

$$\begin{aligned} \varphi'(0) &= u'(x_{T-2} + \beta R \lambda_{T-1} u'(x_{T-1}^*) + \beta R^2 \lambda_T u'(x_T^*)) \\ &= u'(x_{T-2}^* + \lambda_{T-1} \beta \delta R u'(x_{T-1}) + \lambda_T \beta \delta^2 R^2 u'(x_T)) \\ &> u'(x_{T-2}^*) - \max[\beta \delta R u'(x_{T-1}), \beta \delta^2 R^2 u'(x_T)] \\ &\geq 0 \end{aligned}$$

Where the last inequality follows from the initial hypothesis. This shows that the optimum is strictly positive on Φ : i.e. $\tau(\lambda) > 0$, for all $\lambda \in \Phi$. Since Φ is a compact set, $\tau(\lambda)$ attains its minimum on Φ by Weierstrass theorem. Let $v = \min_{\lambda \in \Phi} \tau(\lambda)$, and take any $\bar{\tau} \in (0, v)$.

Let $s_t(\omega)$ be the consumption strategy of self t when cash on hand at that period is equal to ω , and define $\Delta_{T-1} = s_{T-1}^*(\omega_{T-1} - \bar{\tau}) - s_{T-1}^*(\omega_{T-1})$. By the argument above, $s_{T-2}^*(\omega_{T-2}) + \bar{\tau}$ is an optimal deviation if $(1, \frac{\Delta_{T-1}}{\bar{\tau}}, \frac{\Delta_{T-1}}{\bar{\tau}} - 1) \in \Phi$. In period T , the agent will consume all resources left. Thus, her equilibrium strategy is given by $s_T(\omega) = \omega$, so all we need to show is that $\frac{\Delta_{T-1}}{\bar{\tau}} \in [0, 1]$. In period $T-1$, there is no dynamic inconsistency, so the optimal strategy is obtained by solving

$$s_{T-1}(\omega) = \arg \max u(x_{T-1}) + \beta \delta u(x_T)$$

subject to the constraint $x_{T-1} + R^{-1} x_T = \omega$. First order conditions are given by

$$u'(s_{T-1}(\omega)) = \beta \delta R u'(\omega - s_{T-1}(\omega))$$

Differentiating with respect to ω at both sides of the equality, and after some algebraic manipulations, we obtain

$$\frac{s'_{T-1}(\omega)}{1-s'_{T-1}(\omega)} = \frac{u''(\omega - s_{T-1}(\omega))}{u''(s_{T-1}(\omega))}$$

Since $u(\cdot)$ is strictly concave, we must have $s'_{T-1}(\omega) \in (0, 1)$. By the Mean Value Theorem, there exists $\eta \in (\omega_{T-1} - \bar{\tau}, \omega_{T-1})$, such that $\frac{\Delta_{T-1}}{\bar{\tau}} = s'(\eta) \leq 1$. Hence $(1, \frac{\Delta_{T-1}}{\bar{\tau}}, \frac{\Delta_{T-1}}{\bar{\tau}} - 1) \in \Phi$ and the result follows. \square

Proof of Proposition 2.1: The result is a direct consequence of Lemmas 2.3 and 2.4. \square

Proof of Lemma 2.1: First, we show that $\tau = x^*$ arises as the equilibrium allocation. Let s^* be a Markov perfect equilibrium of the post-transfer game. In order to prove the result, it suffices to show that $s_t^*(\tau_t) = \tau_t$ for all t , where $\tau = x^* \in \mathbb{R}_+^T$ is some efficient allocation. For period T , it is trivially true that this is the best strategy since player T will consume all resources on hand. Next, assume $s_j^*(\tau_j) = \tau_j$ for all $j > t$. I claim that optimal strategy for player t implies $s_t^*(\tau_t) = \tau_t$. Assume, towards a contradiction, that $s_t^*(\tau_t) = \tau_t - \varepsilon$, for some $\varepsilon > 0$, and let $(\tau_t - \varepsilon, x'_{t+1}, \dots, x'_T)$ be the new consumption allocation from period t to T . Since players $t + 1$ to T are playing the strategy $s_j(\tau_j) = \tau_j$ by hypothesis, notice that, for any $\varepsilon > 0$, all of them have the option of obtaining a utility of at least $U_j(\tau_j, \tau_{j+1}, \dots, \tau_T)$, and hence $U_j(x'_j, x'_{j+1}, \dots, x'_T) \geq U_j(\tau_j, \tau_{j+1}, \dots, \tau_T)$, for all $j \in \{t, \dots, T\}$, with strict inequality for player $t + 1$ since $u(\cdot)$ is strictly monotone and $\omega_{t+1} = \tau_{t+1} + \varepsilon$. This implies that $(\tau_1, \dots, \tau_{t-1}, \tau_t - \varepsilon, x'_{t+1}, \dots, x'_T)$ Pareto dominates (τ_1, \dots, τ_T) , a contradiction. Next, we show uniqueness. I claim

that any efficient allocation $x^* \in R_+^T$ satisfies:

$$u'(x_t^*) \geq \beta(\delta R)^\tau u'(x_{t+\tau}^*)$$

for all $t, \tau \geq 1$. Assume, towards a contradiction, that there exist τ' and t' such that this condition does not hold. Then

$$u'(x_{t'}^*) < \beta(\delta R)^\tau u'(x_{t'+\tau'}^*)$$

This implies that x^* cannot be an efficient allocation since, by concavity of $u(\cdot)$, there exists $\varepsilon > 0$ satisfying

$$u'(x_{t'}^* - \varepsilon) \leq \beta(\delta R)^\tau u'(x_{t'+\tau'}^* + \varepsilon)$$

Which is clearly a Pareto improvement. Uniqueness follows from Theorem 1 in Laibson (1997b). \square

Proof of Lemma 2.2. Define $V(x) = \beta\delta v_{t+1}(x)$, and notice that $\tau'_i = R(B_t - \tau_i)$ and $u(\tau_i + y_i) = -u(\tau_i)u(y_i)$, $i = L, H$. Assume, towards a contradiction, that $\tau(\theta_L) > \tau(\theta_H)$. From the incentive compatible constraints we have

$$u(y_H) \leq \frac{V(\tau'_H) - V(\tau'_L)}{u(\tau_H) - u(\tau_L)}$$

and

$$u(y_L) \geq \frac{V(\tau'_H) - V(\tau'_L)}{u(\tau_H) - u(\tau_L)}$$

Hence $u(y_H) \leq u(y_L)$, a contradiction. \square

Proof of Proposition 2.3. It is easier to analyze the problem if we first define some variables. Let $\mu = E_j[-u(y_t)]$, $\varphi_L = -u(y_L)$, and $\varphi_H = -u(y_H)$. Moreover, make the following change of variables: instead of having τ_L and τ_H as our decision variables, let the decision variables be $u_H = u(\tau_H)$, $u'_H = v(\tau'_H)$, $u_L = u(\tau_L)$ and $u'_L = v(\tau'_L)$. Since an exponential utility function can be decomposed as $u(y+x) = -u(y)u(x)$ and after some algebraic manipulations, the problem becomes

$$\begin{aligned} \max \quad & \gamma[\varphi_L u_L + \delta u'_L] + (1 - \gamma)[\varphi_H u_H + \delta u'_H] \\ \text{s.t.} \quad & \end{aligned}$$

$$\varphi_L u_L + \beta \delta u'_L - \varphi_L u_H - \beta \delta u'_H \geq 0$$

$$u_H + \beta \delta u'_H - \varphi_H u_L - \beta \delta u'_L \geq 0$$

$$B - V_1(u_L) - R^{-1}V_2(u'_L) \geq 0$$

$$B - V_1(u_H) - R^{-1}V_2(u'_H) \geq 0$$

where V_1 and V_2 are the inverse functions of $u(\cdot)$ and $v(\cdot)$, respectively. The Lagrangean for this problem is given by the function

$$\mathcal{L} = \phi(u_L, u_H, u'_L, u'_H) + \sum_{i=1}^4 \lambda_i \phi_i(u_L, u_H, u'_L, u'_H)$$

where $\phi(\cdot)$ represents the objective function, while ϕ_i corresponds to constraint i , starting from above. Because $u(\cdot)$ and $v(\cdot)$ are concave functions, $V_i : R \rightarrow R$, $i = 1, 2$, is convex. Hence, ϕ_i , $i = 1, \dots, 4$ is concave.

Notice that the objective function is linear, so by the Theorem of Kuhn and Tucker, $u^* = (u_H, u'_H, u_L, u'_L)$ is a solution to the problem above if and only if there is $\lambda^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*) \in R_+^4$, where λ_i^* is the corresponding multiplier for constraint i , $i = 1, \dots, 4$, such that the following Kunh-Tucker first order conditions hold:

$$\gamma \varphi_L + \lambda_1 \varphi_L - \lambda_2 \varphi_H - \lambda_3 V'_1(u_L) = 0 \quad (2.10)$$

$$\gamma + \lambda_1 \beta - \lambda_2 \beta - \lambda_3 R^{-1} \delta^{-1} V'_2(u'_L) = 0 \quad (2.11)$$

$$(1 - \gamma) \varphi_H - \lambda_1 \varphi_L + \lambda_2 \varphi_H - \lambda_4 V'_1(u_H) = 0 \quad (2.12)$$

$$(1 - \gamma) - \lambda_1 \beta + \lambda_2 \beta - \lambda_4 R^{-1} \delta^{-1} V'_2(u'_H) = 0 \quad (2.13)$$

In a pooling equilibrium the policymaker solves

$$\max_{\tau} \gamma[u(\tau + y_L) + \delta v_{t-1}(R(B - \tau))] + (1 - \gamma)[u(\tau + y_H) + \delta v_{t-1}(R(B - \tau))]$$

After some algebraic manipulations, this problem is equivalent to solve

$$\max_{\tau} \mu u(\tau) + \delta v_{t-1}(R(B - \tau))$$

The solution is implicitly given by $\mu u'(\tau) = \delta R v'(\tau')$, where $\tau' = R(B - \tau)$. This implies

$$\delta^{-1} R^{-1} V_2' = \mu^{-1} V_1'$$

Given this condition, the Kuhn-Tucker first order conditions for a pooling equilibrium can be rewritten as follows

$$\gamma \varphi_L + \lambda_1 \varphi_L - \lambda_2 \varphi_H = v_1 \lambda_3 \quad (2.14)$$

$$\gamma + \lambda_1 \beta - \lambda_2 \beta = \mu^{-1} v_1 \lambda_3 \quad (2.15)$$

$$(1 - \gamma) \varphi_H - \lambda_1 \varphi_L + \lambda_2 \varphi_H = v_1 \lambda_4 \quad (2.16)$$

$$(1 - \gamma) - \lambda_1 \beta + \lambda_2 \beta = \mu^{-1} v_1 \lambda_4 \quad (2.17)$$

From equations 14)-16) or 17)-18) we have:

$$\lambda_2 = \frac{\gamma(\mu - \varphi_L)}{\mu\beta - \varphi_H} + \frac{\mu\beta - \varphi_L}{\mu\beta - \varphi_H} \lambda_1 \quad (2.18)$$

From where it can be concluded that necessary and sufficient conditions for having positive λ_1 and λ_2 multipliers are

$$\varphi_L > \mu\beta \quad (2.19)$$

$$\lambda_1 \geq \frac{\gamma(\mu - \varphi_L)}{\varphi_L - \mu\beta} \quad (2.20)$$

Positive λ_3 and λ_4 are obtained if and only if the following conditions are satisfied

$$\gamma \varphi_L + \lambda_1 \varphi_L - \lambda_2 \varphi_H \geq 0 \quad (2.21)$$

$$\gamma + \lambda_1 \beta - \lambda_2 \beta \geq 0 \quad (2.22)$$

$$(1 - \gamma) \varphi_H - \lambda_1 \varphi_L + \lambda_2 \varphi_H \geq 0 \quad (2.23)$$

$$(1 - \gamma) - \lambda_1 \beta + \lambda_2 \beta \geq 0 \quad (2.24)$$

equivalently

$$\lambda_2 \leq \frac{\varphi_L}{\varphi_H} + \frac{\varphi_L}{\varphi_H} \lambda_1 \quad (2.25)$$

$$\lambda_2 \leq \frac{\gamma}{\beta} + \lambda_1 \quad (2.26)$$

$$\lambda_2 \geq -(1 - \gamma) + \frac{\varphi_L}{\varphi_H} \lambda_1 \quad (2.27)$$

$$\lambda_2 \geq -\frac{1 - \gamma}{\beta} + \lambda_1 \quad (2.28)$$

Condition (25) implies condition (26). Therefore, expressions 18)-20), 25), and 27)-28) together provide a set of necessary and sufficient parametric restrictions for a pooling equilibrium to exist.

Conditions 18) and 25) imply:

$$\lambda_1 \geq \frac{\varphi_L \mu \beta - \varphi_H}{\beta \mu (\varphi_H - \varphi_L)} \quad (2.29)$$

which is trivially satisfied for any $\lambda_1 \geq 0$. Conditions 18) and 27)-28) are satisfied if and only if:

$$\lambda_1 \leq (1 - \gamma) \frac{\varphi_H}{\varphi_H - \varphi_L} \frac{1 - \beta}{\beta} \quad (2.30)$$

Therefore, all of these conditions above are satisfied if and only if

$$(1 - \gamma) \frac{\varphi_H}{\varphi_H - \varphi_L} \frac{1 - \beta}{\beta} \geq \frac{\gamma(\mu - \varphi_L)}{\varphi_L - \mu\beta} \quad (2.31)$$

and condition (19) are satisfied.

Define the function

$$\phi(\beta) = \frac{1 - \beta}{\beta} (\varphi_L - \mu\beta)$$

Condition (31) can be rewritten as follows

$$\phi(\beta) \geq \frac{\gamma}{1 - \gamma} \frac{\varphi_H - \varphi_L}{\varphi_H} (\mu - \varphi_L) \quad (2.32)$$

Define by β^* the value of β that makes equation (31) hold with equality. After some manipulations, we have $\beta^* = \frac{\varphi_L}{\varphi_H}$. Since $\phi(\cdot)$ is strictly decreasing in the set

$[0, \frac{\varphi_L}{\mu})$, and $\beta^* < \frac{\varphi_L}{\mu}$, we have that condition (32) is satisfied if and only if $\beta \leq \beta^*$.

This completes the proof. \square

Chapter 3
Does Conditionality Generate
Heterogeneity and Regressivity in
Program Impacts?
The Progresa Experience

3.1 Introduction

Nowadays, conditional cash transfer schemes (CCTS) constitute a key element of many anti-poverty programs around the world. Following Das, Do, and Oler (2004), a conditional cash transfer scheme can be broadly defined as "...any scheme requiring a specified course of action in order to receive a benefit as a conditional cash transfer". Examples of programs implementing CCTS are Oportunidades in Mexico, Red de Proteccion Social in Nicaragua, and Bolsa Familia in Brazil. The aim of these programs is to alleviate today's poverty by transferring money to poor families, and to short-circuit tomorrow's, by making the transfers conditional. The conditionality usually operates through lower bounds on human capital investment which takes the form of requiring a minimum attendance rate to school, and constant health monitoring for the children.

What is the rationale behind imposing a conditionality to the beneficiaries of a social program? if individuals are rational, there are no externalities, and policymakers have full information, then there is no case for implementing CCTS. However, these conditions are rarely met. If individuals are not fully rational, imposing a conditionality may help them to increase their own welfare. For instance, if a beneficiary is time inconsistent, then it may be optimal to impose the condition that transfers should be received in several payments.¹ If information is asymmetric in the sense that the policymaker does not have some relevant information of the beneficiaries such as income and asset holdings, then CCTS can be used as a screening mechanism with the specific purpose of improving the targeting efficiency of the program. For example, if the conditioned-on good is inferior, richer households are more likely to be screened out of the program (Besley and Coate

¹See chapter 2 in this dissertation.

1992).

There is a third rationale for CCTS. In the presence of externalities, individuals do not internalize the effect of their choices on others. By imposing a conditionality, policymakers may be able to move individuals towards a more efficient equilibrium. One notorious case is that of child labor and human capital investment: parents usually decide children's time allocation between education and work. Since the economic benefits of child labor are immediately felt, and the economic benefits of education are only feasible in the long run, parents may not internalize the benefits of human capital investment in their children. Therefore, CCTS can be used in this case to restore efficiency by imposing lower bounds on variables such as school attendance.

Although CCTS may help policymakers to reach a more efficient economy and to reduce poverty in the long run by increasing investments on human capital today (so the children of the poor may escape poverty in the future), they could imply a tradeoff between the equity and efficiency goals of policymakers, at least in the short run. In particular, if the conditioned-on good is normal, then worse off households may be receiving less "effective" transfers than other groups of beneficiaries if participating in the program imposes some sort of opportunity cost such as foregone wages from child labor.

A good example of this tension is The Female Stipend Program in Bangladesh. This program gives stipends to girls who attend at least 85% of classes at a secondary school level with the explicit goal of increasing investment on human capital. All girls can participate in this program independently of their socioeconomic background. Since education is usually a normal good, richer households are more likely to enroll their daughters in secondary schools than households in the low

tail of the income distribution. Besides, the opportunity cost of enrolling a child into school or making the 85% lower bound on school attendance is more likely to exceed the benefits obtained from the stipend for the poorest households. Khandker et al (2003) notice that the "...untargeted stipend disproportionately affects the school enrollments of girls from households with larger land wealth. Targeting towards the land poor may reduce the overall enrollment gains of the program while equalizing enrollment effects across landholding classes."

Despite the potential effects that CCTS programs have on the distribution of treatment effects, most existing research on program evaluation of anti-poverty programs focuses on mean impacts. There are, however, some studies for developed countries that take the issue of heterogeneity in program impacts into account. In an excellent study about heterogeneity in program impacts, Heckman, Smith and Clements (1997) find strong evidence of heterogeneous impacts when evaluating the US Job Training Partnership Act. In a similar spirit, Bitler, Gelbach, and Hoynes (2003) study the Connecticut's Job First program; they conclude that this welfare program exhibits a lot of heterogeneity in program impacts, just as predicted by standard labor supply theory.

In this paper, we study the distribution of program impacts in Progresa, recently renamed Oportunidades. This anti-poverty program was introduced by the Mexican government in 1997 and provides conditional cash transfers to poor families. Similar to The Female Stipend Program in Bangladesh, the conditioned-on good is school attendance which, not surprisingly, is a normal good in the case of Mexico (Lopez-Acevedo and Salinas 2000). We take advantage of the experimental design of the evaluation sample to identify the parameters of interest for this study.

Our empirical findings can be summarized in two main points. First, there is

strong evidence that heterogeneity in program impacts is a common phenomenon in Progresa. Second, under the assumption of perfect positive dependence, and consistent with the model developed in this paper, better off households tend to receive larger positive program impacts than poorer households.

The paper is organized as follows. Section 2 describes Progresa, the evaluation sample, and the selection of beneficiaries. Section 3 develops a simple household bargaining model of child labor and human capital accumulation, and discusses its connection with CCTS. Section 4 briefly analyzes the evaluation problem and presents average treatment effects of the program as a benchmark case. Section 5 develops some tests for homogeneity in program impacts. Section 6 imposes a specific type of monotonicity assumption: rank preservation, and makes this assumption operational through the estimation of quantile treatment effects (QTE). Section 7 concludes. Mathematical details, algorithms, and proofs are in the Appendices.

3.2 Progresa

In 1997, the Mexican government introduced the Programa de Educacion, Salud y Alimentacion (the Education, Health, and Nutrition Program), better known as Progresa, and recently renamed Oportunidades, as an important element of its more general strategy to eradicate poverty in Mexico. The program is characterized by a multiplicity of objectives such as improving the educational, health, and nutritional status of poor families.

Progresa provides cash transfers, in-kind health benefits, and nutritional supplements to beneficiary families. Moreover, the delivery of the cash transfers is exclusively through the mothers, and is linked to children's enrollment and school

attendance. The conditionality works as follows: in localities where Progresa operates, those households classified as poor with children enrolled in grades 3 to 9 are eligible to receive the grant every two months. The average bi-monthly payment to a beneficiary family amounts to 20 percent of the value of bi-monthly consumption expenditures prior to the beginning of the program. Moreover, these grants are estimated taking into account the opportunity cost of sending children to school, given the characteristics of the labor market, household production, and gender differences. By the end of 2002, nearly 4.24 million families (around 20 percent of all Mexican households) were incorporated into the program. These households constitute around 77 percent of those households considered to be in extreme poverty.

3.2.1 Data: A Quasi-Experimental Design

Because of logistical and financial constraints, the program was introduced in several phases. This sequentiality in the implementation of the Progresa was capitalized by randomly selecting 506 localities in the states of Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi and Veracruz. Of the 506 localities, 320 localities were assigned to the treatment group and the rest were assigned to the control group. In total, 24,077 households were selected to participate in the evaluation sample. The first evaluation survey took place in March 1998, 2 months before the distribution of benefits started. 3 rounds of surveys took place afterwards: October/November 1998, June 1999, and November 1999. The localities that served as control group started receiving benefits by December 2000. For the empirical application of the methodologies developed in this paper, we will make use of the June 1999 round.

3.2.2 Progresa's selection of localities and beneficiary households

Progresa's methodology to identify potential beneficiaries consists of two two main stages: (1) the selection of localities; (2) the selection of beneficiary households within selected localities. For the first stage, a marginality index was constructed for each locality in Mexico. Based on this index, localities deemed to have a high marginality level and with more than 50 and less than 2,500 inhabitants were considered priorities for the program. Finally, budgetary constraints as well as program components that require the presence of school and clinics for the implementation of the program were considered to select the group of localities to be covered by the Progresa. For the second stage, a census, ENCASEH (Encuesta de Características Socioeconómicas de los Hogares), was conducted in each of the selected localities. Using this data, a measure of monthly per capita income per household was constructed subtracting child income from total household income. A poverty line of 320 pesos per capita per month was employed to create a new binary variable taking the value of 1 if household's monthly per capita income was below 320 pesos and 0 otherwise. Finally, discriminant analysis was employed for each geographical region. By doing so, it was possible to identify the variables that discriminate best between poor and non-poor households, and a rule to classify households as poor or non-poor was developed by estimating a discriminant score for each household.

3.3 A Simple Model of Human Capital Investment and CCTS

In this section, we present a simple model of child labor and human capital investment. Our objective is to shed some light on the connection between these variables and CCTS. We build this model as an extension of Baland and Robinson (2000), although we do not adopt a unitary view of the household. Similar to Kanbur and Haddad (1997) and Martinelli and Parker (2003), we adopt a bargaining perspective for the intra-household resource allocation problem.

3.3.1 One-Sided Altruism

We consider a one-good economy. The single good in this economy is produced with the linear technology

$$Y = L \tag{3.1}$$

where L is labor input measured in efficiency units of labor. We assume that the labor market is perfectly competitive.

There is a continuum of households who live for two periods, $t = 1, 2$. Each of these households is composed by a man, a woman, and a child. We will refer to the man and the woman together as the parents for the rest of the section. In period 1, parents are characterized by their income generating ability a , where a also represents efficiency units of labor. We assume that households are distributed uniformly on $[\underline{a}, \bar{a}]$, with each household inelastically supplying a efficiency units of labor per period.

In period 1, the child is endowed with one unit of time. Parents decide how to allocate the child's time between child labor, l , and human capital accumulation,

h . They also decide how much to leave as a bequest to the child, b . For the sake of simplicity, it is assumed that l is measured in efficiency units of labor, so the child is endowed with one efficiency unit of labor in period 1. In the second period, the child's income generating ability is given by $\phi(h)$, where $h = 1 - l$ and $\phi(\cdot)$ is C^2 , strictly increasing, and strictly concave function defined on $[0, 1]$, with $\phi(0) = 1$, $\phi'(1) < 1$, and $\phi'(0) > 1$. This technology implies that the efficient investment level on human capital, h^o , is given implicitly by $\phi'(h^o) = 1$.²

Let (x_{1f}, x_{2f}) and (x_{1m}, x_{2m}) denote the consumption levels of the father and the mother for periods 1 and 2, respectively. The child is assumed to consume only in period 2, with consumption level denoted by x_c . The woman cares only about her own consumption and the consumption of the child; similarly, the man cares only about his own consumption and the consumption of the child. The father's preferences are represented by

$$W_f = \alpha(\ln x_{1f} + \ln x_{2f}) + (1 - \alpha) \ln x_c \quad (3.2)$$

and the mother's preferences are given by

$$W_m = \beta(\ln x_{1m} + \ln x_{2m}) + (1 - \beta) \ln x_c \quad (3.3)$$

where $1 > \alpha > \beta > 0$.

Besides choosing the time allocation of the child, parents can also decide to make positive bequests to him. We denote these bequest by $b \in \mathbb{R}_+$. Parents have access to a storage technology, so they can transfer resources between periods by saving. We denote the household's saving level by s . Households are borrowing constrained in the sense that parents can save but not borrow. Therefore, parents

²In other words, h^o is the level of human capital that maximizes the household's intertemporal income.

face the budget constraints

$$x_{1f} + x_{1m} = a + l - s \quad (3.4)$$

$$x_{2f} + x_{2m} = a + s - b, \quad (3.5)$$

and

$$x_c = \phi(1 - l) + b \quad (3.6)$$

Decisions about $x_{1f}, x_{2f}, x_{1m}, x_{2m}, x_c, b$, and s are made by the parents in the first period by solving a *generalized Nash bargaining problem* with solution given by the following program

$$\text{Max}(x_{1f}^\alpha x_{2f}^\alpha x_c^{1-\alpha} - u_f)^\gamma (x_{1m}^\beta x_{2m}^\beta x_c^{1-\beta} - u_m)^{1-\gamma} \quad (3.7)$$

The parameter $\gamma \in (0, 1)$ introduces asymmetry into the model. The ratio $\frac{\gamma}{1-\gamma}$ can be interpreted as the relative bargaining power of the father with respect to the mother. u_f and u_m are referred to as threat points or disagreement points. For the rest of the analysis we assume $u_f = u_m = 0$.

Proposition 3.1 *If savings and bequests are interior, then parents are investing the efficient level of human capital on the child. Moreover, human capital is a normal good.*

PROOF: See Appendix.

3.3.2 Two-Sided Altruism

We now introduce a particular form of altruism from children to parents. We will show that the results we obtained above can be extended to this new setting.

We assume that children derive utility both from consumption in the second period and from any transfer to their parents:

$$W_c = \pi \ln x_c + (1 - \pi) \ln \tau^c \quad (3.8)$$

where x_c is child's consumption when adult, τ^c is the transfer given to the parents, and $\pi \in (0, 1)$.

Household choices are timed as follows. In period 1, parents choose investment on human capital and saving. Period 2 is divided in two subperiods. In the first subperiod, they choose the level of bequests. In the second subperiod, children decide how much to transfer to their parents. Therefore, they face the following budget constraint:

$$x_c + \tau^c = \phi(1 - l) + b \quad (3.9)$$

We solve for the equilibrium by backward induction. For the second subperiod, it is easy to show that children choose the following levels of own consumption and transfers to their parents:

$$x_c = \pi(\phi(1 - l) + b) \quad (3.10)$$

$$\tau^c = (1 - \pi)(\phi(1 - l) + b) \quad (3.11)$$

Since both parents are assumed to be forward looking, parents anticipate the effect that their current decisions have both on child consumption and the transfers received from him. Therefore, the solution to the Nash bargaining problem is given by the solution to

$$Max(x_{1f}^\alpha x_{2f}^\alpha (\phi(1 - l) + b)^{1-\alpha})^\gamma (x_{1m}^\beta x_{2m}^\beta (\phi(1 - l) + b)^{1-\beta})^{1-\gamma} \quad (3.12)$$

subject to the constraints

$$x_{1f} + x_{1m} = a + l - s \quad (3.13)$$

$$x_{2f} + x_{2m} = a + s + (1 - \pi)\phi(1 - l) - \pi b \quad (3.14)$$

Proposition 3.2 *In the model with two-sided altruism, if savings and bequests are interior, then parents invest the optimal level of human capital on the child. Moreover, human capital is a normal good.*

PROOF: See Appendix.

3.3.3 Conditional Cash Transfers: Efficiency vs Equity

We now introduce a social planner whose objective is to help households to invest the optimal amount of human capital h^o on the child. The planner implements the following policy: It provides a transfer $\bar{\tau}$ to all households that invest at least the optimal level of human capital. Formally,

$$\tau = \begin{cases} \bar{\tau} & \text{if } h \geq h^o \\ 0 & \text{otherwise} \end{cases}$$

Let $V(\bar{\tau}, a, l^o)$ denote the indirect utility of a household with income generating ability a if it accepts the conditionality imposed by the policymaker. Similarly, let $V(0, a, l^*(a))$ denote its indirect utility if it does not, where $l^*(a)$ is the optimal choice of child labor for a household that does not participate in the program. Clearly, a household will accept the conditionality if $V(\bar{\tau}, a, l^o) - V(0, a, l^*(a)) > 0$. If child labor is an inferior good, or equivalently, human capital investment is a normal good, it can be shown that this difference is increasing on a , so better off households are more likely to accept the conditionality. Since the opportunity

cost of participating in the program is given by the foregone income coming from child labor, $l^*(a) - l^o$, the effective transfer received by a household with income generating ability a is given by:

$$\tau^e(a) = \begin{cases} \bar{\tau} + l^o - l^*(a) & \text{if } V(\bar{\tau}, a, l^o) > V(0, a, l^*(a)) \\ 0 & \text{otherwise} \end{cases}$$

Clearly, effective transfers $\tau^e(a)$ are non-decreasing on a for households participating in the program.³ Therefore, within this group, better off households tend to receive a larger positive impact from the program.

More generally, we can distinguish three types of households with choices depending on their income generating ability a . The first type of household invests less than the optimal level of human capital h^o even when the CCTS is available, so it does not receive any transfer at all. The second type of household was investing less than the optimal level h^o before the scheme was available, but increases its investment level to h^o once he becomes a beneficiary of the program. Finally, the third type of household was already investing the optimal level of human capital, so it always participate in the program since it represents a pure income transfer to the household.

3.4 The Evaluation Problem

Although randomization helps to answer many of the questions raised by policymakers, there are many other questions that remain unanswered, in particular those related with the distribution of program impacts across the population of beneficiaries.

³Given the assumptions of the model, in particular the concavity of $\phi(\cdot)$, for any level of generating ability a , l^o is a lower bound for the optimal choice of child labor: i.e. $l^*(a) \geq l^o$

To formalize the inferential problem, let each member j of population J be exposed to a mutually exclusive and exhaustive binary set of treatments $T = \{0, 1\}$, and have a response function $y_j(t) : T \rightarrow \mathbb{R}$ mapping treatments into outcomes. The population is a probability space (J, Ω, P) and $y(\cdot) : J \rightarrow \mathbb{R} \times \mathbb{R}$ is a random variable mapping the population into their response functions. Therefore, there exist two potential states of the world for each member j of J : $(y_j(0), y_j(1))$. Lets denote program participation by the indicator variable d_j , where $d_j = 1$ indicates program participation, and $d_j = 0$ otherwise. The analyst observes d_j , but he cannot observe $y_j(0)$ and $y_j(1)$ simultaneously. More formally, he observes $y_j = d_j y_j(1) + (1 - d_j) y_j(0)$. The fact that one cannot observe both outcomes for each individual is known as the evaluation problem.

3.4.1 Average Treatment Effects

Following the traditional approach in the program evaluation literature, the average treatment effect on the treated (ATE) is given by

$$\tau = E[y(1) - y(0) | d = 1] \tag{3.15}$$

Randomization guarantees the identification of ATE since we have $P(y_0 | d = 1) = P(y_0 | d = 0)$. In fact, it turns out that ATE can be consistently estimated under the weaker assumption that d is independent of $y(0)$.⁴

Columns 2 and 3 reports estimated mean outcomes for treatment and control samples, respectively. The first two rows concern total per capital expenditure and total per capita purchase of food items. The fourth column provides average treatment effects of the program on each of these variables. These results show

⁴To see this point, decompose the difference $E[y(1)] - E[y(0)]$ as follows $E[y | d = 1] - E[y | d = 0] = E[Y(0) | d = 1] - E[y(0) | d = 0] + \tau = \tau$.

that the effect of Progresa on total monthly per capita expenditure was about 26 pesos (a 15% mean effect), while its ATE on total monthly per capita food purchase was about 20 pesos. These treatment effects are statistically significant at the 1% level.

Table 3.1: Per Capita Mean Outcomes and Treatment Effects

Outcome	Treatment (n=6946)	Control (m=4098)	τ	95% CI
Total Expenditure	202.815	176.341	26.474 (3.410)	[20.731,32.217]
Food Purchase	148.385	128.4156	19.970 (2.028)	[16.149,23.790]

From the discussion on the program evaluation problem, we know that the identification of the joint distribution $P(y_1, y_0)$ is, in general, not possible. There is a case, however, where one can identify the distribution of program impacts $P(y_1 - y_0)$. The dummy-endogenous-variable model (Heckman 1978) assumes that

$$y_j(1) = y_j(0) + \tau$$

Defining τ as the treatment effect, this assumption implies homogeneous treatment responses. Therefore, the distribution of program impacts is the Dirac measure at τ :

$$P(y(1) - y(0) \mid d = 1) = P(\tau \mid d = 1) \tag{3.16}$$

Under random treatment selection, we have $\tau = E[y(1) - y(0) \mid d = 1]$, which is identified. Hence, the dummy-endogenous variable model identifies the distribution of program impacts.

3.4.2 Fréchet Space

Because of the evaluation problem, one cannot observe an individual's outcome in both treatment and control states. Therefore, it is not possible to identify the distribution of program impacts without imposing more structure on the problem at hand. However, we may be able to partially identify some features of the distribution of the random vector $(y(0), y(1))$ when $P(y(1))$ and $P(y(0))$ are identified.⁵

Let us introduce the following notation. H denotes the bi-dimensional cumulative distribution function of the random vector $(y(0), y(1))$, where $H(\mathbf{t}) = P(y_0 \leq t_1, y_1 \leq t_2)$, with $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$. \mathcal{H} denotes the *Fréchet Space* given the marginals, that is $\mathcal{H}(F_0, F_1)$ is the space of all cumulative distribution functions $H(\mathbf{t})$ on \mathbb{R}^2 with fixed marginal cumulative distribution functions $F_0(t_1) = P(y_0 \leq t_1)$ and $F_1(t_2) = P(y_1 \leq t_2)$. We denote by E_H the expectation operator under the joint distribution H .

Fréchet (1951) showed that the distribution $H(x_1, x_2)$ belongs to $\mathcal{H}(F_0, F_1)$ if and only if

$$H_-(t_1, t_2) \leq H(t_1, t_2) \leq H_+(t_1, t_2) \quad (3.17)$$

for all $(t_1, t_2) \in \mathbb{R}^2$, where

$$H_-(t_1, t_2) = \max\{F_1(t_2) + F_0(t_1) - 1, 0\} \quad (3.18)$$

$$H_+(t_1, t_2) = \min\{F_0(t_1), F_1(t_2)\} \quad (3.19)$$

More recently, Ruschendorf (1981) showed that these bounds are sharp.

Tchen (1980) has established a result that will be proved to be very useful for the purposes of the present analysis. This result states that nonnegative and convex functions are monotone on the *Fréchet Space*:

⁵For a review of the partial identification approach see Manski (2003).

Lemma 3.1 (Tchen 1980) For any convex nonnegative function ψ defined on \mathbb{R} ,

$$E_H\psi(y_1 - y_0) \in [E_{H_+}\psi(y_1 - y_0), E_{H_-}\psi(y_1 - y_0)]$$

for all $H \in \mathcal{H}(F_1, F_2)$.

3.4.3 Partial Identification of Mobility Treatment Effects

Because of the evaluation problem, many distributional scenarios are consistent with the data at hand. Could it be possible to "measure" this multiplicity of scenarios through some statistic? In this section we provide a way to do it by applying the same kind of logic we can find in studies of economic mobility.

While the goal of analyzing treatment effects is to predict the outcomes that would occur if different treatment rules were applied to the population (Manski 2003), the study of economic mobility centers on quantifying the movement of the units of analysis through the distribution of economic well-being over time. More precisely, research on economic mobility tries to connect past and present, "establishing how dependent one's current economic position is on one's past position..." (Fields 2001). In this sense, the analysis of economic mobility does not have to face the evaluation problem since both states, past and present, are observed in principle.

Suppose for a moment that we were able to identify counterfactual outcomes for two individuals. One of the individuals experiences a program impact of +100, the other individual experiences a decrease in the outcome of interest of -100. Keeping everything else constant, how much outcome *movement* has taken place? The standard approach to answer this question is to estimate the average treatment effect, so the net effect of the treatment is zero. After this simple exercise,

one is left with the feeling that overall the treatment effect has been totally neutral. However, the fact that the two individuals considered in this simple example registered changes in their outcomes implies that the treatment is not neutral at all.

Fields and Ok (1996) define a measure of mobility that considers symmetric income movements⁶ as $\int |w - z| dH(w, z)$, where w_i and z_i are the incomes of individual i at two different points on time. We can extend this measure to the context of program evaluation by redefining these variables, so our measure of mobility treatment effects would be given by

$$m = \int |y_1 - y_0| dH(y_1, y_0) \quad (3.20)$$

In contrast to mobility analysis, when analyzing treatment effects one has no information on counterfactual outcomes for the treated population, so we cannot identify this measure. However, we can partially identify m since the absolute value function is convex and positive (Lemma 3.1).

One complication arises since most data sets, and the Progres data set is not the exception, have unbalanced sample sizes, that is to say, the number of observations in the treatment group is not the same than the number of observations in the control group. We circumvent this problem by using quantiles of the empirical distributions \hat{F}_0 and \hat{F}_1 ⁷. Table presents some estimations of m based on 100, 500, and 900 quantiles. The bootstrap confidence intervals were estimated using 2000 bootstrap replications. The ratio m_{H_+}/m_{H_-} is around six to one, which indicates that a great number of distributional scenarios are compatible with the data at

⁶Symmetric outcome movement arises when individuals' outcomes change from one state to another and one is concerned about the *magnitude* of these fluctuations but not their direction.

⁷The applied algorithm is described in more detail in Appendix B.

hand. Because of the evaluation problem, one cannot discard the possibility of

Table 3.2: Mobility Treatment Effects

Number of Quantiles	$[m_{H_+}, m_{H_-}]$	95% Normal CI	90% Percentile CI
100	[25.719, 147.524]	[21.167, 151.148]	[22.280, 152.076]
500	[26.461, 156.838]	[22.156, 161.246]	[23.030, 160.954]
900	[26.004, 158.575]	[21.609, 163.339]	[23.178, 162.985]

having an important subset of the treated population receiving negative treatment effects when ATE are strictly positive. Let $\mathcal{L} = \{j \in J : y_j(1) < y_j(0)\}$ denote the set of members of population J that register a loss as a result of participating in the program. Although it is not possible to identify the set of individuals who belong to this set in general, we can partially identify a parameter that may shed some light on the potential negative effects of being exposed to the treatment, at least in average sense.⁸ We define the average loss of participating in the program as follows⁹

$$L_H = \int \mathbf{1}_{\mathcal{L}}(y(1) - y(0))dH \quad (3.21)$$

$$= \int \min(y(1) - y(0), 0)dH \quad (3.22)$$

Lemma 3.2 *Sharp bounds on L are given by $[L_{H_-}, L_{H_+}]$.*

PROOF: See Appendix.

Table 3 presents some estimations of L based on 100, 500, and 900 quantiles (the bootstrap confidence intervals were also estimated using 2000 bootstrap replications). Even though these worst case bounds may seem exaggerated at a first

⁸Notice that these are worst case bounds. Monotonicity assumption motivated my program design and economic theory can be proved to be very helpful to improve inference.

⁹ $\mathbf{1}_{\mathcal{L}}$ is the indicator function which is equal to one if $j \in \mathcal{L}$, and 0 otherwise.

Table 3.3: Average Loss

Number of Quantiles	$[L_{H_-}, L_{H_+}]$	95% Normal CI	90% Percentile CI
100	$[-61.589, 0.000]$	$[-64.752, 0.000]$	$[-64.326, 0.000]$
500	$[-66.062, -.724]$	$[-69.784, 0.000]$	$[-69.083, -.004]$
900	$[-66.607, -.317]$	$[-70.356, 0.000]$	$[-69.947, -.020]$

sight, especially the lower bound, they are a reminder that ATE may be missing a lot of relevant information. This empirically corroborates the fact that the evaluation problem generally implies the existence of multiple distributional scenarios consistent with the data generating process.

3.5 Testing for Homogeneity in Program Impacts

In this section, we apply a partial identification approach that will allow us to develop simple tests to evaluate the hypothesis of homogeneous treatment effects on the treated.

Consider testing

$$H_0 : y(1) - y(0) = c \text{ a.s.}$$

versus

$$H_1 : y(1) - y(0) \neq c \text{ a.s.}$$

For some real number $c = E(y_1) - E(y_0)$.

Define the functional

$$\Phi(F_0, F_1) = \int \psi(y(1) - y(0)) dH_+ - \psi\left(\int y(1) dF_1 - \int y(0) dF_0\right) \quad (3.23)$$

where $\psi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}_+$ belongs to the class of nonnegative and strictly convex real valued functions. The following result will be proved to be very helpful for testing the hypothesis of homogeneous program impacts:

Proposition 3.3 *Let $\psi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}_+$ be any nonnegative and strictly convex real valued function. If $\Phi(F_0, F_1) > 0$, then $y(1) - y(0) \neq c$ a.s.*

PROOF: See Appendix.

Therefore, we could test the hypothesis of homogeneity in program impacts through testing the hypothesis $H_0 : \Phi(F_0, F_1) = 0$. As an example, let \mathcal{W}_α denote the family of functionals defined by

$$\{\Phi_\alpha(F_0, F_1) : \Phi_\alpha = \int (|y(1) - y(0)|^\alpha dH_+ - |\int y(1)dF_1 - \int y(0)dF_0|^\alpha, \alpha \geq 2)\}$$

It can be shown that $\psi(x) = |x|^\alpha$ is a strictly convex function¹⁰ for $\alpha \geq 2$ (See Appendix). Therefore, $\Phi_\alpha(F_0, F_1) > 0$ implies $y(1) - y(0) \neq c$ a.s.

From here, we can derive an indirect way of testing the null hypothesis by statistically comparing the hypothesis

$$H_0 : \Phi_\alpha(F_0, F_1) = 0$$

versus

$$H_1 : \Phi_\alpha(F_0, F_1) \neq 0$$

Corollary 3.1 *The hypothesis of homogeneous treatment effects can be rejected if*

$$Var(Y(1)) \neq Var(Y(0))$$

PROOF: See Appendix.

Corollary 3.1 can be proved to be very helpful if we impose more structure on the problem. Let $Y_i(1) \sim N(\mu_1, \sigma_1^2)$ and $Y_j(0) \sim N(\mu_0, \sigma_0^2)$, $i = 1, \dots, n$, $j = 1, \dots, m$, be two independent random samples. Notice that

¹⁰Notice that $Var_{H_+}(Y(0) - Y(1))$ is a member of \mathcal{W}_α since $Var_{H_+}(Y(0) - Y(1)) = \Phi_2$.

$$\frac{S_1^2/\sigma_1^2}{S_0^2/\sigma_0^2} \sim F_{n-1,m-1}$$

where S_i^2 , $i = 0, 1$, is the sample variance, and $F_{n-1,m-1}$ is the F distribution with $n - 1$ and $m - 1$ degrees of freedom. Therefore, if the populations are normally distributed, we can test H_0 by statistically testing the hypothesis $\frac{\sigma_1}{\sigma_0} = 1$.

Table 3.4: F test for $H_0 : \frac{\sigma_1}{\sigma_0} = 1$

S_0^2	S_1^2	$f = \frac{S_1^2}{S_0^2}$	n	m	$P(F_{n-1,m-1} < f)$
18791.13	27779.93	1.478	6946	4098	1

From table 3.4, it is clear that we can reject the hypothesis that both populations share the same standard deviation. Consequently, the hypothesis of homogeneity in program impacts is also rejected. However, this test is not accurate unless the distributions of the populations are close to normal.¹¹

We also apply other tests for equality of variances that are less sensitive to departures from normality. Levene's test (1960) tends to be more robust than the F test when the distribution is not Gaussian. Brown and Forsythe (1974) extended Levene's test to use either the median or the trimmed mean instead of the mean. Let W_0 , $W_{.50}$, and $W_{.10}$ denote, respectively, the original Levene's statistic, the Levene's statistic replacing the mean by the median, and the Levene's statistic with a 10% trimmed mean.¹² All of these tests reject the null hypothesis at the 1% level (see table 3.5).¹³

¹¹A skewness and kurtosis test rejects the hypothesis of normality. In fact, the hypothesis of symmetry can also be easily rejected.

¹²Brown and Forsythe(1974) reached the conclusion that using the trimmed mean performed best when the underlying data followed a Cauchy distribution (i.e., heavy-tailed) and the median performed best when the underlying data followed a (i.e., skewed) distribution. Using the mean provided the best power for symmetric, moderate-tailed, distributions

¹³The Levene's test rejects the null hypothesis if $W > F_{1,n+m-2}$ where $F_{1,n+m-2}$

Table 3.5: Levene's statistics

W_0	$W_{.10}$	$W_{.50}$
9.003	7.796	8.019

Another alternative is to use bootstrap methods to estimate the distribution of the statistic $\hat{\Phi}$ by resampling the evaluation samples (Efron 1979). Let $P(\hat{\Phi} - \Phi | F_0, F_1)$ denote the exact, finite sample distribution of $\hat{\Phi} - \Phi$. Using standard notation from the bootstrap literature, let $\hat{\Phi}^* - \hat{\Phi}$ be computed from observations obtained according to the empirical distributions \hat{F}_0 and \hat{F}_1 in the same way $\hat{\Phi} - \Phi$ is computed from the true observations $Y_i(1) \sim F_1$ and $Y_j(0) \sim F_0$, $i = 1, \dots, n$, $j = 1, \dots, m$. Finally, let Q_β^* denote the β -quantile of the CDF of $\hat{\Phi}^* - \hat{\Phi}$. That is

$$Q_\beta^* = \inf\{\hat{\Phi}^* : P(\hat{\Phi}^* - \hat{\Phi} | \hat{F}_0, \hat{F}_1) \geq \beta\} \quad (3.24)$$

A commonly applied method to test H_0 is to assume that $\hat{\Phi} - \Phi$ is normally distributed, and then to use the bootstrap estimate of the standard deviation as an approximate estimator of the true sample variance. That is

$$\frac{\hat{\Phi} - \Phi}{\hat{\sigma}^*} \sim N(0, 1) \quad (3.25)$$

However, under the null, Φ is at the boundary of the parameter space since $\Phi \geq 0$. This implies that the random quantity $\hat{\Phi}/\hat{\sigma}^*$ is always positive, and hence it cannot be normally distributed with mean zero and variance one.

One possible solution for this problem is to follow Efron (1987) by assuming the existence of a monotone increasing transformation $\varphi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\varphi(\hat{\Phi}) - \varphi(\Phi) \sim N(z_\Phi, \sigma_\Phi^2) \quad (3.26)$$

is the upper critical value of the F distribution for some predetermined significance level.

for every choice of Φ (z_Φ is known as the bias correction term). For the purpose of the present study, we can weaken this assumption by requiring just symmetry for the distribution of $\varphi(\hat{\Phi}) - \varphi(\Phi)$.

Proposition 3.4 *Suppose there exists a strictly increasing function $\varphi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ such that*

$$\varphi(\hat{\Phi}) - \varphi(\Phi) \mid F_0, F_1 \sim V$$

$$\varphi(\hat{\Phi}^*) - \varphi(\hat{\Phi}) \mid \hat{F}_0, \hat{F}_1 \sim V$$

where V is continuously and symmetrically distributed about $\gamma \in \mathbb{R}$, satisfying

$$F_V(2\gamma + F^{-1}(\beta)) > 0$$

for some $\beta \in (0, 1/2)$. Then

$$\text{Reject } H_0 \text{ if } \min \hat{\Phi}^* > 0$$

is a level β test.

PROOF: See Appendix.

We estimate the bootstrap cdf of $\hat{\Phi}_2^*$ using $B = 2000$ bootstrap replications. From table 3.6, it can be inferred that the null hypothesis of homogeneous treatment effects can be easily rejected under the assumptions of the proposition. For instance, if $V \sim N(-\sigma\gamma, \sigma^2)$, we have

$$\begin{aligned} P(\varphi(\hat{\Psi}^*) - \varphi(\hat{\Psi})) &= P(\sigma Z - \sigma\gamma < 0) \\ &= P(Z < \gamma) \\ &= F_Z(\gamma) \end{aligned}$$

A plug in estimator for γ is therefore given by

$$\begin{aligned}\hat{\gamma} &= F_Z^{-1} \left(\frac{\#\{\varphi(\hat{\Phi}^*) < \varphi(\hat{\Phi})\}}{B} \right) \\ &= F_Z^{-1} \left(\frac{\#\{\hat{\Phi}^* < \hat{\Phi}\}}{B} \right)\end{aligned}$$

where the last equality follows from the monotonicity of $\varphi(\cdot)$. As expected, the sign of this parameter is strictly negative, taking values in the range $(F_Z^{-1}(.25), F_Z^{-1}(.50))$ for $\#q \in \{100, 300, 600, 800, 1000\}$, where $\#q$ indicates the number of quantiles used in the estimation. Therefore, we can reject the null hypothesis at the 1% level under the assumption of normality.

Andrews (2000) argues that the bootstrap may not be consistent when the parameter of interest is on a boundary of the parameter space. One possible solution is to draw subsamples of size $k < \min(n, m)$ from the original data with replacement. This sampling method is identical to the standard bootstrap in every aspect, but the size of each replication.¹⁴ Another possible advantage of this method is that we can estimate the bootstrap distribution of $\hat{\Phi}^*$ without using the quantiles of the empirical distributions as the original data. Table 3.7 presents several quantiles of the bootstrap distribution of $\hat{\Phi}_2^*$ for $k \in \{1000, 2000, 3000, 3500, 4000\}$. Under the assumption of normality and bias correction, the hypothesis of homogeneity in program impacts can be rejected at the 1% level.

¹⁴Bickel et al (1997) discuss a number of resampling schemes under which the size of the sample replication is smaller than the original sample size. They argue that the k out of n sampling scheme works very well in all known realistic examples of bootstrap failure.

Table 3.6: Summary statistics for the bootstrap distribution of $\hat{\Phi}_2^*$ using empirical quantiles of F_0 and F_1 .

Number of Quantiles	$Q_{.01}^*$	$Q_{.05}^*$	$Q_{.10}^*$	$Q_{.50}^*$	Mean	sd	Min $\hat{\Phi}_2^*$	$\hat{\Phi}_2$
100	36.481	54.2097	67.888	150.27	178.528	113.198	15.455	88.487
300	53.022	80.851	101.669	285.202	474.225	519.634	30.183	95.315
600	78.661	122.863	162.053	502.72	678.995	592.624	39.900	405.422
800	96.059	150.736	215.238	580.472	809.775	949.181	57.945	309.866
1000	109.783	190.4319	258.657	681.292	1060	1523.616	50.742	346.014

Table 3.7: Summary statistics for the bootstrap distribution of $\hat{\Phi}_2^*$ using a $k/\min(n, m)$ resampling scheme.

k	$Q_{.01}^*$	$Q_{.05}^*$	$Q_{.10}^*$	$Q_{.50}^*$	Mean	sd	Min $\hat{\Phi}_2^*$
1000	175.041	371.705	558.809	2206.117	5502.247	6903.977	84.790
2000	239.864	411.428	556.666	3621.683	4463.259	4163.627	105.5011
3000	282.169	476.443	650.213	3327.524	4048.062	3188.16	131.8912
3500	299.809	482.117	680.626	3265.586	3951.013	2970.799	152.982
4000	302.032	486.642	754.739	3254.25	3835.516	2728.157	102.811

3.6 Identification of Program Impacts under Monotonicity Assumptions

The bounds implied by the Fréchet Space of bivariate distributions proved to be very helpful for developing a test for homogeneity of program impacts. However, without further assumptions, it is an impossible task to pin down the actual distribution of treatment effects even in the case of a random experiment.

Inference on the distribution of program impacts may be improved by imposing assumptions implied by economic theory or any other mechanism related with the data generating process such as program design. Manski (1997) investigates what may be learned about treatment response under the assumptions of monotone, semi-monotone, and concave-monotone response functions. He shows that these assumptions have identifying power, particularly when compared to a situation where no prior information exists (worst case bounds). Typically, the type of monotonicity assumptions applied by econometricians dealing with partially identified parameters take some form of stochastic dominance. For instance, in a missing treatments environment, Molinari (2005b) shows that one can extract information from the observations for which treatment data are missing using monotonicity assumptions. Specifically, one could assume that the effect of a social program on the outcome of interest cannot be negative. This is equivalent to assume that for each j in J we have

$$\tau_j = \max\{y_j(1) - y_j(0), 0\} \tag{3.27}$$

Given the design of Progresa, this seems to be a reasonable assumption. One can expect a positive effect on the outcome of interest, in our case consumption, for treated individuals. Moreover, this type of assumption implies first order stochastic

dominance (FSD) of distribution $P(y(1))$ over distribution $P(y(0))$: i.e. $P(y_1 \leq x) \leq P(y_0 \leq x)$ for all $x \in \mathbb{R}$. Actually, this assumption is stronger than FSD. It implies that, for all $t \in \mathbb{R}$, we have

$$P(y(1) \geq t \mid y(0) = t) = 1 \quad (3.28)$$

Notice that the converse is not always true, that is to say, stochastic dominance does not necessarily imply monotonicity.¹⁵

From section 3.3, we know that if human capital is a normal good, a policy-maker implementing a CCTS faces a dilemma: on the one hand, it may represent a very helpful policy tool for achieving an efficient level of human capital. On the other hand, this policy instrument may be at odds with a more equal distribution of *effective* benefits. In particular, it was argued that when the conditioned-on good is normal, better-off households tend to receive larger "effective" benefits, once the opportunity cost of foregone earnings from child labor is deducted. Unfortunately, because of the evaluation problem, we cannot test this hypothesis without imposing more assumptions. We circumvent this problem by establishing a different type of monotonicity assumption, one that will allow us to test the hypothesis of regressivity in program impacts.

We assume the existence of a non-decreasing real valued function $\phi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$y_j(1) = \phi(y_j(0)) \quad (3.29)$$

Notice that function $\phi(\cdot)$ is not indexed, and in consequence this assumption implies *rank preservation* among the members of population J . More precisely, in the

¹⁵To see that, just observe the random vector $(y(0), y(1))$ whose support consists of two points: (1,3) and (2,1). Clearly $y(1)$ stochastically dominates $y(0)$, but the monotonicity assumption is violated.

contest of program evaluation, rank preservation means that, for some outcome of interest Y , the rank of a particular unit of observation i with respect to any other observation j is the same in both treatment and control states. More formally, rank preservation implies that, for any two members i and j of population J , the following relation holds:

$$(y_i(0), y_i(1)) \geq (y_j(0), y_j(1))$$

This untestable assumption is also a necessary condition for the existence of regressive program impacts¹⁶.

Let $\tau(y(0)) = \phi(y(0)) - y(0)$. We say that there is regressivity in program impacts whenever $\tau(y(0))$ is a non-decreasing and non-trivial function of $y(0)$, that is, for any $i, j \in J$, such that $y_i(0) > y_j(0)$, we have¹⁷

$$\frac{\phi(y_i(0)) - \phi(y_j(0))}{y_i(0) - y_j(0)} \geq 1 \tag{3.30}$$

In order to make this result operational, and to test the hypothesis of regressivity in program impacts, we will use quantile treatment effects¹⁸(QTE), which are a natural extension of rank preservation to the analysis of distribution of treatment effects. Let us introduce this concept more formally. The q^{th} quantile of distribution $F_i(y)$, $i = 0, 1$ is defined as:

$$y_i(q) = \inf\{y : F_i(y) \geq q\}$$

¹⁶Let $(y_i(0), y_i(1))$ and $(y_j(0), y_j(1))$ be the outcomes in both states for $i, j \in J$. Without loss of generality, let $y_i(0) > y_j(0)$. Regressivity in program impacts is equivalent to $y_i(1) - y_i(0) > y_j(1) - y_j(0)$, which implies $y_i(1) > y_j(1)$

¹⁷More precisely, there is regressivity in program impacts if $\tau(y(0))$ is a non-decreasing and non-trivial function almost everywhere.

¹⁸See Koenker and Bassett (1978) for an application of quantile estimation to a regression setting. Abadie, Angrist, and Imbens (2002) extend their idea to the estimation of quantile treatment effects. See Appendix C for a description of the QTE estimator.

The following result will be proved to be useful for the empirical application. It shows the existence of the function $\phi(\cdot)$ under some mild continuity assumption:

Lemma 3.3 *If $y_0(q)$ is a continuity point of F_0 , then there exists a non-decreasing function $\phi(q) : (0, 1) \rightarrow \mathbb{R}$ such that $y_1(q) = \phi(y_0(q))$; moreover, there exists a unique function $\tau(q)$ satisfying $\tau(q) = y_1(q) - y_0(q)$.*

PROOF: See Appendix.

The QTE for quantile q can be defined as the difference in treatment status between quantile q of treatment group and quantile q of control group. Formally, QTE for quantile q is given by

$$\tau(q) = y_1(q) - y_0(q)$$

Therefore, QTE represent an alternative way for testing for regressivity in program impacts. A non decreasing and non-trivial QTE function is strong evidence for regressive program impacts under the assumption of rank preservation, where for rank preservation we mean rank preservation in terms of quantiles of the distributions F_1 and F_0 .

Tables 3.8 and 3.9 introduce the QTE estimator for per capita total expenditures and per capita food purchase, respectively, for several quantiles. These quantiles were estimated simultaneously, so statistical comparisons can be made among them. Empirical variance of QTE was calculated by means of 200 bootstrap replications of the quantile treatment effect.

We plot these QTE in figures 3.1 and 3.2. For comparison purposes, we plot the average treatment effect as a horizontal dashed line. Dotted lines surrounding the ATE line represent a 95% confidence interval. Clearly, the variation of the treatment effects across the different quantiles is both economically and statisti-

Table 3.8: Quantile Treatment Effects: Total Expenditure

q	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
$\tau(q)$	9.487 (1.870)	13.800 (1.515)	17.194 (0.985)	18.304 (1.355)	20.250 (1.693)	22.608 (1.602)	24.141 (1.636)	23.824 (1.765)	24.590 (1.597)	24.815 (2.141)
q	.55	.60	.65	.70	.75	.80	.85	.90	.95	
$\tau(q)$	25.708 (2.120)	25.722 (2.337)	27.984 (2.499)	27.821 (2.813)	28.564 (3.077)	28.687 (3.411)	33.174 (4.533)	40.471 (6.816)	47.476 (11.740)	

Table 3.9: Quantile Treatment Effects: Food Purchase

q	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
$\tau(q)$	4.728 (1.430)	9.630 (1.187)	11.134 (1.014)	11.825 (1.228)	13.698 (1.265)	15.065 (1.055)	15.832 (1.118)	16.935 (1.215)	18.117 (1.379)	19.129 (1.203)
q	.55	.60	.65	.70	.75	.80	.85	.90	.95	
$\tau(q)$	20.029 (1.397)	20.966 (1.273)	22.823 (1.637)	22.946 (2.012)	26.071 (2.117)	27.286 (2.266)	28.504 (2.678)	33.757 (3.581)	39.332 (6.754)	

cally significant, particularly at the extremes of the QTE plot, although for a broad band treatment effects are statistically homogeneous.

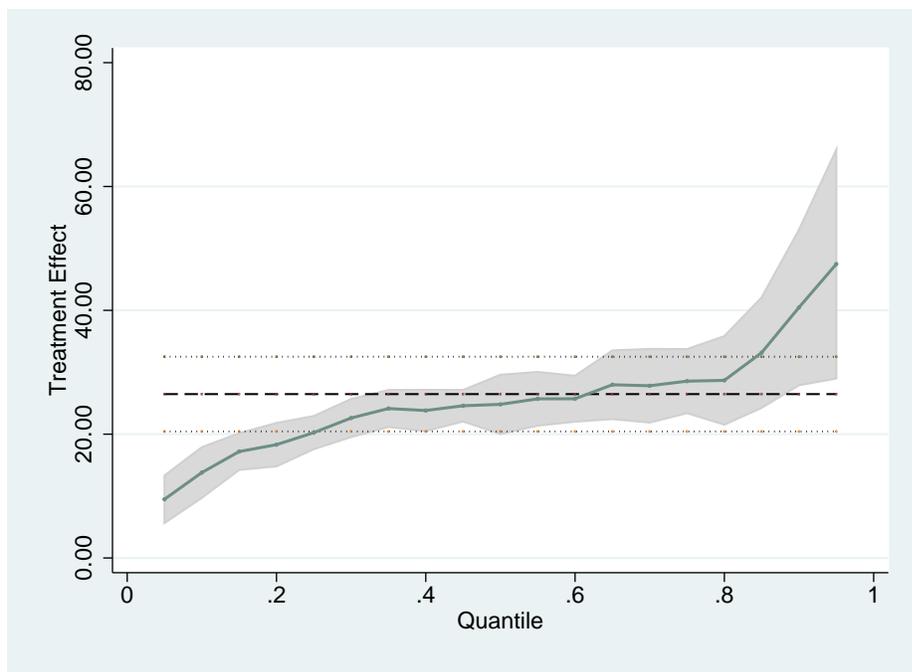


Figure 3.1: Quantile Treatment Effects: Total Expenditure

The QTE estimators are consistent with the predictions of the theoretical model: better off households tend to receive larger positive impacts from the program which generates the monotonically increasing shape of the QTE. This characteristic of the QTE for PROGRESA is more remarkable when one contrast que treatment effect between the lower 20 and the upper 20 centiles. For instance, in the case of total expenditure, the treatment effect for the 95th centile is about five times the treatment effect estimated for the 5th centile. This gap is about 10 times between the same quantiles in the case of food purchase.

QTE estimation also represents an alternative method to test for homogeneity in program impacts under some mild regularity conditions. This assertion is formalized in the following Lemma:

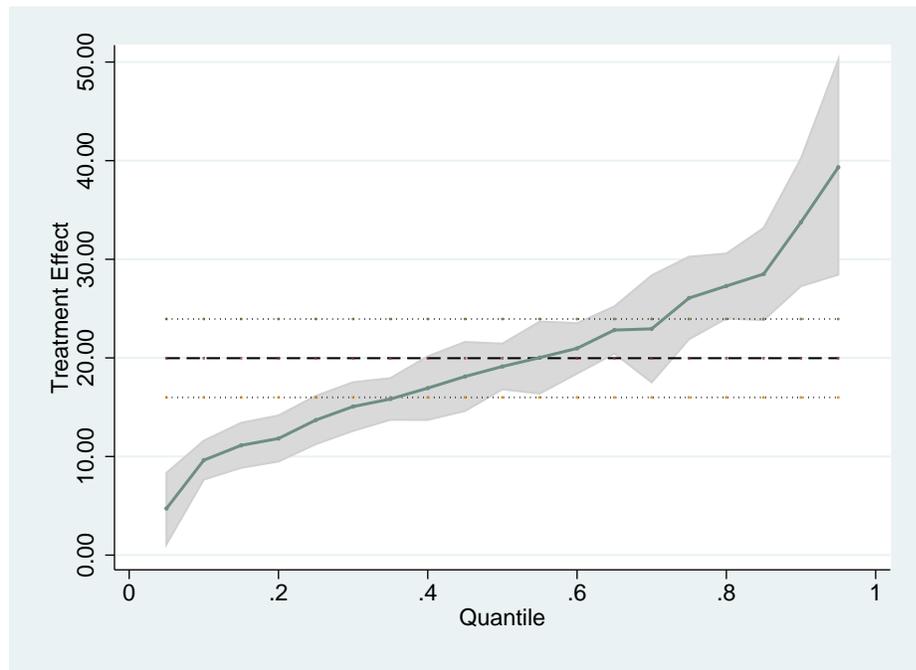


Figure 3.2: Quantile Treatment Effects: Food Purchase

Lemma 3.4 *If treatment effects are homogeneous across the population, then $\tau(q) = \tau$ for all $q \in [0, 1]$ such that $y_0(q)$ is a continuity point of F_0 .*

PROOF: See Appendix.

From tables 3.7 and 3.8, we can conclude that the hypothesis of homogeneity in program impacts can be rejected under the conditions of Lemma 3.4.

3.7 Conclusions

Conditional cash transfers represent an important policy tool for fighting poverty, particularly when there is some type of externality that prevents the poor from reaching more efficient equilibria. Human capital investment is just one example of an activity generating positive externalities. Correcting for these externalities is then an important step to break the circle of intergenerational poverty.

There are some issues, however, that should be considered by policymakers

implementing this type of programs. If the conditioned-on good is normal, then it is very likely that the distributional effects of the program will be far from being distributionally neutral. In fact, as we saw in the empirical analysis, heterogeneous treatment effects are pervasive, at least in the case of the Progresa evaluation sample.

Under the assumption of rank preservation, program impacts tend to be distributionally regressive for the population participating in Progresa. As it was argued in the text, this finding is also consistent with the fact that the conditioned-on good is normal. Therefore, if the assumption of rank preservation is correct, the poorest of the poor may not be receiving as much benefits as policymakers believe they are. This has important implications for the design of antipoverty policies: policymakers should consider the existent tradeoff between equity and efficiency of outcomes in order to better understand the consequences and limitations of CCTS like Progresa. The final answer will much depend on the benefits and costs of improving the targeting efficiency of a program.

3.8 Appendix A: Proofs and Derivations

Proof of Proposition 3.1: The household bargaining model is solved through the program

$$\begin{aligned} \mathcal{L} = & (x_{1f}^\alpha x_{2f}^\alpha (\phi(1-l) + b)^{1-\alpha})^\gamma (x_{1m}^\beta x_{2m}^\beta (h(1-l) + b)^{1-\beta})^{1-\gamma} + \\ & \lambda_1(a + l - s - x_{1f} - x_{1m}) + \\ & \lambda_2(a + s + (1 - \pi)\phi(1-l) - \pi b - x_{2f} - x_{2m}) + \\ & \lambda_3 s + \lambda_4 b \end{aligned}$$

from where we can obtain the following first order conditions:

$$\begin{aligned} \frac{z_1}{x_{1f}} &= \lambda_1 \\ \frac{z_1}{x_{2f}} &= \lambda_2 \\ \frac{z_3}{x_{1m}} &= \lambda_1 \\ \frac{z_4}{x_{2m}} &= \lambda_2 \\ \frac{-z_2\phi'(1-l)}{\phi(1-l) + b} + \lambda_1 - \lambda_2(1-\pi)\phi'(1-l) &= 0 \\ \frac{z_2}{\phi(1-l) + b} - \lambda_2\pi + \lambda_4 &= 0 \\ -\lambda_1 + \lambda_2 + \lambda_3 &= 0 \end{aligned}$$

where $z_1 = \alpha\gamma$, $z_2 = (1 - \alpha)\gamma + (1 - \beta)(1 - \gamma)$, and $z_3 = \beta(1 - \gamma)$. From the first order conditions we have

$$\begin{aligned} x_{1m} &= \frac{z_3}{z_1} x_{1f} \\ x_{2m} &= \frac{z_3}{z_1} x_{2f} \\ \frac{z_2\phi'(h)}{\phi(h) + b} &= \frac{z_1}{x_{1f}} \end{aligned}$$

and

$$\phi'(h) \geq 1$$

with the last condition holding with equality if $(b, s) \in R_{++}^2$.

For the second part, assume the household is both savings and bequest constrained, so $b = s = 0$. Assume also that it receives an exogenous transfer of income $\omega > 0$ in period 1. From the first order conditions

$$\frac{\phi'(1-l)}{\phi(1-l)} = \frac{z_1}{z_2 x_{1f}} = \frac{z_3}{z_2 x_{1m}}$$

Since the household is both bequest and savings constrained, we have $l^* < l^o$ and $x_{1f} + x_{1m} = a + \omega + l$. Assume, towards a contradiction, that child labor does not decrease: i.e. $\Delta l \geq 0$. Hence, either x_{1f} or x_{1m} increases. This fact and the condition above together imply an increase in child labor, a contradiction. Therefore, human capital is a normal good. \square

Proof of Proposition 3.2: The proof for the first part of the proposition is along the lines of the case with one-sided altruism. To prove that human capital is a normal good, assume the household receives an exogenous positive transfer in period 1, say $\Delta\omega > 0$. If the household is both saving and bequest constrained, then the budget constraint in period 1 is given by $x_{1m} + x_{1f} = a + \Delta\omega + l$. Assume, towards a contradiction, that child labor increases. From the first order conditions, both x_{1m} and x_{1f} increase in equilibrium. From the first order conditions, we also have:

$$\frac{z_3}{x_{1f}} = \left[z_2 + \frac{\left[z_4 \left(\frac{z_4 + z_1}{z_1} \right) \right] (1 - \pi)}{a + (1 - \pi)} \right] \frac{\phi'(1-l)}{\phi(1-l)}$$

This is clearly a contradiction since the left-hand side of the equation strictly decreases, while the right-hand side increases or remains constant. \square

Lemma 3.5 *Sharp bounds on the correlation coefficient ρ_{Y_0, Y_1} are given by*

$$[\rho_{H_+}^{Y_0, Y_1}, \rho_{H_-}^{Y_0, Y_1}]$$

PROOF:

$$\begin{aligned}\rho_{Y_0, Y_1} &= \frac{\text{Var}(Y_1 - Y_0) - \text{Var}(Y_1) - \text{Var}(Y_0)}{2\sqrt{\text{Var}(Y_0)\text{Var}(Y_1)}} \\ &= \frac{E(Y_1 - Y_0)^2 - (E(Y_1) - E(Y_0))^2 - \text{Var}(Y_1) - \text{Var}(Y_0)}{2\sqrt{\text{Var}(Y_0)\text{Var}(Y_1)}}\end{aligned}$$

Since $\varphi(x) = x^2$ is a convex function, the result follows from Lemma 3.1. Sharpness follows from the fact that H_- and H_+ are sharp bounds on the Fréchet Space. \square

Proof of Lemma 3.2 : Notice that the absolute value of the program impact can be decomposed as follows

$$|y(1) - y(0)| = y(1) - y(0) - 2\min(y(1) - y(0), 0)$$

Taking expectations at both sides of the equality, we have

$$m = E(y(1) - y(0)) - 2L$$

From where

$$L = \frac{E(y(1) - y(0)) - m}{2} \tag{3.31}$$

Since m is positive, the result follows from Lemma 3.1. Alternatively, we can apply Lemma 3.1 directly by noticing that

$$\min(y(1) - y(0), 0) = -\max(y(0) - y(1), 0)$$

and $\varphi(x) = \max(x, 0)$ is a convex function. \square

Proof of Proposition 3.3: By Lemma 3.1 and the Frechet bounds, we have

$$\int \psi(y_1 - y_0)dH - \psi\left(\int y_1 dF_1 - \int y_0 dF_0\right) \geq \Phi(F_0, F_1)$$

for all $H \in \mathcal{H}(F_1, F_0)$. Define a random variable $Z = Y_1 - Y_0$. By Jensen's

inequality

$$\begin{aligned}
\int \psi(z)dH_+ &\geq \psi\left(\int zdH_+\right) \\
&= \psi\left(\int y_1dH_+ - \int y_0dH_+\right) \\
&= \psi\left(\int y_1dF_1 - \int y_0dF_0\right)
\end{aligned}$$

which is equivalent to $\Phi(F_0, F_1) \geq 0$. The result follows by using the fact that Jensen's inequality holds with equality for the case of strictly convex functions if and only if $Y_1 - Y_0$ is a constant with probability 1. \square

Proof of Corollary 3.1: Notice that

$$\begin{aligned}
\Phi_2 &= \text{Var}_{H_+}(Y(1) - Y(0)) \\
&= \sigma_1^2 + \sigma_0^2 - 2\rho_{H_+}\sigma_1\sigma_0 \\
&\geq \sigma_1^2 + \sigma_0^2 - 2\sigma_1\sigma_0 \\
&= (\sigma_1 - \sigma_0)^2
\end{aligned}$$

where ρ_{H_+} is the correlation coefficient evaluated at H_+ . Hence $\sigma_1 \neq \sigma_0$ implies $\Phi_2 > 0$, and the result follows from Proposition 3.3. \square

Proof of Corollary 3.1 when $y(1)$ and $y(0)$ are members of the same location-scale family. Since y_0 and y_1 are members of the same location-scale family, we have

$$y_i = \sigma_i Z + \mu_i$$

for $i = 0, 1$, where $Z \sim f(z)$. Because for any location-scale family it is possible to choose $f(z)$ such that $EZ = 0$ and $EZ^2=1$, without loss of generality, we choose these values for the first and second moment of Z . Notice that the extreme joint distribution H_+ is obtained when there is maximum correlation between y_1 and

y_0 . This occurs when high values of y_1 are "matched" with high values of y_0 . This is equivalent to form the pairs $(\sigma_0 z + \mu_0, \sigma_1 z + \mu_1)$ for all z in the support of Z .

Hence

$$\begin{aligned} \int (y(1) - y(0))^2 dH_+ &= \int [(\sigma_1 - \sigma_0)z + (\mu_1 - \mu_0)]^2 df(z) \\ &= E_Z [((\sigma_1 - \sigma_0)^2 Z^2 + 2(\sigma_1 - \sigma_0)(\mu_1 - \mu_0)Z + (\mu_1 - \mu_0)^2)] \\ &= (\sigma_1 - \sigma_0)^2 + (\mu_1 - \mu_0)^2 \end{aligned}$$

The result follows from Proposition 3.3. \square

Proof of Proposition 3.4: Notice that to test the hypothesis $H_0 : \Phi = 0$ is equivalent to test $H_0^\varphi : \varphi(\Phi) = \varphi(0)$. A level $\beta \in (0, 1/2)$ for the latter hypothesis is given by

$$\text{Reject } H_0^\varphi: \varphi(\Phi) = \varphi(0) \text{ if } V_{1-\beta} - \gamma < \hat{\varphi} - \varphi(0)$$

where $V_\beta = F^{-1}(\beta)$. This a straightforward result since under the null we have

$$P(V_{1-\beta} - \gamma < \hat{\varphi} - \varphi(0)) = \beta$$

I will refer to this test as T_1 for the rest of the proof. Let $G(s) = P(\varphi^* < s)$ be the bootstrap cdf of φ^* . Since $\varphi^* = \hat{\varphi} - \gamma + V$, we have

$$\begin{aligned} G(s) &= P(V < s - \hat{\varphi} + \gamma) \\ &= F_V(s - \hat{\varphi} + \gamma) \end{aligned}$$

with inverse $G^{-1}(\beta) = F_V^{-1}(\beta) + \hat{\varphi} - \gamma$. I claim that the test T_2 defined as

$$\text{Reject } H_0^\varphi \text{ if } \varphi(0) < G^{-1}(F_V(2\gamma + V_\beta))$$

is equivalent to T_1 . This is true since

$$\begin{aligned} G^{-1}(F_V(2\gamma + V_\beta)) &= 2\gamma + V_\beta + \hat{\varphi} - \gamma \\ &= \gamma - V_{1-\beta} + \varphi \end{aligned}$$

It follows that the test T_3

$$\text{Reject } H_0^\varphi \text{ if } \varphi(0) < \min \varphi(\hat{\Phi}^*)$$

is a level β test since, for some $\beta \in (0, 1/2)$

$$\begin{aligned} \varphi(0) &< \min \varphi(\hat{\Phi}^*) \\ &\leq G^{-1}(F_V(2\gamma + V_\beta)) \end{aligned}$$

Finally, let $H(s) = P(\hat{\Psi}^* < s)$ be the bootstrap cdf of $\hat{\Psi}^*$. Since $\varphi(\cdot)$ is a strictly increasing transformation, the quantiles of $\varphi(\hat{\Psi}^*)$ coincide with those of $\hat{\Psi}^*$. Hence, T_3 is equivalent to

$$\text{Reject } H_0 : \Psi = 0 \text{ if } 0 < \min \hat{\Psi}^*$$

since $\min \varphi(\hat{\Psi}^*) = \varphi(\min \hat{\Psi}^*)$. This completes the proof. \square

Lemma 3.6 $\psi(x) = |x|^\alpha$ is a strictly convex function for $\alpha \geq 2$

PROOF: For $\alpha = 2$, the result is immediate since $\psi(x) = x^2$, and $\psi'' > 0$. For $\alpha > 2$, we make use of Pecaric and Dragomir's inequality, which indicates that if $pq(q+p) > 0$, $z_1, z_2 \in \mathbb{R}$, and $\alpha \geq 1$, then

$$\frac{|z_1 + z_2|^\alpha}{p+q} \leq \frac{|z_1|^\alpha}{p} + \frac{|z_2|^\alpha}{q}$$

w.l.g. define $z_1 = \lambda x$, $z_2 = (1-\lambda)y$, $x, y \in \mathbb{R}$, $p = \lambda$, $q = 1-\lambda$, and $\lambda \in (0, 1)$.

Then we have $\lambda(1-\lambda) > 0$, and hence

$$\begin{aligned} |\lambda x + (1-\lambda)y|^\alpha &\leq \frac{|\lambda x|^\alpha}{\lambda} + \frac{|(1-\lambda)y|^\alpha}{1-\lambda} \\ &= \lambda^{\alpha-1} |x|^\alpha + (1-\lambda)^{\alpha-1} |y|^\alpha \\ &< \lambda |x|^\alpha + (1-\lambda) |y|^\alpha \end{aligned}$$

Where I have used the fact that $\lambda^{\alpha-1} < \lambda$ and $(1-\lambda)^{\alpha-1} < (1-\lambda)$, for $\alpha > 2$. \square

Proof of Lemma 3.3: Let $\tau(y_0(q))$ and $\phi(\cdot)$ be defined, respectively, by

$$\tau(y_0(q)) = \inf\{\xi : q \leq F_1(y_0(q) + \xi)\}$$

and

$$\phi(y_0(q)) = F_1^{-1}(F_0(y_0(q)))$$

From the quantile function, we have

$$y_1(q) = F_1^{-1}(q) = \inf\{x : F(y(1) \leq x) > q\}$$

Hence,

$$y_1(q) = \phi(y_0(q)) = F_1^{-1}(q) = \tau(y_0(q)) + y_0(q)$$

The result follows by noticing that $\phi(y_0(q))$ is a non-decreasing function of $y_0(q)$.

For a proof of uniqueness see Doksum (1974).□

Proof of Lemma 3.4: Doksum (1974) shows that if $\tau(x) = \tau$ for x in the support of $y(0)$, then $F_0(x) = F_1(x + \tau)$ for all x . Therefore

$$F_0(y_0(q)) = F_1(y_0(q) + \tau)$$

From the proof of Lemma 3.3, we have

$$y_1(q) = F_1^{-1}(F_0(y_0(q))) = y_0(q) + \tau$$

The result follows.□

3.9 Appendix B: Estimation and Bootstrap Algorithm using the Empirical Quantiles

The objective is to estimate bootstrap confidence intervals for the parameters $\theta_- = E_{H_-}[\phi(y_1 - y_0)]$ and $\theta_+ = E_{H_+}[\phi(y_1 - y_0)]$, for some measurable function $\phi(\cdot)$. The data in this problem consists of two independent random samples drawn $Y_i(1) \sim F_1$ and $Y_j(0) \sim F_0$, $i = 1, \dots, n$, $j = 1, \dots, m$. Let \hat{F}_1 and \hat{F}_0 denote the empirical distribution functions implied by these samples.

1. Estimation of θ_- and θ_+

1) Estimate $b = \lceil \gamma \min\{n, m\} \rceil$ empirical quantiles for F_1 and F_0 , where $\gamma \in (0, 1)$ and $\lceil \cdot \rceil$ is the integer function. More precisely, for each $t \in \{t_1, \dots, t_b\}$, $i = 1, 2$, we estimate

$$q_i^{t_j} = \inf\{x : \hat{F}_i(y(i) \leq x) \geq t_j\} \quad (3.32)$$

2) Let \hat{Q}_1 and \hat{Q}_0 be the empirical distribution function of the quantiles estimated above, that is to say, a distribution placing a probability mass $\frac{1}{b}$ to each of these quantiles:

$$\hat{Q}_i(x) = \frac{1}{b} \sum_{j=1}^b \mathbf{1}(q_i^{t_j} \leq x) \quad (3.33)$$

3) For all $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, define

$$\hat{H}_-(x_1, x_2) = \max\{\hat{Q}_0(x_1) + \hat{Q}_1(x_2) - 1, 0\} \quad (3.34)$$

$$\hat{H}_+(x_1, x_2) = \min\{\hat{Q}_0(x_1), \hat{Q}_1(x_2)\} \quad (3.35)$$

4) Estimate θ using plug-in estimators: $\hat{\theta}_- = \theta(\hat{H}_-)$ and $\hat{\theta}_+ = \theta(\hat{H}_+)$.

2. Estimation of the Extreme distributions H_- and H_+

Define the sequences of quantiles of F_1 and F_0 , respectively, by $\{q_1^{t_j}\}$ and $\{q_0^{t_j}\}$. Let $\mu_i = E_{Q_i}[q_i^t]$ denote the expected value of the chosen quantiles under probability measure Q_i . The correlation coefficient between q_1^t and q_0^t is given by

$$\rho(q_0^t, q_1^t) = \frac{1}{b} \sum (q_1^{t_j} - \mu_1)(q_0^{t_j} - \mu_0) \quad (3.36)$$

By Lemma 3.5, we know that this coefficient is at its minimum when it is evaluated at H_+ , and is at its maximum when evaluated at H_- . We can estimate the extreme distributions H_- and H_+ by applying the following result:

Lemma 3.7 (*Hardy, Littlewood, and Polya 1952*) *The sum of products $\sum_i x_i y_i$ is a maximum when both $\{x_i\}$ and $\{y_i\}$ are increasing, and a minimum when one is increasing and the other is decreasing.*

Therefore, by defining $x_j = (q_1^{t_j} - \mu_1)$ and $y_j = (q_0^{t_j} - \mu_0)$, it follows that H_+ is obtained by pairing the largest quantile of F_1 with the largest quantile of F_0 , the second largest quantile of F_1 with the second largest quantile of F_0 , and so on. To construct H_- , we just need to pair the largest quantile of F_1 with the smallest quantile of F_0 , the second largest quantile of F_1 , with the second smallest quantile of F_0 , and so on.

3. Bootstrap

5) Generate bootstrap random samples from \hat{F}_1 and \hat{F}_0 : $Y_i^*(1) \sim F_1$ and $Y_j^*(0) \sim F_0$, $i = 1, \dots, n$, $j = 1, \dots, m$.

Let F_1^* and F_0^* denote the empirical distributions implied by the bootstrap random samples. That is

$$F_i^*(x) = \frac{1}{b} \sum_j \mathbf{1}(y_{ij} \leq x) \quad (3.37)$$

6) Replicate steps 1)-4) above for the bootstrap distributions F_1^* and F_0^* . That is:

6a) Estimate b empirical quantiles for F_1^* and F_0^*

$$q_i^{t_j^*} = \inf\{x : F_i^*(y^*(i)) \leq x\} \geq t_j\} \quad (3.38)$$

6b) Let Q_1^* and Q_0^* be the empirical distribution of the quantiles estimated above, that is to say, a distribution placing a probability mass $\frac{1}{b}$ to each of these quantiles.

$$Q_i^*(x) = \frac{1}{b} \sum_{j=1}^b \mathbf{1}(q_i^{t_j^*} \leq x) \quad (3.39)$$

6c) Define

$$H_-(x_1, x_2) = \max\{Q_0^*(x_1) + Q_1^*(x_2) - 1, 0\} \quad (3.40)$$

$$H_+(x_1, x_2) = \min\{Q_0^*(x_1), Q_1^*(x_2)\} \quad (3.41)$$

6d) Estimate θ_-^* and θ_+^* , respectively, by $\theta_-^* = \theta(H_-^*)$ and $\theta_+^* = \theta(\hat{H}_+^*)$.

7) Repeated independent generation of F_1^* and F_0^* yields a sequence of independent realizations of θ_+^* and θ_-^* , which can be used to approximate their actual bootstrap distribution.

3.10 Appendix C: Quantile Treatment Effects

Let $Q_q(Y | T)$ be the conditional quantile function of the conditional distribution $F(Y | T)$, where $T \in \{0, 1\}$ is the binary variable indicating treatment status: it takes the value of 1 if treated, and 0 otherwise. Assume $F(Y | T)$ is continuous and strictly increasing, and that $Q_q(Y | T)$ is linear:

$$Q_q(Y | T) = \alpha_q + \beta_q T$$

It can be shown that the parameters α_q and β_q can be characterized as follows (Koenker 1978)

$$(\alpha_q, \beta_q) = \arg \min_{(\alpha, \beta) \in \mathbb{R}^2} E[\rho_q(Y - \alpha - \beta T)]$$

where $\rho_q(u) = u(q - I(u < 0))$ is the check function. Let α^* and β^* be the solution to this problem. Then it is easy to show that the QTE can be recovered from here since

$$\tau(q) = y_1(q) - y_0(q) = Q_q(Y | T = 1) - Q_q(Y | T = 0) = \beta_q^*$$

For the estimation, let $(y_i, T_i)_{i=1}^n$ be a sample from the population. Then we can apply the analog principle and follow Koenker and Bassett (1978) to estimate α and β :

$$(\hat{\alpha}_q, \hat{\beta}_q) = \arg \min_{(\alpha, \beta) \in \mathbb{R}^2} n^{-1} \sum_{i=1}^n \rho_q(Y_i - \alpha - \beta T_i)$$

BIBLIOGRAPHY

- Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimation of quantile treatment effects. *Econometrica* 68, 399–405.
- Ainslie, G. (1992). *Picoeconomics*. Cambridge University Press.
- Andrews, D. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 70, 91–117.
- Angeletos, G.-M., D. Laibson, A. Repetto, J. Tobacman, and S. Weinberg (2001). The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *Journal of Economic Perspectives* 15(3), 47–68.
- Ashraf, N., N. Gons, D. Karlan, and W. Yin (2003). A review of commitment savings products in developing countries. Asian Development Bank Economics and Research Department Working Paper, 45.
- Ashraf, N., N. Gons, D. Karlan, and W. Yin (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *Quarterly Journal of Economics*. Forthcoming.
- Baland, J. M. and J. Robinson (2000). Is child labor inefficient? *Journal of Political Economy* 108(4), 663–679.
- Besley, T. and S. Coate (1992). Workfare versus welfare: Incentive arguments for work requirement in poverty-alleviation programs. *American Economic Review* 82(1), 249–261.
- Bickel, P., F. Gotze, and W. V. Zwet (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica* 7, 1–31.

- Bitler, M., J. Gelbach, and H. Hoynes (2003). What mean impacts miss. Rand Labor and Population Discussion Series.
- Blackorby, C. and D. Donaldson (1988). Cash versus kind, self-selection, and efficient transfers. *American Economic Review* 78, 691–700.
- Brown, M. B. and A. B. Forsythe (1974). Robust tests for equality of variances. *Journal of the American Statistical Association* 69, 364–367.
- Buddelmeyer, H. and E. Skoufias (2004). An evaluation of the performance of regression discontinuity design on progressa. World Bank Policy Research Working Paper, 3386.
- Chakravarty, S. (1983). A new index of poverty. *Mathematical Social Sciences* 6, 761–766.
- Chesher, A. and C. Schluter (2002). Welfare measurement and measurement error. *Review of Economic Studies* 69, 357–378.
- Clark, S., R. Hemming, and D. Ulph (1981). On indices for the measurement of poverty. *Economic Journal* 91, 307–313.
- Cowell, F. A. and M.-P. Victoria-Feser (1996). Poverty measurement with contaminated data: A robust approach. *European Economic Review* 40, 1761–1771.
- Cropper, M. and D. Laibson (1998). The implications of hyperbolic discounting for project evaluation. World Bank, working paper 9356.
- Das, J., Q. T. Do, and B. Ozler (2004). Conditional cash transfers and the equity-efficiency debate. World Bank paper series, 3280.
- de Wet, T. (1976). Berry-esseen results for the trimmed mean. *S. Afr. Statist. J.* 10, 77–96.

- DellaVigna, S. and U. Malmendier (2003). Overestimating self-control: Evidence from the health-club industry. mimeo, University of California Berkeley.
- Doksum, K. (1974). Empirical probability plots and statistical inference for non-linear models in the two-sample case. *The annals of statistics* 2, 267–277.
- Dominitz, J. and R. Sherman (2005). Identification and estimation of bounds on school performance measures: a nonparametric analysis of a mixture model with verification. Mimeo, California Institute of Technology.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7, 1–26.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82, 171–185.
- Fang, H. and D. Silverman (June, 2004). Measuring time-inconsistency: Evidence from work-welfare decisions in the nsly. mimeo, Yale University.
- Fields, G. (2001). *Distribution and Development* (First ed.). Russell Sage Foundation-MIT Press.
- Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica* 52, 761–766.
- Frechet, M. (1951). Sur les tableaux de correlation dont les marges sont donnees. *Annals University Lyon: Series A* 14, 53–77.
- Frederick, S., G. Loewenstein, and T. O’Donoghue (2002). Time discounting and time preference: a critical review. *Journal of Economic Literature* 40, 351–401.
- Goldman, S. (1979). Intertemporally inconsistent preferences and the rate of consumption. *Econometrica* 47, 621–626.

- Goldman, S. (1980). Consistent plans. *Review of Economic Studies* 47, 533–401.
- Gruber, J. and B. Koszegi (2001). Is addiction rational? theory and evidence. *Quarterly Journal of Economics* 116, 1261–1303.
- Hagenaars, A. (1987). A class of poverty indices. *International Economic Review* 28, 583–607.
- Hampel, F. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association* 69, 383–393.
- Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel (1986). *Robust Statistics*. Wiley, New York.
- Hardy, G., J. Littlewood, and G. Polya (1952). *Inequalities* (Second ed.). Cambridge University Press.
- Heckman, J. (1978). Dummy endogenous variable in a simultaneous equation system. *Econometrica* 46, 931–961.
- Heckman, J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64(4), 487–535.
- Horowitz, J. E. and C. F. Manski (1995). Identification and robustness with contaminated and corrupted data. *Econometrica* 63, 281–302.
- Huber, P. (1981). *Robust Statistics*. Wiley, NY.
- Imbens, G. and C. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica* 72, 1845–1857.
- INEGI (2002). *Encuesta Nacional de Ingresos y Gastos de los Hogares de 2002*. Instituto Nacional de Estadística, Geografía, e Informática, Mexico.

- Kakwani, N. (1980). On a class of poverty measures. *Econometrica* 48, 437–446.
- Kakwani, N. (1993). Statistical inference in the measurement of poverty. *The Review of Economics and Statistics*, 632–639.
- Kanbur, R. and L. Haddad (1997). Are better off households more unequal or less unequal? *Oxford Economic Papers* 46(3), 445–458.
- Kanbur, R., J. Pirttila, and M. Tuomala (2004). Non-welfarist optimal taxation and behavioral public economics. Mimeo.
- Khandker, Shahidur, and F. Pitt (2003). Subsidy to promote girl's secondary education the female stipend program in bangladesh. Discussion Paper.
- Koenker, R. (1978). Regression quantiles. *Econometrica* 46, 33–50.
- Kreider, B. and J. Pepper (2004). Inferring disability status from corrupt data. Mimeo.
- Laibson, D. (1997a). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* 112(2), 443–477.
- Laibson, D. (1997b). Hyperbolic discounting and golden eggs. *Quarterly Journal of Economics* 112(2), 443–447.
- Laibson, D. (1998). Life cycle consumption and hyperbolic discount functions. *European Economic Review* 42, 861–871.
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics*. Stanford University Press.
- Lopez-Acevedo, G. and A. Salinas (2000). Marginal willingness to pay for education and the determinants of enrollment in mexico. World Bank, Working Paper Series, 2405.

- Lowenstein, G. and R. Thaler (1989). Anomalies: Intertemporal choice. *Journal of Economic Perspectives* 3, 181–193.
- Manski, C. (2003). *Partial Identification of Probability Distributions* (First ed.). Springer-Verlag.
- Manski, C. F. (1997). Monotone treatment response. *Econometrica* 65, 1311–1334.
- Martinelli, C. and S. Parker (2003). Should transfers to poor families be conditional on school attendance? a household bargaining perspective. *International Economic Review* 44(2), 523–544.
- Molinari, F. (2005a). Identification of probability distributions with misclassified data. Mimeo, Cornell University.
- Molinari, F. (2005b). Missing treatments. Working Paper, Cornell University.
- O’Donoghue, T. and M. Rabin (1993). Doing it now or later. *American Economic Review* 89(1), 103–124.
- O’Donoghue, T. and M. Rabin (2001). Choice and procrastination. *Quarterly Journal of Economics* 116(1), 121–160.
- Peleg, B. and M. Yaari (1973). On the existence of a consistent course of action when tastes are changing. *Review of Economic Studies* 40, 391–401.
- Phelps, E. and R. Pollak (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies* 35(1), 185–199.
- Pratt, J. (1964). Risk aversion in the small and the large. *Econometrica* 32, 122–136.
- Ravallion, M. (1994). Poverty rankings using noisy data on living standards.

- Economics Letters* 45, 481–485.
- Ruschendorf, L. (1981). Sharpness of frechet-bounds. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57(4), 293–302.
- Rutheford, S. (1999). Savings and the poor: The methods, use and impact of savings by the poor of east africa. Report prepared for MicroSave-Africa.
- SEDESOL (2002). Nota tecnica para la medicion de la pobreza con base en los resultados de la encuesta nacional de ingresos y gastos de los hogares, 2002. Secretaria de Desarrollo Social, Mexico.
- Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica* 44, 219–231.
- Shapiro, J. M. (2005). Is there a daily discount rate? evidence from the food stamp nutrition cycle. *Journal of Public Economics* 89, 303–325.
- Shavell, S. and L. Weiss (1979). The optimal payment of unemployment insurance benefits over time. *Journal of Political Economy* 87, 1347–1362.
- Sierpinski, W. (1928). Un theoreme general sur les familles d'ensembles'. *Fundamenta Mathematicae* 12, 206–210.
- Stephens, M. (2002). Paycheck receipt and the timing of consumption. NBER, working paper 9356.
- Stephens, M. (2003). 3rd of the month: do social security recipients smooth consumption between checks? *American Economic Review* 93, 406–422.
- Stigler, S. (1973). The asymptotic distribution of the trimmed mean. *The Annals of Statistics* 1, 472–477.
- Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization.

Review of Economic Studies 23(3), 165–180.

Szekely, M., N. Lustig, M. Cumpa, and J. Mejia (2000). Do we know how much poverty there is. Mimeo, Inter-American Development Bank.

Tchen, A. (1980). Inequalities for distributions with given marginals. *The Annals of Probability* 8(4), 814–827.

Watts, H. (1968). An economic definition of poverty. In D. Moynihan (Ed.), *On Understanding Poverty*. New York: Basic Books.