Heterogeneous Program Impacts in PROGRESA

Habiba Djebbari
University of Maryland
IZA
hdjebbari@arec.umd.edu

Jeffrey Smith
University of Maryland
NBER and IZA
smith@econ.umd.edu

Abstract

The "common effect" model in program evaluation assumes that all treated individuals have the same impact from a program. Does the program have the same effect on everyone? Will some groups benefit more from the program than others? We investigate heterogeneity in program impact for the Mexican social program PROGRESA, which is a means-tested conditional cash transfer program implemented in rural regions of the country. The design of the program provides a theoretical motivation for exploring heterogeneity in program impacts. We examine the program targeting mechanism and find heterogeneity in the eligible population along the criteria used for beneficiary selection. We also investigate the overall heterogeneity of program impacts, which includes both observed and unobserved heterogeneity. Experimental data are sufficient to identify mean program impacts or impacts on subgroups, but do not identify unobserved heterogeneity in impacts. Using a non-parametric technique, we find evidence against the "common effect" model. This result does not rely on any assumption and thus is particularly strong evidence of heterogeneous treatment effects. Additional assumptions allow us to further analyze the distribution of impacts.

First Version November 2003
This Version July 2004

1. Introduction

The most commonly used estimators in program evaluation provide information on the program mean impact. Yet, this is only a narrow answer to the question of how well a program works. Exploring heterogeneity in program impacts provide information on the distributional effects of policy interventions in a way that goes beyond mean impacts. It can also help to go inside the “black box” of mean program impacts and learn more about how policies generate their mean effects (see Heckman and Smith 1995 for a discussion of the main criticism towards experimental methods).

The purpose of this paper is to investigate the importance of heterogeneity in program impacts for the Mexican program for education, health and nutrition PROGRESA. Does the program have the same effect on everyone? Will some groups benefit more from the program than others? Most of the existing literature on heterogeneity of treatment effects mainly looks at the heterogeneous effect that vary with observed characteristics, i.e. the heterogeneity on subgroups of the population. We explore the heterogeneity of impacts as a function of the criteria used by PROGRESA to select beneficiaries. Yet, we also investigate the overall heterogeneity of program impacts, which includes both observed and unobserved heterogeneity. Four other studies explore similar aspects of the distribution of impacts in the US context (Heckman, Smith and Clements 1997, Black, Smith, Berger and Noel 2003, Bitler, Gelbach and Hoynes 2004, and Biddle, Boden and Reville 2003). This is the first paper that investigates the overall heterogeneity in program impacts for a policy intervention in a developing country.

In the next section, we describe the Mexican program for education, health and nutrition PROGRESA. The Mexican social program PROGRESA is a means-tested conditional cash transfer program. It targets the poorest households in the most remote rural regions of Mexico. The PROGRESA program is designed as a cash transfer payment given to the mother in the household upon compliance with a defined set of requirements (e.g. regular child attendance to school and frequent visits to health centers). Thus, the program has both a short-term poverty reduction objective and a longer-term objective in terms of investment in human capital. In addition, we describe the data and the experimental design. The latter consists of an experiment with randomly assigned treatment and control localities. Within the treatment localities, eligible households are offered the program benefits. Within the control localities, eligible households do not receive the benefits during the two-year evaluation period. Data collected in November 1998, i.e. six months after the start of the program, as well as data collected in June 1999 and in November 1999 are used in this study. By sending conditional cash transfers to households, the PROGRESA program's main objectives are to reduce household poverty, to increase household food consumption, to encourage investment in human capital. Thus, the outcomes of interest include household total expenditures and value of consumption, household food expenditures and value of food consumption, and the children's time spent in schooling activities, income-generating activities and domestic activities. A summary of the evaluation of PROGRESA can be found in Skoufias (2001).

In the third section, we present the theoretical framework for our investigation. Based on the program benefit scheme and on how well the program was implemented, we discuss the case for heterogeneous program impacts versus the "common effect" model in PROGRESA. Then, we analyze the heterogeneity in program impacts on household welfare, consumption and time allocation that is raised by the conditionality of the cash transfer.

In the fourth section, we explore the heterogeneity of impacts along observable characteristics. We focus on the variation in impacts as a function of the two criteria used by PROGRESA to select beneficiaries, i.e. a village marginality index a household poverty index. Both indices are constructed by the program officials using information collected prior to the intervention (Skoufias *et al.*, 1999). This analysis allows us to assess the effectiveness of the targeting mechanism. Are the poorest households in the most marginal villages getting a greater program impact than less poor households from less marginal places? In order to answer this question, we estimate treatment effects on subgroups, which is the most common way to investigate the distributional impacts of a program. We estimate program impacts along the targeting criteria for household total expenditures and value of consumption, household food expenditures and value of food consumption, and the children's time spent in schooling activities, income-generating activities and domestic activities.

In the two subsequent sections, we investigate the overall heterogeneity of program impacts, which includes both observed and unobserved heterogeneity. Experimental data are sufficient to identify mean program impacts or impacts on subgroups, but do not identify unobserved heterogeneity in impacts.

In the fifth section, we derive a lower bound for the total variance of impacts using results from classical probability theory. We test whether the lower bound is significantly different from zero. This allows us to test whether the net effect of the program is homogeneous over both observed and unobserved characteristics. Additional assumptions are required to further analyze the distribution of impacts. We assume perfect positive dependence between untreated outcome levels and treated outcome levels. This assumption holds if the ranks of the households are unaffected by the program. It allows us to estimate quantile treatment effects (QTE). The QTE estimation provides information on how the impact varies at different quantiles of the untreated distribution.

In the sixth section, moments of the distribution of impacts are identified under the assumption that the untreated outcome levels and program impacts are independent. This assumption means that households do not anticipate gains from the program at the time they decide to participate in the program, which is likely to occur when households are randomly assigned to a treatment and a control group. We test this assumption and we identify the first two moments of the distribution of impacts using a parametric random coefficient model. We assume a normal distribution and estimate the distribution of impacts. Alternatively, when the independence assumption holds, we identify the first four moments of the distribution of impacts using a more flexible parametric approach. We approximate the distribution of impacts assuming that it belongs to the Pearson family of distribution.

We find evidence against the "common effect" assumption, which is robust to the assumption adopted. From the analysis on subgroups, we find evidence that there is variation in impacts along the targeting criteria in the treated population, that the geographic targeting is effective but the beneficiary selection on household poverty within poor localities is not. The last two findings are consistent with the findings in Skoufias *et al.* (2001) on the program targeting mechanism.

The bounding analysis provides evidence that the net variance of impacts along observed and unobserved characteristics is different from zero. This result does not rely on any assumption and thus is particularly strong evidence of heterogeneous treatment effects. We also find from the bounding analysis that the experimental data are consistent with a large range of impact distributions. From the QTE estimation, we learn that the program impact on wealth and nutrition is lower for households who were at a lower level of wealth and nutrition prior to the intervention in the last two rounds. This is consistent with the theoretical prediction that treatment effect on wealth and nutrition are higher for the households whose cost of complying with the program requirements is the lower. The impacts are positive at each decile of the untreated distribution of outcomes. In the first round, the program impact does not vary with the untreated outcome distribution.

Under the assumption of independence between gains from the program and the untreated outcome levels and when impacts follow a normal distribution, we find that the fraction of the treated population with a positive impact ranges is higher for the latest rounds. This finding suggest that in the first round, some households may have supported the cost of the program requirements without receiving the program benefits because of implementation problems. Yet, for most of the outcomes of interest, the assumption of independence between program impacts and untreated outcome levels fails to hold. This suggests selective compliance to program requirements of the eligible population.

2. The PROGRESA program and data.

2.1. Description of the program

The PROGRESA program targets poor rural households in Mexico. It has been implemented since 1998. At the end of 1999, it covered 2.6 million families, i.e., about 40% of all rural households and one ninth of all families in Mexico. In 1999, the annual program budget was approximately \$777 million, which corresponds to 0.2% of Mexico's GDP (Skoufias, 2001). In January 2002, the Inter-American Development Bank approved its largest loan ever to Mexico for expanding PROGRESA to urban areas of the country. Despite the important recent political changes, PROGRESA remains in place, under the new name Opportunidad.

The actual targeting of PROGRESA involves two stages: (1) the selection of the localities where PROGRESA operates, (2) the selection of beneficiary households within the selected localities. The most remote localities with a minimum of infrastructure (e.g. at least one primary school and a health center) are selected. Within each selected locality, the selection of households is based on pre-program survey information on household wealth that includes income and relevant household characteristics. A household is eligible to the program if it falls below a poverty line defined by income and other relevant socio-economic attributes. On average, 78% of each locality population is found to be poor, i.e. eligible for the program benefits. The program benefits are conditional income transfers and they are comprised of two components:

1. Educational grants given to families with children in the last three years of primary school and secondary school children. The grant amounts vary by grade and gender in favor of girls and of the most advanced children and reflect opportunity costs. The grants are given upon attendance to school.[1] A complex system of verification based on forms completed and signed by teachers and school directors ensures that the attendance requirement is met before sending money to the households.
2. All selected households can benefit from a monetary transfer designed to help them improve their nutrition. This component is commonly called the food cash transfer. But, although households are encouraged to spend the money on food, they are not required to do so. In order to receive this cash transfer, they are required to make regular visits to health centers and to participate to health talks. Only one visit per year to a health center is required for adults, two to five visits a year for pregnant and breast-feeding women and two to seven visits a year for infants and children. In addition, in-kind nutritional supplements are provided to under-nourished children and infants and to pregnant and breast-feeding women.

The average transfer from October 1998 to November 1999 is about 197 pesos per household per month[2], which is equivalent to 20% of the mean value of consumption of a poor household. An additional requirement of PROGRESA is that households

[1] Children are required to attend school with an 85% monthly attendance rate. If a child fails to meet this requirement, families stop receiving the educational grant for that child, but remain eligible for other program benefits.
[2] The figures are in November 1998 pesos and the value is approximately $20 U.S.

withdraw from other assistance programs that share similar objectives with PROGRESA, such as Ninos de Solidaridad, DICONSA, LICONSA and INI. [3]

2.2. The experimental design

The evaluation sample is designed as an experiment with randomization of localities into treatment and control groups. The randomization is at the village level rather than at the household level in order to avoid contamination bias. Only eligible households in treatment localities actually receive any transfer. Eligible households in control localities are denied these transfers until the end of the evaluation period in 2000.[4] The random assignment of localities in treatment and control groups was conducted in order to evaluate the impacts of the program on a range of outcomes.[5]

Behrman and Todd 1999 evaluate the randomness of the sample and find that the treatment and control groups mean outcomes at the locality level are similar before the intervention. However, they find small differences at the household and individual level.

Note also that the existing alternative assistance programs do not operate anymore in all of the PROGRESA localities, i.e. the treatment and control localities. Thus, comparing treatment localities average outcomes and control localities average outcomes gives an estimate of the mean program impact.

All households in the selected localities are interviewed before and at several points in time after the start of the program. The evaluation dataset consists of repeated observations (panel data) for 24,000 households from 506 localities (320 assigned to the treatment group and the remaining 186 to the control group) over five rounds of survey (baseline: October 1997 and March 1998; follow-ups: October 1998, May 1999 and November 1999).[6]

---

[3] Ninos de Solidaridad provide educational grants. DICONSA maintains subsidized prices for basic food items. LICONSA provide poor families with one free kilogram of tortillas and subsidize the price of milk. INI is targeted to indigenous people and provide lodging and food or educational grants to students.

[4] Extending the PROGRESA program to all the eligible population had to be done by phases because of the size of the program. For a random subset of the villages, the incorporation to the program was postponed for two years. Thus, this subset of the population acts as a control group.

[5] These impacts are assessed using estimators commonly used in program evaluation (e.g. the difference-in-difference estimator). Detailed data description and mean program impacts are presented in a series of research reports (see http://www.ifpri.org/themes/progresa.htm).

[6] A time allocation questionnaire was administred once in November 1999. It collected very detailed information on each individual's activities.

2.3. The outcomes of interest

The outcomes of interest in the empirical section include the wealth of the household measured by household total expenditures and by the value of consumption of all goods and services (Deaton, 1997). Improving household nutrition is a key objective of the program. Household food expenditures and the value of food consumption provide quantitative measures of nutrition. In addition, we examine the time children spent in schooling activities, income-generating activities and domestic activities using a detailed module on time allocation. We denote income-generating activities as all of the activities involving work outside the house, including wage labor (for an employer, on one's own firm/farm with salary or other paid casual work) and non-wage labor (as an aide, on one's own firm/farm and other non-paid casual work). Domestic activities include all activities that take place at home and could have been realized by someone else the household would have hired, such as cleaning the house, washing, sewing and ironing clothes, shopping for the household, preparing meals and washing dishes, fetching water or wood, disposing garbage, taking care of animals and fields, looking after children including taking them to school, or looking after elderly or sick people. A third category of activity consists of attendance at school and time spent studying outside the classroom.[7] In addition to data on the individual and household outcomes, individual and household characteristics are collected as well.

3. A Theoretical Discussion.

In a paper on the welfare reform in the US, Bitler, Gelbach and Hoynes (2004) find evidence of program impact heterogeneity which is consistent with the theoretical predictions of a standard labor supply model. In this section, we first discuss the case for heterogeneous program impacts versus the "common effect" model in PROGRESA based on the program benefit scheme and on how well the program was implemented. Then, we analyze the heterogeneity in program impacts emerging from the conditionality of the cash transfer.

Households with different preferences, budget sets and production sets are likely to respond differently to a homogeneous treatment. Yet, the design of the program can address this issue by varying the treatment. For example, in the case of Mexico, one expects parents to be less likely to send daughters to school than sons. The PROGRESA program addresses the problem by providing a larger grant amount for girls sent to school. In addition, one expects that the older the child is, the higher his or her opportunity cost of time in the labor market. Again, the program anticipates this source of heterogeneity and provides larger grants for secondary school children. The opportunity cost of time of children is also likely to be higher for households who hold or operate land. Girls who live in large households with many younger children are also more likely to be employed at home. Mothers whose opportunity cost of time in domestic or labor activities is higher are likely to bear a high cost from participating in the program, which would lower the overall impact of the program on their households.

Yet, the program does not control for all potential sources of heterogeneity by varying the treatment. In addition, it is unclear whether the program payment schedule could exactly balance out differences in costs. Thus, we can expect to find heterogeneity in program impacts although the program treatment already varies from a household to another. Finally, although the program started in Spring 1998, 27% of the eligible households in the treatment group did not receive any cash benefits by March 2000 according to administrative data. This is mostly due to administrative problems in the implementation of the program. If these households bore the cost of participating in the program without receiving the program benefits, they are likely to have experienced a loss in welfare and a decrease in consumption. Households who sent children to school and complied with program requirements have experienced delays in the receipt of the cash transfer due to delays in the verification of the requirements or in the delivery of the monetary transfers. Coady and Djebbari (1999) assess the early stages of the implementation of the program.

In the remaining of the section, we discuss the effect of the conditionality of the educational grant on household behavior and welfare. This analysis complements the study by Skoufias and Parker (2001). Skoufias and Parker use a standard labor supply

[7] Finally, leisure time is a residual, i.e. the remaining time in a day after subtracting time spent for income generating activities, domestic activities and school. Observations for which leisure time was less or equal to 8 hours were deleted from the sample.

model and find that the effect of the conditionality on children's time allocation depends on households preferences and initial wealth level.

In a simplified version of the standard economic model of time allocation, a household composed of one adult and one child optimally chooses a composite good z, the child's schooling time t_s, and the adult's and child's leisure time l_a, l_c. The indirect utility function V (.,.) is as follows:

$$V(p,Y) = \underset{X}{Max}\ \{U(X) \quad s.t. \sum pX = Y = I + (w_a + w_c)\, T\}\ \text{, where } X = (z,\ t_s,\ l_c,\ l_a).$$

The vector of prices associated with X is denoted p, household full income Y is composed to household non-labor income I and household value of the common time endowment T at prices w_a and w_c. Suppose that in a first stage, household optimal choices are limited to $(z,\ l_c,\ l_a)$.

Like the indirect utility function, the partial indirect utility function is increasing in Y. For a given Y, the partial indirect utility function is first increasing, and then decreasing in t_s: it admits a maximum in t_s. We denote by t_s^* the optimal schooling time:

$$t_s^* = \underset{t_s}{ArgMax}\, V(p,t_s,Y) = \underset{z,t_s,l_c,l_a}{ArgMax}\ \{U(X) \quad s.t. \sum pX = Y\}.$$

Program participants must be at least as well-off taking up the program than not taking it up. Yet, the increase in participant welfare can reflect an increase in consumption or an increase in leisure time. Because households have to send children to school a minimum of $\bar{t}$ hours, the welfare effects of a pure income transfers vs. tied income transfers are analyzed using unconstrained and constrained optimization problems:

$$V^U(p,Y) = \underset{X}{Max}\ \{U(X) \quad s.t. \sum pX = Y\} = \underset{t_s}{Max}\, V^U(p,t_s,Y),$$
$$V^C(p,Y) = \underset{X}{Max}\ \{U(X) \quad s.t. \sum pX = Y and\ t_s \geq \bar{t}\} = \underset{t_s}{Max}\, V^C(p,t_s,Y).$$

One can distinguish between four types of households based on the constrained and unconstrained partial indirect utility functions. Figure 1 shows how these two functions vary with t_s. Denote by A the optimal schooling time without transfer, B the optimal schooling time with a pure unconditional transfer and C the optimal schooling time with a conditional transfer. Let g be the transfer amount.

1. Type I households are better off accepting a tied income transfer than refusing it because point C is above point A. Without any aid (point A), these households were already sending their children to school at least the minimum of $\bar{t}$ hours. The effect of the program on the child's schooling time is thus minor. Furthermore, the optimal schooling time is the same whether or not the transfer is conditional (B = C):

$$V^{C}(p,Y)=V^{U}(p,Y)<V^{C}(p,Y+g)=V^{U}(p,Y+g),$$
$$t^{U*}(p,Y)=t^{C*}(p,Y)>\overline{t},$$
$$t^{U*}(p,Y+g)=t^{C*}(p,Y+g)>\overline{t}.$$

   For Type I households, the main effect of the program consists of the pure income effect of the cash transfer. We can expect type I households to have higher expenditures after taking-up the program.

2. Type II households also are better off accepting a tied income transfer than refusing it (point C is above point A). Before receiving the aid, these households choose to send children to school less than the minimum of $\bar{t}$ hours (point A). When they get the educational grants, the children attend school more than the minimum required attendance level. Yet, whether they receive a conditional or an unconditional transfer does not affect the optimal schooling time (B = C).

$$V^{C}(p,Y)<V^{U}(p,Y)<V^{C}(p,Y+g)=V^{U}(p,Y+g),$$
$$t^{U*}(p,Y)<t^{C*}(p,Y)=\overline{t},$$
$$t^{U*}(p,Y+g)=t^{C*}(p,Y+g)>\overline{t}.$$

   Like households of type I, type II households only benefit from the pure income effect of the cash transfer. With the additional PROGRESA income, children's time spent in school is increased above the minimal attendance requirement level. We can also expect higher expenditures for these households.

3. Type III households are similar to type II households in that (1) they are better off accepting a tied income transfer than refusing it, (2) before receiving the aid, children were sent to school less than the minimum of $\bar{t}$ hours. However, the optimal schooling time these households would have chosen for their children if given an unconditional transfer would have been smaller than the optimal schooling chosen with a tied income transfer. Point C is to the right of point B. For type III households, the conditionality of the grant matters. Type III households are affected both by the pure income effect of the cash transfer and the effect of a lower price of schooling

driven by the attendance requirement. This price effect has the standard substitution and income effect, which reinforces the pure income effect of the transfer.

$$V^C(p,Y) < V^U(p,Y) < V^C(p,Y+g) < V^U(p,Y+g),$$
$$t^{U*}(p,Y) < t^{C*}(p,Y) = \bar{t},$$
$$t^{U*}(p,Y+g) < t^{C*}(p,Y+g) = \bar{t}.$$

With the PROGRESA conditional transfer, families can exactly meet the minimum attendance requirement. Children spend more time studying than without a transfer. Yet, at income level $(Y+g)$, type III households may have to bear the cost of children's foregone labor earnings due to the implicit reduction in labor time $\bar{t} - t^{U*}(p,Y+g)$ that results from the conditionality of the grant. Thus, the grant impact on household expenditures may be negative for these households. We expect the food cash transfer benefit to have a positive effect on household expenditures. The net effect on expenditures could be either positive or negative.

4. The last type of household (type IV) looks like the type II and III households in that they were sending children to school less than the minimum numbers of hours targeted by the program. In addition, like type III households, the conditionality of the grant affects their potential welfare. However, it also affects the outcome of their choice. When compliance to program requirements is demanded, in their choice between (i) complying and receiving the grant, (ii) not complying and not receiving a grant, these households choose the latter. They are better off not taking the conditional grant than taking it. Point A is above point C. These households are likely to be the poorest or at least the ones who rely the most on child labor. The costs of complying with the program requirements include the foregone earnings of their children. With an unconditional grant, although their demand for schooling would not have attained the minimum of $\bar{t}$ hours, it would still have been greater than it actually is when facing the trade-off implicit in the conditional grant. Point B is at the right of point A.

$$V^C(p,Y) < V^C(p,Y+g) < V^U(p,Y) < V^U(p,Y+g),$$
$$t^{U*}(p,Y) < t^{C*}(p,Y) = \bar{t},$$
$$t^{U*}(p,Y+g) < t^{C*}(p,Y+g) = \bar{t}.$$

Type IV households are not getting the educational grants. The only effect of the program on children's time allocation and household expenditures is through the food cash transfer component.

Two points can be made from this analysis. First, although the treatment is heterogeneous, we cannot rule out the case of heterogeneity in program impacts. Problems in program implementation are likely to be at the origin of lower or even temporary negative program impacts for some participants. Second, identifying the different types of households allows us to form predictions concerning the heterogeneity of program impacts. In particular, we find that impacts on expenditures are greater for households who are meeting or almost meeting program requirements prior to the intervention. We expect program impacts on child's schooling for these households to be small or zero. These are likely to be well-off households or households with young children. We expect to find very small program impacts on expenditures for households whose costs of participating are the highest. These impacts could even be negative for households who meet program requirements but would have still rely on child labor under an unconditional scheme. Yet, program impacts on child's schooling would be the largest for these households. We also expect small or zero program impacts on child's schooling for households who cannot meet program requirements. These are likely to be the poorest households who greatly rely on child labor.

Finally, note that we restricted the previous analysis to households with a single child. In Mexican households with more than a child, young children and boys are more likely to be sent to school than older children and girls. Thus, program impacts on schooling (labor) of older children or girls could even be negative (positive).

4. Heterogeneity of impacts and the targeting of the program.

4.1. Empirical Model

If the goal of the targeting is to increase the efficiency of the program then the treatment should be allocated to those for which the impact is the largest.[8] We explore the heterogeneity of impacts as a function of the two criteria used by PROGRESA to select beneficiaries, i.e. a village marginality index and a household poverty index. Both indices

[8] See section 5.7 of Berger, Black and Smith (2000) on optimal targeting of unemployment insurance in the US.

are constructed by the program officials using information collected prior to the intervention (Skoufias *et al.*, 1999). Both indices are included in the datasets.

The village marginality index is constructed using village-level information on the illiteracy rate of heads of households, on access to basic infrastructure (running water, a drainage system, electricity), on housing characteristics (ratio of household members to rooms in the house, frequency of houses where floor are made of dirt) and the importance of agricultural activities in the village. The higher the value of the marginality index, the more remote is the village. The household poverty index takes into account household characteristics, family assets and per capita income. The higher the poverty score is, the poorer the household is. In each eligible village, households are classified as poor and non-poor based on this score.

We allow the treatment effect to vary with the village marginality index and the household poverty index. If the targeting mechanism is effective, then the poorest households in the most marginal villages get a greater program impact than less poor households from less marginal places.

In order to test this hypothesis, we estimate program impacts on household total expenditures and value of consumption, household food expenditures and value of food consumption, and the children's time spent in schooling activities, income-generating activities and domestic activities (all outcomes are designed by Y in the equations below).

The treatment group households ($T = 1$) are compared to the control group households ($T = 0$). We include interaction terms between the poverty index (*Pindex*) and the village marginality index (*Vindex*) and the treatment indicator (T).

We control for household or individual characteristics in order to obtain more precise estimates. In addition, this should correct for any differences not accounted for by the randomization of localities into treatment and control groups. When estimating program impacts on nutrition and wealth, the control variables include household composition and characteristics of the head of household. For impacts on children's time allocation, the control variables include the child's age, parent's education, the age of the mother and the father, whether the head of household is a female or a male, whether the head of household speaks an indigenous language and variables measuring the demographic composition of the household. Control variables are designed by X.

We estimate the following equation.

$$(1) \quad \begin{aligned} &Y = \alpha + \alpha_1 * Pindex + \alpha_2 * Vindex + \alpha_3 * Pindex * Vindex + \\ &\quad \beta * T + \beta_1 * T * Pindex + \beta_2 * T * Vindex + \beta_3 * T * Pindex * Vindex + X\delta + \varepsilon. \end{aligned}$$

First, we test whether there is any program impact on the outcome by evaluating the following joint hypothesis:

$$H_0 : \beta = \beta_1 = \beta_2 = \beta_3 = 0.$$

Then, we test whether the program impact along the poverty and the village marginality criteria is the same for all households by testing the following hypothesis in equation (1):

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$$

Rejecting this null hypothesis is evidence of heterogeneous program impacts along the program targeting criteria. In order to test whether the impacts are decreasing or increasing along the household poverty and village marginality indices, we examine the sign of the coefficients on the interaction terms in equation (1). We expect the sign of the coefficient on the interaction terms to be positive for the wealth, nutrition and education outcomes, and negative for children labor and domestic activities. We also estimate the fraction of the treated population with a positive program impact for each of the outcomes.

4.2. Results and discussion

Table 1 shows the estimation results for equation (1) for the wealth and nutrition variables. We reject the null hypothesis that the treatment does not vary with the targeting criteria for all outcomes and rounds. The rejection is stronger for the interaction with the village marginality index and weaker for the interaction with the poverty index. We find that the signs of the coefficients on the interaction terms are positive. This is evidence that the program impacts on wealth and nutrition are greater for those who initially live in the most remote villages. The coefficient on the interaction term between treatment and village marginality index decreases with time. The same holds for the coefficient on the interaction term between the treatment indicator and the two indices, which decreases with time. This means that the overall difference in program impacts between households who initially live in the most remote areas and those who initially live in the less marginal areas is getting smaller over time. Table 2 shows that the program has a significant impact on the wealth and nutrition aggregates in all rounds and that the fraction with a positive program impact on wealth and nutrition is also increasing over time. In the second and third round, the program impact on wealth and nutrition is positive for more than 90 percent of the treated group.

Results from the impact of the program on time allocation outcomes for boys of primary school-age and secondary school-age are presented in Table 3. We do not find any program impact on schooling for primary school-age boys. The overall program impact on time spent in domestic chores is insignificant as well. Yet, the fraction of eligible young boys with a positive program impact on time spent in domestic activities is unexpectedly high. We find evidence of a program impact on their labor activities. This impact varies with the program targeting criteria. The probability of participating in the labor market and time spent working is the lowest for the children who live in the most remote villages. Much of the variation in program impacts comes from the interaction with the village marginality index. Program impacts do not vary much with the initial poverty level of households.

For secondary school-age boys, there is a significant program impact on both schooling and labor activities. The program does not have an impact on their domestic activities. We find evidence that the program impact on participation in school and time spent studying is smaller for children who live in the most remote villages. Furthermore, only 63 percent of eligible boys experience a positive impact on schooling time. This is not the expected finding. It is plausible that for the most remote villages, the decision to invest in education beyond primary school is less attractive than in less remote areas. We also find that participation in labor activities decreases more rapidly in the most remote areas than in the least remote areas for eligible households. Yet, the program impact on the time spent working does not vary with the targeting criteria. Nor do program impacts on any of the outcomes vary with the initial poverty level of households.

Table 4 shows the program impacts on girls' time allocation outcomes. For primary school-age girls, the program impacts participation in domestic activities and time spent in domestic chores. But, the overall impact of the program on either schooling or labor is insignificant. We find evidence of variation of program impacts along the targeting criteria for the domestic activities of primary school-age girls. As expected, the program impacts along the targeting criteria are negative for this outcome. This means that young girls from the most marginal areas are benefiting from a larger reduction in domestic activities because of the program than young girls from less marginal areas. Yet, the fraction of eligible young girls with a positive impact of the program on their domestic activities is unexpectedly high. Note also that program impacts do not vary with the initial poverty level of households.

Finally, PROGRESA has a significant impact on schooling and domestic activities for secondary school-age girls. Program impacts on schooling vary with the targeting criteria, but program impacts on domestic chores are independent of the initial poverty and marginality indices. We find that the coefficient on the interaction between treatment and the poverty and marginality indices is negative for the participation in school equation of secondary school-age girls. The sign is also negative for the time spent studying. This is not the expected finding. This means that the program impact is smaller for poorest girls from the most remote villages.

Overall, we find evidence that the program impacts vary with the program targeting criteria for most outcomes. In particular, program impacts depend on the initial place of living. We find that program impacts do not vary much with the initial poverty level of households. In an analysis of the PROGRESA program targeting, Skoufias *et al.* (2001) compare the current targeting mechanism to alternative selection models. They also find that program impacts could have been achieved using the geographic targeting alone. [9] As expected, we find that households from the most marginal villages get a greater program impact on wealth and nutrition than households from less marginal places. Similarly, children from the most remote villages are found to have a larger reduction on work and domestic activities from the program. However, we also find that the program does not affect the schooling of primary school-age children. In addition, both secondary school-age boys and girls have a smaller program impact on their schooling time when they live in the most remote villages. Low impacts on schooling in the primary grades are consistent with the program spending money to "buy the base" (see Todd and Wolpin, 2003, for a similar finding). On average, secondary school-age

[9] They also discuss the social cost raised by targeting within poor villages.

children from the most remote areas may be getting a lower program impact because many of them chose not to attend school and thus do not get the PROGRESA grant.

5. Heterogeneity of impacts and perfect positive dependence.

The existing literature on heterogeneity of treatment effects mainly looks at the heterogeneous effect that vary with observed characteristics, i.e. the heterogeneity on subgroups of the population. In the case of PROGRESA, other papers found evidence of differential impacts on child's schooling and labor for primary school-age children vs. secondary school-age children, for girls vs. boys, for drop-outs children vs. children continuing through school (see Skoufias 2001 for a synthesis of the results).

In this section, we investigate the overall heterogeneity of program impacts, which includes both observed and unobserved heterogeneity. First, we quantify the total variance of impacts and test whether the net effect of the program is homogeneous over both observed and unobserved characteristics. Experimental data are sufficient to identify mean program impacts or impacts on subgroups, but do not identify unobserved heterogeneity in impacts. Second, we assume perfect positive dependence between untreated outcome levels and treated outcome levels in order to explore other distributional aspects of program impacts. Perfect positive dependence occurs when the ranks of the households are unaffected by the program. This assumption means that a high (low) rank household in the untreated state remains a high (low) rank household after the treatment.

5.1. Testing non-parametrically for heterogeneous program impacts

The main issue in program evaluation is a problem of "missing data". One cannot simultaneously observe the outcome of interest in the case of program participation (Y_1) and in the case of non-participation to the program (Y_0) for a given individual. Let T denote participation in the program, with $T = 1$ if a person participates and $T = 0$ otherwise. Because localities are randomly assigned in treatment and control groups, the experiment provides information on the marginal distributions of the outcome, i.e. $F_1(y_1 \mid T=1)$, the participants' outcomes, and $F_0(y_0 \mid T=1)$, what the participants outcomes would have been had they participated (Heckman, Smith and Clements 1997). Although the joint distribution of outcomes is never observed, it can be bounded using classical probability inequalities due to Fréchet (1951) and Hoeffding (1940). Bounds for the correlation between Y_1 and Y_0 can be estimated, and thus the variance of the impact $\Delta = Y_1 - Y_0$ can also be bounded using the Fréchet-Hoeffding inequalities.[10] One can then test whether the minimum variance is statistically different from zero. Rejecting this hypothesis implies that the program impact is heterogeneous for the population covered by the experiment. The Fréchet-Hoeffding bounds are as follows:

$$Max[F_1(y_1 \mid T=1) + F_0(y_0 \mid T=1) - 1, 0)] \leq F(y_1, y_0 \mid T=1)$$
$$\leq Min[F_1(y_1 \mid T=1), F_0(y_0 \mid T=1)].$$

The upper-bound distribution corresponds to the case of perfect positive dependence between Y_0 and Y_1, i.e. the case for which the two marginal distributions are matched in ascending order. The lower-bound distribution in the Fréchet-Hoeffding inequalities corresponds to the case of perfect negative dependence, i.e. the case for which the marginal distributions are matched in the reverse order. Since the sample sizes of the treatment group and control group are different, we use the percentiles of each distribution. In each case, we calculate the outcomes correlation $r_{Y_0Y_1}$ and derive the bounds for $Var(\Delta) = Var(Y_1) + \text{var}(Y_0) - 2r_{Y_0Y_1}\sqrt{Var(Y_0)Var(Y_1)}$. We then test whether the lower bound of the variance, which is derived from the Fréchet-Hoeffding upper-bound distribution, equals to zero. In the common effect model (homogeneous program impacts), the impact is constant and $Var\ (\Delta) = 0$. Thus, rejecting the null hypothesis that the minimum variance is zero implies a rejection of the common effect model and provides evidence of heterogeneous program impacts, under the assumption of perfect positive dependence. In addition, the Fréchet-Hoeffding bounds give an estimate of the range of values for the variance of impacts.

5.2. A semi-parametric analysis

The standard common effect estimator assumes that all treated households receive the same impact from the program. When comparing the distributions of the outcome variable for the treatment group and the control group, the treatment group distribution is

[10] Cambanis *et al.* (1976) showed that if $k\ (Y_1, Y_0)$ is superadditive (or subadditive) then the extreme values of $E(\ k\ (Y_1, Y_0) \mid T = 1\ )$ are obtained by the upper- and lower bounding distributions.

only shifted by a constant factor. [11] We consider a model of heterogeneous impacts on the distribution that is estimated by quantile regression assuming the ranking property holds.

The advantage of quantile regression is that the impact of the program on different quantiles of the outcome of interest does not have to be constant. Thus, the estimation of quantile treatment effects allows testing the hypothesis that the treatment effect is the same for all points of the initial distribution of the outcome by testing whether the impacts are the same across quantiles of the control distribution. [12]

Quantile treatment effects (QTE) are a special case of quantile regression of the conditional mean of Y given $X = x$ where X is a discrete variable indicating whether the observation belongs to the treatment group or the control group. The quantile regression estimator minimizes a weighted sum of absolute residuals (Koenker and Bassett, 1978). Other experimental evaluations have used QTE, e.g. Heckman, Smith and Clements (1997), and Abadie, Angrist and Imbens (2002).

We present the estimates of the QTE, which are impacts conditional on the percentiles of Y_0 for the targeted outcomes: (1) wealth, proxied by the per capita value of consumption, (2) per capita value of food consumption. [13] The results on program impacts are estimated for November 1998, June 1999 and November 1999. Since the program started to send benefits in the summer 1998 and data on consumption is first collected in November 1998, the three cross-sections consist of post-program samples. Thus, we investigate quantile treatment effects in a simple difference model. In addition to estimating QTE on the outcomes *Y*, we also estimate QTE on $\tilde{Y}$ defined below:

$$Y = \alpha_0 + \sum_{i=1}^{K} X_i \alpha_i + v,$$

$$\tilde{Y} = y - [\hat{\alpha}_0 + \sum_{i=1}^{K} X_i \hat{\alpha}_i].$$

The residual $\tilde{Y}$ is obtained by removing the effect of household characteristics *X* from the outcome *Y*. Estimating QTE on $\tilde{Y}$ is similar to estimating experimental mean

[11] Note that the common effect model also assumes that household rankings with respect to the outcome of interest in the treated and untreated states are unaffected by the program.
[12] However, quantile treatment effects estimation does not provide information on the quantiles of the treatment effect distribution.
[13] QTE are also estimated for per capita expenditures and per capita food expenditures. Estimation results are available upon request.

program impacts conditioned on X. Note that QTE estimation on Y and QTE estimation on $\tilde{Y}$ are both consistent, but the latter is more efficient.

What should we expect? Both the Fréchet-Hoeffding lower bound of the variance of impacts and the QTE analysis correspond to the case of perfect positive dependence. The Fréchet-Hoeffding bounding analysis can inform us on the existence of heterogeneity in program impacts. The QTE estimation provides information on how the impact varies at different points of the untreated distribution, e.g. the first decile, the median, the last decile. The estimation of quantile treatment effects also allows us to estimate the variance of impacts over the quantiles of the untreated outcome distribution, which should be close to the lower-bound of the variance estimated using the Fréchet-Hoeffding inequalities.

Based on the theoretical framework, we expect the treatment effect on wealth and nutrition to be higher for the households whose cost of complying with the program requirements is the lower. In particular, children are required to attend school on a regular basis and their foregone earnings are an additional cost for households that relied on child labor before the start of the program.

5.3. Results and discussion.

Table 5 provides evidence of the heterogeneity of program impacts on per capita expenditures, per capita value of consumption, per capita food expenditures and per capita value of food consumption using the Fréchet-Hoeffding inequalities. The standard errors are obtained from the bootstrap (Efron and Tibshirani, 1994). We find that program impacts standard deviations can range anywhere from 4 to 12 pesos (minimum) to 130 to 260 pesos (maximum). These values can be compared to the average monthly per capita cash transfer amount that eligible households are entitled to receive when they fulfill program requirements, i.e. 32 pesos. [14] In all cases, we reject the null hypothesis that the minimum standard deviation is equal to zero at the 1% significance level. Using this non-parametric technique, we find that the experimental data are consistent with a large range of impact distributions.

Figure 2-7 present the difference in quantiles from the two marginal distributions of the wealth and nutrition aggregates conditional on a set of observable characteristics for each round. The associated 90 percent pointwise confidence intervals are obtained

[14] The average household size is 6.

from the bootstrap with 200 replications. Overall, the impacts are positive at each decile of the untreated distribution of outcomes.

For both outcomes, the difference overall increases from the lowest percentile to the highest percentile of the control group distribution. It suggests the program impact on wealth and nutrition is lower for households who were at a lower level of wealth and nutrition prior to the intervention.

In November 1998, the program impact on per capita value of consumption conditioned on X varies from about 3 pesos for the lowest decile of the untreated distribution to 9 pesos for the higher decile, i.e. increases by a factor three (Figure 2). This impact is low compared to the amount a PROGRESA household is eligible to receive upon compliance with program requirements. The impact at the median is about 7 pesos, compared to a mean impact of 8 pesos. This suggests that some households may not be getting the maximum benefit amount in November 1998, either because they have not fulfill all program requirements or because of implementation problems.

In June 1999, the impact ranges from 9 pesos to 28 pesos, with a median at 19 pesos and a mean at 17 pesos (Figure 3). Although the impact at the highest decile is close to the average per capita benefit amount, the impact at the first decile is still low. Thus, it is possible that the poorest households are still not getting the maximum benefit amount in June 1999. In November 1999, the impact ranges from 11 pesos to 22 pesos, with a median and mean at 16 pesos (Figure 4).

The mean impact between November 1998 and November 1999 has increased by a factor two, while the impact at the lowest decile increased by a factor four and the impact at the highest decile increased by less than a factor three in a year period. The impacts on per capita food consumption are similar in magnitude to the impacts of per capita consumption (Figures 5-7). This suggests that PROGRESA affects household total consumption mainly through higher food consumption.

Finally, Table 6 shows evidence of a variation in program impacts from the QTE analysis, except for the first round for which impacts are homogenous along the quantiles of the untreated distribution. In addition, Table 6 shows the program impact standard deviation from the QTE estimation. When the program impacts are found to be heterogeneous, impact standard deviations range from about 2 pesos to 6 pesos, which are similar to the lowest range of values estimated using the Fréchet-Hoeffding bounds.

6. Heterogeneous program impacts under the assumption of independence between program impacts and untreated outcome levels.

In this section as in the previous section, we also investigate the overall heterogeneity of program impacts, which includes both observed and unobserved heterogeneity. In this section, the main assumption is that the untreated outcome levels and program impacts are independent. This assumption means that households do not anticipate gains from the program at the time they decide to participate in the program. This is likely to occur when households are randomly assigned to a treatment and a control group. How plausible is the assumption in the case of PROGRESA? In the Mexican program, random assignment is not at the household level but at the village level and treatment group households can choose not to fulfill program requirements. We test the independence assumption. When this assumption holds and under additional assumptions, we can plot the distribution of impacts.

Note that if the assumption of perfect positive dependence holds, then program impacts should not vary along the quantiles of the untreated distribution. Yet, since program impacts are found to vary with the untreated distribution, the QTE results imply that the assumption of independence between program impacts and the untreated outcome levels does not hold. Yet, if the assumption of perfect positive dependence does not hold, then the variation in impacts along the quantiles of the untreated distribution is not consistent.

6.1. The Hildreth-Houck random coefficient model

As previously, denote by Y_0 the outcome in the untreated state, and Y_1 the outcome in the treated state. Any individual can only be observed in one or the other state. The observed outcome Y is a function of Y_0, Y_1 and the treatment indicator *T*. Suppose that the treatment effect varies for each household. We denote the household-specific program impacts by β_i. We assume that program impacts and untreated outcome levels Y_0 are uncorrelated. We have:

$$(2) \qquad Y_i = (1-T_i)\,Y_{i0} + T_i\,Y_{i1} = Y_{i0} + T_i\,(Y_{i1} - Y_{i0}).$$

Denote by $\hat{Y}_0$ the expected outcome for the control group given a vector of individual characteristics *X*, i.e. :

$$\hat{Y}_0 = E(Y \mid X, T = 0).$$

We have from equation (2) that:

$$T_i = 1 \quad \Rightarrow \quad Y_{i1} = \hat{Y}_0 + \overline{\beta} + v_{i1},$$
$$T_i = 0 \quad \Rightarrow \quad Y_{i0} = \hat{Y}_0 + v_{i0},$$

where $\overline{\beta}$ is the mean treatment effect, v_{i0} and v_{i1} are respectively the individual deviations of Y_{i0} and of Y_{i1} with respect to their means.

Thus, we have from equation (2):

$$Y_i = \hat{Y}_0 + T_i\ (\overline{\beta} + (v_{i1} - v_{i0})) + v_{i0}.$$

The household-specific program impacts β_i are as follows:

$$\beta_i = \overline{\beta} + (v_{i1} - v_{i0}).$$

Household-specific program impacts are such that:

$$\mathrm{E}(\beta_i) = \overline{\beta}, \quad \beta_i = \overline{\beta} + \sigma_i.$$

We assume that $\mathrm{E}(\sigma_i) = 0, \quad \mathrm{E}(v_{i0}) = 0.$

Equation (2) becomes:

$$(3) \qquad Y_i = \hat{Y}_0 + T_i\ \beta_i + v_{i0}.$$

Equation (3) has the structure of the Hildreth-Houck random coefficient model (1968). The independence between program impacts and untreated outcome levels implies that σ_i are independent of v_{i0}.

Let $\varepsilon_i = T_i\sigma_i + v_{i0}$. Then equation (3) can be written as follows:

$$Y_i = \hat{Y}_0 + T_i\ \overline{\beta} + \varepsilon_i.$$

The Hildreth-Houck model is a heteroscedastic error structure model:

$$Var(\varepsilon_i \mid T_i = 0) = Var(v_{i0}),$$
$$Var(\varepsilon_i \mid T_i = 1) = Var(\sigma_i + v_{i0}).$$

Under the assumption that σ_i and v_{i0} are uncorrelated, we have:

$$Var(\sigma_i + v_{i0}) = Var(\sigma_i) + Var(v_{i0}).$$

Once we assume that program impacts and the untreated outcome levels are independent, testing for heterogeneous program effect consists in testing that the variance of the error term depends on whether the household belongs to the treatment or the control group. This is done using the Breusch-Pagan test for heteroscedasticity and the LR test for groupwise heteroscedasticity (Judge *et al.* 1985). We treat the existence of negative values for $(Var(\sigma_i+v_{i0})\text{-}Var(v_{i0}))$ as a test of the assumption of independence between program impacts and untreated outcome levels. We also estimate the difference in the variance of the OLS residuals for the treatment and the control groups. Under the assumptions of the random coefficient model, this difference should represent the variance of impacts.

6.2. Additional assumptions on the distribution of program impacts.

If we assume that the impacts are normally distributed, then the estimation of the mean and variance of impacts is sufficient to plot the distribution of impacts. It also allows us to compute the percentage of the treated population that experienced a positive impact from the program.

Alternatively, we assume that the distribution of program impacts belongs to the Pearson family of distributions. This family of distributions includes as special cases the normal, chi-square, beta and gamma distributions. It only allows for one mode but includes bell curve shapes, as well as J-shaped or U-shaped curves (Kendall and Stuart, 1963). From the theoretical model that assumes that untreated outcome levels and gains from the program are uncorrelated, we have:

$$Y_{0i} + \beta_i = Y_{1i}.$$

Since we assume that Y_{0i} and β_i are uncorrelated, we can estimate by deconvolution the four first moments of the distribution of impacts using the moments of the distribution of Y_1 and the moments of the distribution of Y_0. [15] Note that the estimation of the four moment of the distribution of impacts by deconvolution allows to further test the independence assumption. Finding a negative value for the estimated fourth moment is an indication of failure of the independence assumption. The first four moments are then used to approximate the distribution of impacts, assuming that it belongs to the

[15] The derivation of the first four moments from deconvolution are in the Appendix.

Pearson family of distributions (see also Biddle, Boden and Reville 2003 who approximate the distribution of the effects of work-related injury on the subsequent earnings by a Pearson distribution).

Using the estimated first four moments, we test whether the estimated distribution exists. It can be shown that all frequency distributions satisfy the following relation between the second moment about the mean μ_2, the third moment about the mean μ_3 and the fourth moment about the mean μ_4:

$$\beta_2 - \beta_1 - 1 > 0,$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

Rejection is interpreted as a failure of the independence assumption. When the independence assumption is not rejected, this method allows us to obtain a density function for the distribution of impacts from the Pearson family of distributions.

6.3. Results and discussion.

All outcomes previously analyzed are per capita measures. We also estimate the random coefficient model using per capita outcomes although we do not report the estimates. [16] For per capita measures of wealth and nutrition, we find negative value for the estimated $(Var(\sigma_i + v_{i0}) - Var(v_{i0}))$. Thus, we reject the assumption of independence between program impacts and untreated outcome levels for per capita measures. These results are consistent with the variation in impacts along the quantiles of the untreated distribution found from the QTE analysis under the assumption of perfect positive dependence.

In Table 7, we present tests of heterogeneity in the variance of the error term using the Hildreth-Houck random coefficient model for household total wealth and nutrition outcomes. The first column in Table 7 provides the Breusch-Pagan test results and the second column provides the LR test for groupwise heteroscedasticity results. The results are consistent in both columns. We reject the null hypothesis of no heterogeneity in most cases. As in the case of the semi-parametric estimation, rejection is less frequent when using the first round of post-program data. The third column of Table 7 show the estimated standard deviation of impacts associated with a standard error. The estimated standard deviation of impacts is higher in the second round. The program impacts standard deviations vary from 31 to 178 pesos for household wealth and from 50 to 150 pesos for nutrition. Recall that these impacts concern household outcome levels and therefore cannot be directly compared to the impacts on per capita outcomes that are reported in the previous analyses. Yet, they can be compared to the average monthly benefit that a household is untitled to receive upon compliance with program requirements, i.e. about 200 pesos.

Using the estimated mean and variance of impacts and assuming a normal distribution, we plot the distribution of impacts on household consumption (Figure 8) and food consumption (Figure 9). The last column of Table 7 shows that the percentage of the treated population with a positive impact ranges from 63% to 81%. It is higher for the latest rounds. If we believe that the normality assumption holds, then this finding suggest that, in the first round, household may have supported the cost of the program requirements without receiving the program benefits because of implementation problems. Recall that in the empirical section, we consider the effect of the offer to treat which is likely to be a lower bound for the effect of the program on the treated population because some households may not be getting the program benefits.

We estimate the first four moments of the distribution of impacts. We find negative values for the estimated fourth moment for many of the outcomes, except consumption in November 1998 and November1999, food expenditures in November 1998 and November 1999 and food consumption in November 1999. This result is interpreted as a failure of the independence assumption. Furthermore, the first four moments for food expenditures in November 1999 do not correspond to a well-behaved frequency distribution. Thus, we are left with estimating Pearson distributions for the program impacts on four of the outcome variables. Yet, the third column in Table 7 shows that the standard deviation is not significantly different from zero for one of these four outcomes, i.e. November 1998 consumption.

We find that for three of the four outcomes, i.e. November 1998 consumption, June 1999 food expenditures and November 1999 food consumption, the distribution of impacts is approximated by an L-shaped Gamma distribution (Figures 10, 12 and 13).

---

[16] Estimates are available upon request.

First, this shape is clearly not similar to that of a normal distribution as assumed in Figure 8 and 9, although both shapes belong to the Pearson family. Second, all households have a positive program impacts on consumption and food consumption and expenditures. Third, these distributions of impacts have a vertical asymptote at less than 100 pesos. Most of the households have a small program impact and a few households have a large impact. Finally, the distribution of impacts on November 1999 consumption is approximated by a Type IV Pearson distribution (Figure 11). This distribution is less skewed than the estimated Gamma distribution for the other outcomes. Moreover, xx % of the population experienced negative impacts.

Failure of the independence assumption for most of the outcomes of interest cast doubt on the validity of the independence assumption. Some eligible households are likely to choose not to comply with the children school attendance requirement although they would still receive the food cash transfer. Thus, failure of the independence assumption is consistent with selective compliance to program requirements of the eligible population.

7. Conclusion

In this paper, we assess the importance of heterogeneity in impacts from the PROGRESA program. In the case of PROGRESA, experimental data help measure the effect of the "offer to treat". Thus, this commonly used experimental estimator under-estimates the value of the mean program impact for those who actually take-up program benefits. Finding a positive value is sufficient evidence that the program works.

Yet, theoretical predictions based on the program benefit scheme indicate that program impacts on wealth and nutrition are higher for households whose cost of complying with program requirements is lower. Program impacts on education are higher when the conditionality of program is binding. In addition, program impacts on welfare are likely to be small or even negative because of failures in the implementation of the program.

First, we find that the program selection mechanism is only partially effective in capturing the heterogeneity in program impacts. The geographic targeting is effective but we find little if any benefit from targeting within the poor villages. The proportion of beneficiaries with a positive program impact on wealth and nutrition ranges from 63% to

99% depending on the outcome considered. This proportion is the lowest in the first six months after the start of the program. This suggests that some households may have bear the cost of complying without receiving the benefits because of early implementation delays in sending the cash transfers. We also find that secondary school-age children from the most remote areas get the lowest program impact. This suggests that many of them may have chosen not to attend school and do not get the PROGRESA grant.

Second, we investigate the overall heterogeneity of program impacts, which includes both unobserved and systematic heterogeneity. Methodologically, experimental data help identify subgroups effects such as heterogeneity along the targeting criteria, but do not directly identify unobserved heterogeneity in impacts. Using the Fréchet-Hoeffding inequalities from classical probability theory, we find evidence against the homogeneous impact assumption as in Heckman, Smith and Clements (1997). This result does not rely on any assumption and thus is particularly strong evidence of heterogeneous treatment effects. Yet, many distributions of impacts are consistent the estimated variance of impacts from the bounding analysis.

Additional assumptions are required to analyze the distribution of impacts. We consider two assumptions. The first one concerns perfect positive dependence between potential outcomes in the treatment and non-treatment state. The second relates to whether program participants anticipate the gains from the program. The second assumption is that program impacts and untreated outcome levels are independent.

Under both assumptions, we find evidence that the program does not have the same impact on everyone. Thus, we reject the “common effect” assumption underlying most of the impact evaluation research as do Heckman, Smith and Clement (1997), Abadie, Angrist and Imbens (2002), Black, Smith, Berger and Noel (2003), Biddle, Boden and Reville (2003), Bitler, Gelbach and Hoynes (2004) for different treatment effects in the US context.

Under the former assumption, we estimate distribution of program impacts along the quantiles of the untreated distribution. As expected, we find that program impacts on wealth and nutrition are greater for the households who were at a higher level of wealth and nutrition prior to the intervention. We reject the latter assumption for most of the outcomes. This is consistent with the prediction from the analysis of the conditionality effect, i.e. of a selective compliance to program requirements.

References

Abadie, Alberto, Joshua Angrist and Guido Imbens. 2002. "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings." *Econometrica*, 70(1): 91-117.

Behrman, Jere and Petra Todd. 1999. "Randomness in the Experimental Samples of PROGRESA." International Food Policy Research Institute, Washington D.C..

Black, Dan A., Jeffrey A. Smith, Mark C. Berger and Brett J. Noel. 2003. "Is the threat of reemployment services more effective than the services themselves? Experimental evidence from the UI system." *American Economic Review*, 93 (3): 1313-1327.

Biddle, Jeff, Les Boden and Robert Reville. 2003. "A method for estimating the full distribution of a treatment effect, with application to the impact of workfare injury on subsequent earnings." Mimeo.

Bitler, Marianne, Jonah Gelbach and Hilary W. Hoynes. 2004. "What mean impact miss: distributional effects of welfare reform experiments." Mimeo.

Cambanis, Stamatis, Gordon Simons and William Stout. 1976. "Inequalities for *E(k (X, Y))* when the marginals are fixed." *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*. 36(4): 285-294.

Coady, David and Habiba Djebbari. 1999. "A preliminary process evaluation of the Education, Health and Nutrition program (PROGRESA) of Mexico." The International Food Policy Research Institute, Washington D.C..

Deaton, Angus. 1997. "The analysis of household surveys: a microeconometric approach to development policy." Johns Hopkins University Press: 25-32.

Efron, Bradley and Robert Tibshirani. 1994. "An introduction to the bootstrap." CRC Press.

Fréchet, Maurice. 1951. "Sur les tableaux de corrélation dont les marges sont données." Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie. 14: 53-77.

Heckman, James J. and Jeffrey A. Smith. 1995. "Assessing the case for social experiments." *The Journal of Economic Perspectives*. 9(2): 85-110.

Heckman, James J., Jeffrey Smith and Nancy Clements. 1997. "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in program impacts." *The Review of Economic Studies*. 64(4): 487-535.

Hildreth, Clifford and James P. Houck. 1968. "Some estimators for a linear model with random coefficients." *Journal of the American Statistical Association*. 63(322): 584-595.

Hoeffding, Wassily. 1940. "Scale-invariant correlation theory." In N.I. Fisher and P.K. Sen, The collected works of Wassily Hoeffding, Springer Series in Statistics: Perspective in Statistics, 1994: 57-107.

Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lutkepohl, Tsoung-Chao Lee. 1985. "The theory and practice of Econometrics." 2nd Edition.

Kendall, Maurice and Stuart Alan. 1963. "The theory of advanced statistics. Volume 1: Distribution theory." 2nd edition.

Koenker, R. and G. Basset. 1978. "Regression quantiles." *Econometrica*, 46: 33-50.

Skoufias, Emmanuel, Benjamin Davis and Sergio de la Vega. 2001. "Targeting the poor in Mexico: an evaluation of the selection of households for PROGRESA." *World Development*. 29(10): 1769-1784.

Skoufias, Emmanuel, Susan Parker. 2001. "Conditional Cash Transfers and their Impact on Child Work and School Enrollment: Evidence from the PROGRESA program in Mexico." *Economia*. 2 (1): 45-96.

Todd, Petra and Kenneth I. Wolpin. 2003. "Using experimental data to validate a dynamic behavioral model of child schooling and fertility: assessing the impact of a school subsidy program in Mexico." Mimeo.

Table 1: Program Impacts on wealth and nutrition along program targeting criteria.

	November 1998	June 1999	November 1999
P.C. Consumption			
T	10.09 (15.26)	38.69** (16.27)	61.87*** (12.76)
T*Pindex	0.01 (0.02)	0.03 (0.02)	0.07*** (0.02)
T*Vindex	91.71*** (10.83)	46.44*** (11.84)	28.29*** (8.54)
T*Pindex*Vindex	0.12*** (0.02)	0.06*** (0.02)	0.04*** (0.01)
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$ p-value	0.0001	0.0001	0.0001
P.C. Expenditures			
T	6.75 (12.89)	32.16** (14.52)	39.65*** (10.99)
T*Pindex	0.01 (0.02)	0.03 (0.02)	0.05*** (0.02)
T*Vindex	97.19*** (8.84)	58.60*** (10.93)	49.58*** (7.41)
T*Pindex*Gindex	0.14*** (0.01)	0.09*** (0.02)	0.06*** (0.01)
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$ p-value	0.0001	0.0001	0.0001
P.C. Food Consumption			
T	18.86* (11.57)	28.79*** (10.58)	41.31*** (9.86)
T*Pindex	0.02 (0.02)	0.03* (0.02)	0.04*** (0.01)
T*Vindex	52.10*** (7.76)	17.64** (7.82)	8.12 (6.37)
T*Pindex*Vindex	0.07*** (0.01)	0.02 (0.01)	0.01 (0.01)
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$ p-value	0.0001	0.0001	0.0001
P.C. Food Expenditures			
T	13.73 (8.89)	22.17*** (7.47)	22.11*** (6.97)
T*Pindex	0.02 (0.01)	0.02* (0.01)	0.02** (0.01)
T*Vindex	60.29*** (5.86)	23.61*** (6.16)	33.17*** (4.97)
T*Pindex*Vindex	0.09*** (0.01)	0.03*** (0.01)	0.04*** (0.01)
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$ p-value	0.0001	0.0001	0.0001

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Robust Standard Errors in parentheses.

Table 2: Fraction with a positive program impact on wealth and nutrition using treatment effects on subgroups defined by the targeting criteria

	H_o: any program impact p-value	Fraction with a positive program impact
November 1998		
P.C. Consumption	0.0001	0.81
P.C. Expenditures	0.0001	0.67
P.C. Food Consumption	0.0001	0.83
P.C. Food Expenditures	0.0001	0.76
June 1999		
P.C. Consumption	0.0001	0.992
P.C. Expenditures	0.0001	0.97
P.C. Food Consumption	0.0001	0.9995
P.C. Food Expenditures	0.0001	0.9975
November 1999		
P.C. Consumption	0.0001	0.96
P.C. Expenditures	0.0001	0.93
P.C. Food Consumption	0.0001	0.987
P.C. Food Expenditures	0.0001	0.987

Table 3: Program Impacts on boys' time allocation outcomes.

	T Coefficient (std. err.)	T*Pindex Coefficient (std. err.)	T*Vindex Coefficient (std. err.)	T*Pindex*Vindex Coefficient (std. err.)	Fraction with a positive program impact	H_o: any impact p-value $H_0: \beta_1 = \beta_2 = \beta_3 = 0.$ p-value
Primary school age boys						
Participation in school	0.064 (.085)	0.00006 (0.0001)	-0.032 (0.063)	-0.00006 (0.00009)	0.97	0.413 0.816
Time spent in school	34.52 (53.39)	0.037 (0.078)	-20.12 (39.84)	-0.043 (0.062)	0.96	0.436 0.796
Participation in labor activities	0.036 (.033)	0.0001* (0.00005)	-0.074** (0.03)	-0.0001*** (0.00004)	0.17	0.001*** 0.004***
Time spent working	148 (178)	0.39 (0.27)	-376*** (143)	-0.74*** (0.22)	0.15	0.0009*** 0.002***
Participation in domestic activities	0.11 (0.07)	0.0001 (0.0001)	-0.16*** (0.06)	-0.0002*** (0.0001)	0.23	0.084* 0.055*
Time spent in domestic activities	21.92 (33.45)	0.033 (0.049)	-46.8* (25.1)	-0.072* (0.039)	0.51	0.3195 0.319
Secondary school age boys						
Participation in school	0.038 (.09)	0.00006 (0.0001)	-0.148** (0.069)	-0.0003*** (0.0001)	0.65	0.007*** 0.005***
Time spent in school	23.8 (76.5)	0.043 (0.11)	-136** (58)	-0.27*** (0.08)	0.63	0.005*** 0.003***
Participation in labor activities	0.011 (0.069)	0.00007 (0.0001)	-0.095* (0.053)	-0.0001** (0.00008)	0.013	0.007*** 0.025**
Time spent working	28 (142)	0.15 (0.20)	-210* (109)	-0.36** (0.16)	0.02	0.007*** 0.186
Participation in domestic activities	0.11 (0.08)	0.0001 (0.0001)	0.004 (0.067)	0.00001 (0.0001)	0.47	0.691 0.524
Time spent in domestic activities	14.6 (36)	0.014 (0.052)	21.5 (27.5)	0.037 (0.042)	0.85	0.771 0.771

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Robust standard error in parentheses. In the participation equations, coefficients reported measure change in the probability of participating.

Table 4: Program Impacts on girls' time allocation outcomes.

	T Coefficient (std. err.)	T*Pindex Coefficient (std. err.)	T*Vindex Coefficient (std. err.)	T*Pindex*Vindex Coefficient (std. err.)	Fraction with a positive program impact	H_o: any impact p-value $H_0 : \beta_1 = \beta_2 = \beta_3 = 0.$ p-value
Primary school age girls						
Participation in school	0.013 (.087)	0.00003 (0.0001)	-0.012 (0.063)	-0.00005 (0.0001)	0.58	0.732 0.610
Time spent in school	1.97 (53.41)	0.0007 (0.078)	-16.38 (39.35)	-0.05 (0.06)	0.76	0.428 0.478
Participation in labor activities	0.026 (.026)	0.00003 (0.00004)	-0.04* (0.02)	-0.00005 (0.00003)	0.90	0.238 0.354
Time spent working	207 (209)	0.19 (0.31)	-289* (151)	-0.39 (0.24)	0.89	0.169 0.232
Participation in domestic activities	0.08 (0.08)	0.0001 (0.0001)	-0.22*** (0.06)	-0.0003*** (0.0001)	0.37	0.01*** 0.005***
Time spent in domestic activities	46 (40)	0.06 (0.06)	-108*** (30)	-0.15*** (0.04)	0.64	0.009*** 0.003***
Secondary school age girls						
Participation in school	0.26*** (.08)	0.0003** (0.0001)	-0.26*** (0.069)	-0.0004*** (0.0001)	0.93	0.0001*** 0.0008***
Time spent in school	239*** (82)	0.28** (0.11)	-230*** (63)	-0.36*** (0.09)	0.92	0.0001*** 0.001***
Participation in labor activities	0.003 (0.003)	0.00003 (0.00005)	-0.034 (0.029)	-0.00003** (0.00004)	0.46	0.496 0.340
Time spent working	125 (223)	0.14 (0.32)	-168 (182)	-0.14 (0.28)	0.45	0.593 0.437
Participation in domestic activities	0.013 (0.09)	0.00008 (0.0001)	-0.11 (0.06)	0.00017* (0.0001)	0.001	0.03** 0.412
Time spent in domestic activities	-38 (45)	-0.027 (0.066)	-39 (35)	-0.06 (0.05)	0.008	0.04** 0.561

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Robust standard error in parentheses. In the participation equations, coefficients reported measure change in the probability of participating.

Table 5: Bounds on the Variation of the Impacts using the Fréchet-Hoeffding Inequalities

Impact Standard Deviation (Standard error)	From the upper bound distribution	From the lower bound distribution
November 1998		
Per capita household expenditures	4.8*** (.02)	221*** (1.1)
Per capita value of consumption	12.8*** (0.28)	258*** (1.3)
Per capita food expenditures	4.36*** (.07)	161*** (0.8)
Per capita value of food consumption	6.32*** (0.13)	202*** (1.1)
June 1999		
Per capita household expenditures	8.49*** (0.23)	203*** (1.1)
Per capita value of consumption	7.68*** (0.09)	241*** (1.3)
Per capita food expenditures	7.36*** (0.15)	137*** (0.7)
Per capita value of food consumption	8.42*** (.08)	181*** (1.07)
November 1999		
Per capita household expenditures	6.82*** (0.15)	193*** (0.98)
Per capita value of consumption	6.20*** (0.12)	226*** (1.2)
Per capita food expenditures	3.33*** (0.04)	128*** (0.6)
Per capita value of food consumption	3.81*** (.18)	163*** (0.8)

Note: *** 1% significance level, Bootstrap S.E. in parentheses (500 replications). Results obtained using the percentiles of the two empirical c.d.f.s.

Table 6: Standard Deviation of Impacts from the QTE estimation

	H_o: any program impact p-value*	Standard Deviation of program impacts
November 1998		
P.C. Consumption	0.2665	2.16
P.C. Expenditures	0.5834	1.51
P.C. Food Consumption	0.1355	1.87
P.C. Food Expenditures	0.7935	0.69
June 1999		
P.C. Consumption	0.0001	5.67
P.C. Expenditures	0.0002	3.75
P.C. Food Consumption	0.0001	5.11
P.C. Food Expenditures	0.0001	3.15
November 1999		
P.C. Consumption	0.0049	3.31
P.C. Expenditures	0.0216	2.07
P.C. Food Consumption	0.0001	4.36
P.C. Food Expenditures	0.0051	1.86

The set of covariates X includes household size, household composition and characteristics of the head of household.
* p-value associated with the null hypothesis that impact is constant across deciles of the untreated distribution.

Table 7: Testing the presence of heteroscedasticity from the Hildreth-Houck Random Coefficient Model for Household-level Outcome Variables.

	Breusch-Pagan LM-stat (p-value)	LR-stat for groupwise heteroscedasticity (p-value)	S.D.(Δ)	Percentage with a positive impact †
Household Expenditures November 98	0.067 (0.79)	0.4 (0.527)	31.63 (74.4)	81
Household Expenditures June 99	8.78*** (0.003)	11.39*** (0.001)	107* (77.1)	77
Household Expenditures November 99	1.18 (0.27)	0.45 (0.5)	60.77 (71.5)	71
Household value of consumption November 98	5.83*** (0.01)	7.08*** (0.008)	118 (99)	63
Household value of consumption June 99	25.58*** (0.0001)	33.78*** (0.0001)	178*** (106)	72
Household value of consumption November 99	11.58*** (0.0001)	4.179** (0.041)	132* (98)	76
Household food expenditures November 98	1.67 (0.19)	2.605 (0.107)	50.45 (51)	66
Household food expenditures June 99	40.87*** (0.0001)	54.71*** (0.0001)	107*** (53)	71
Household food expenditures November 99	27.08*** (0.0001)	15.47*** (0.0001)	89*** (46)	74
Household value of food consumption November 98	3.82** (0.05)	4.126** (0.042)	84 (84)	64
Household value of food consumption June 99	37.46*** (0.0001)	49.57*** (0.0001)	149*** (87)	70
Household value of food consumption November 99	54.56*** (0.0001)	43.39*** (0.0001)	147*** (81)	71

Note: *** indicates 1% significance level, ** indicates 5% significance level.

† Percentage with a positive impact is derived assuming a normal distribution.

Figure 1: Types of households based on their indirect utility levels

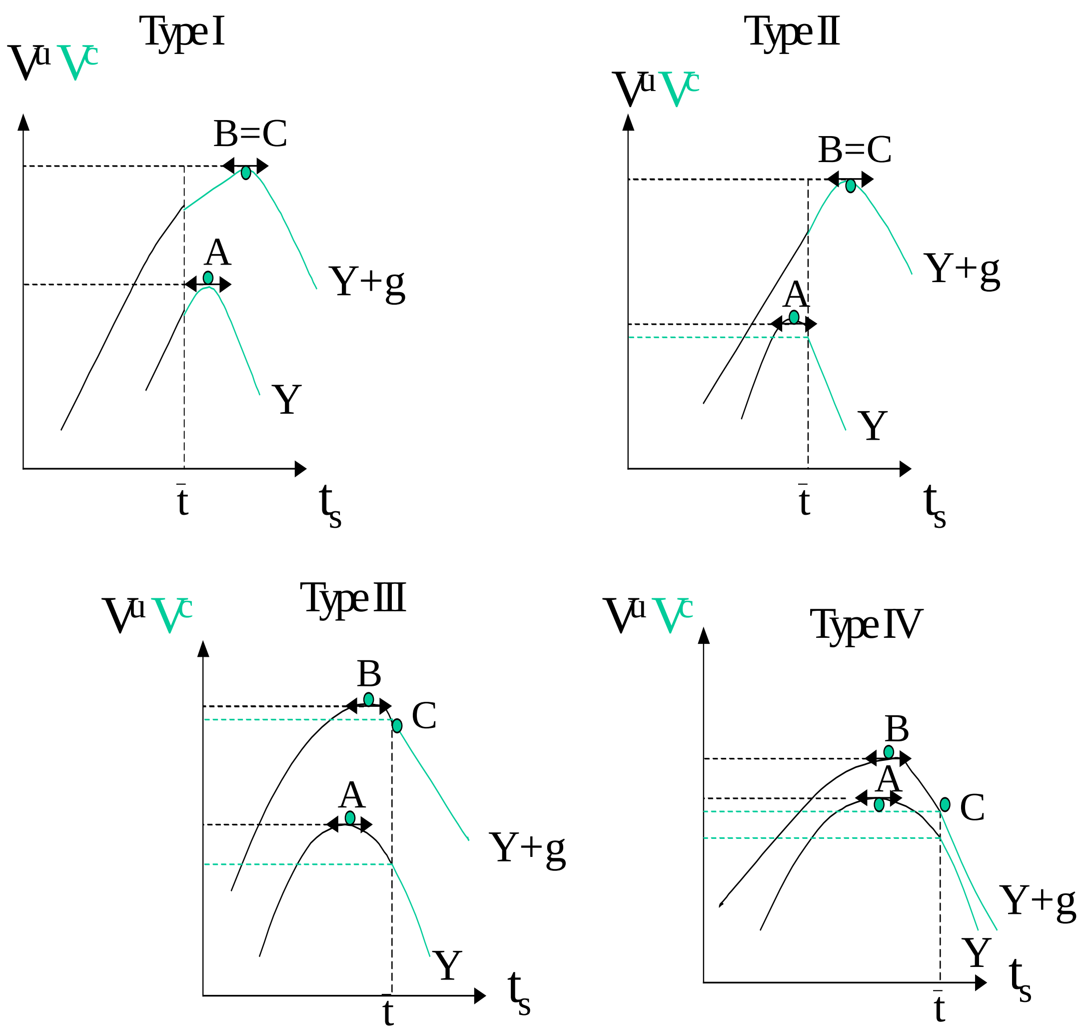


Figure 2: Difference in per capita value of consumption in the treatment and control group in November 1998 (Nov 1998 pesos) controlling for covariates

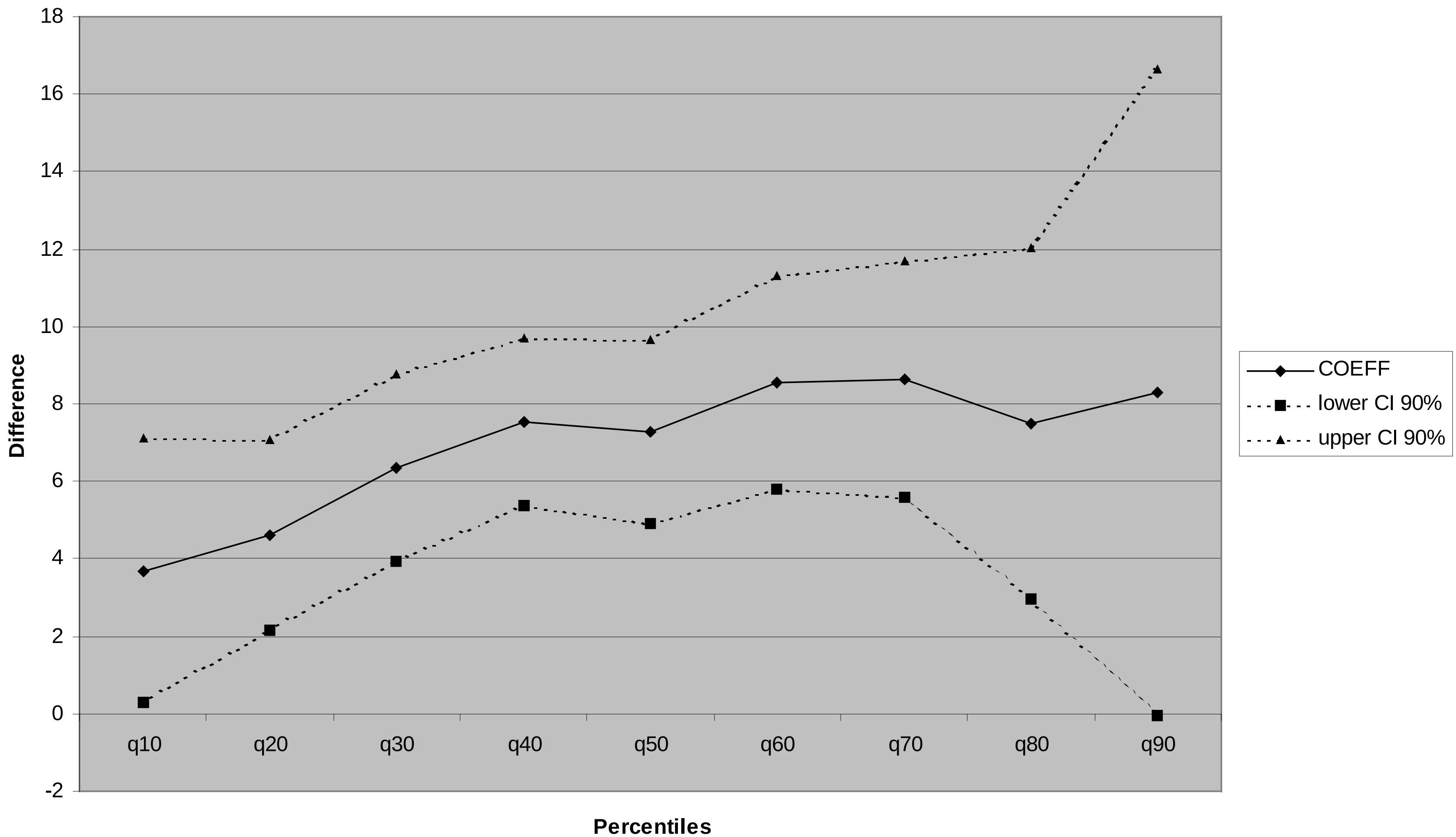


Figure 3: Difference in per capita value of consumption in treatment and control in June 1999 (Nov. 1998 pesos) controlling for covariates

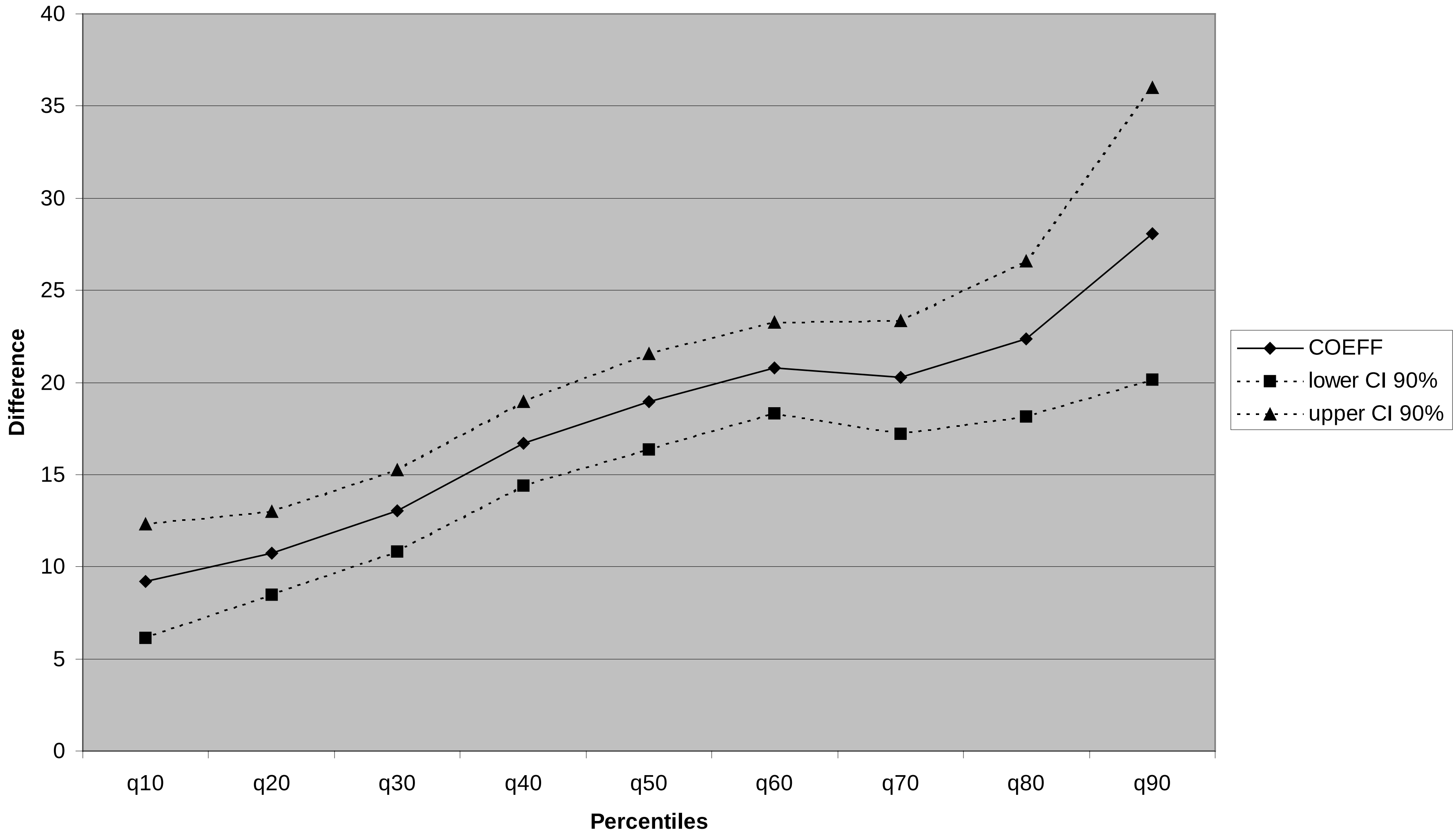


Figure 4: Difference in per capita value of consumption in treatment and control group in November 1999 (Nov. 1998 pesos) controlling for covariates

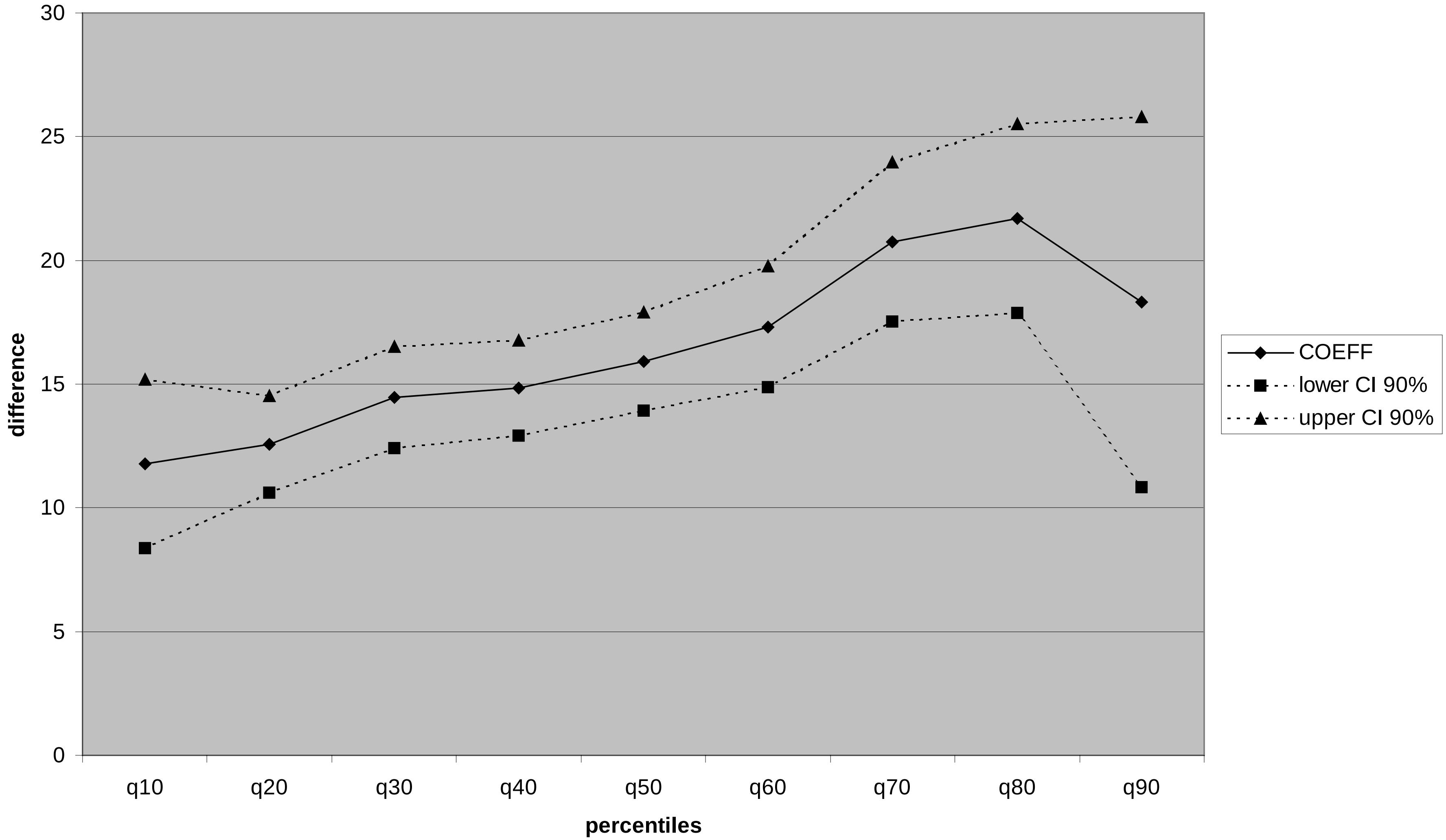


Figure 5: Differences in per capita value of food consumption in treatment and control groups in November 1998 (in Nov. 98 pesos) controlling for covariates

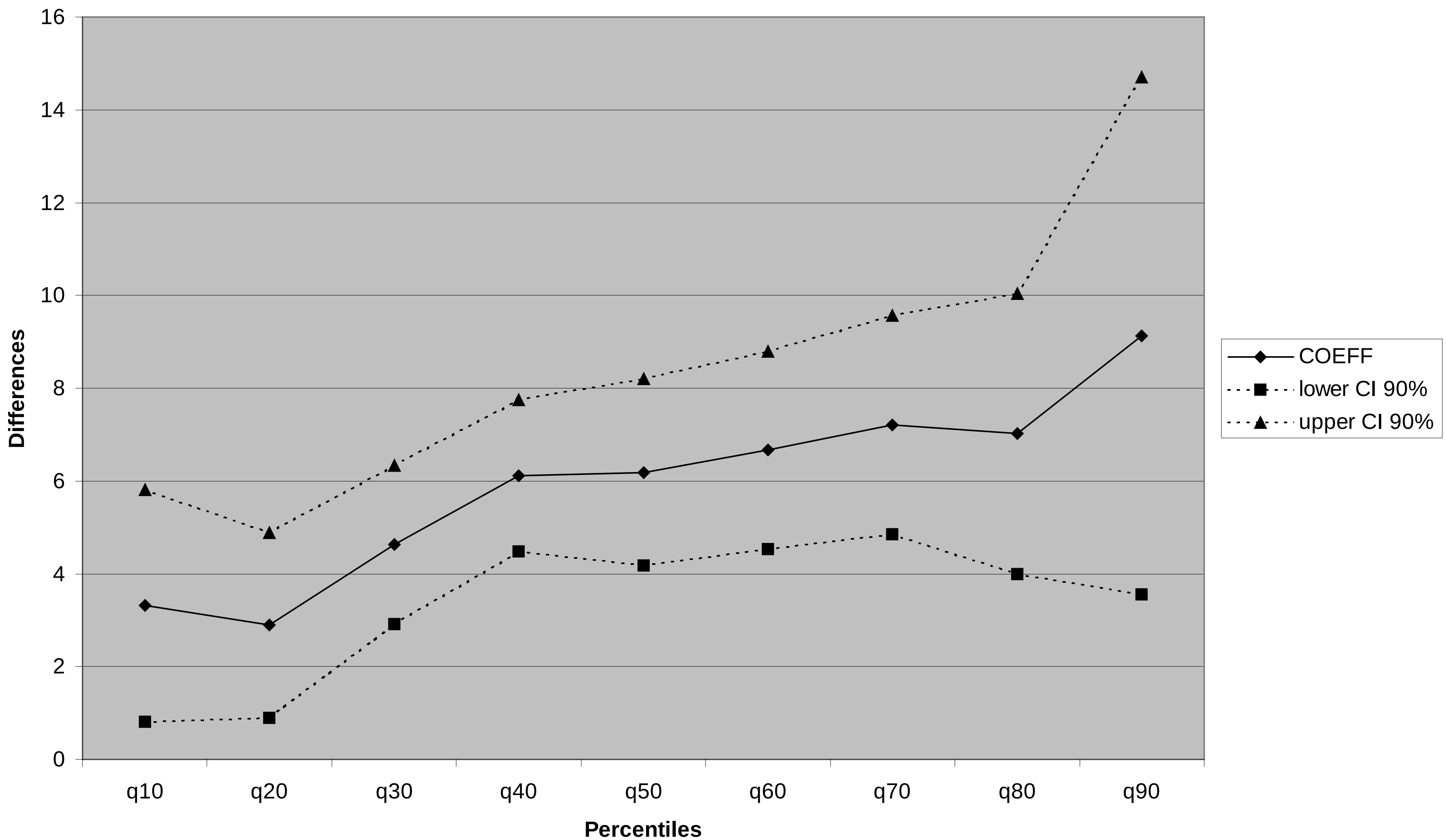


Figure 6: Difference in per capita value of food consumption in treatment and control groups in June 1999 (in Nov 98 pesos) controlling for covariates

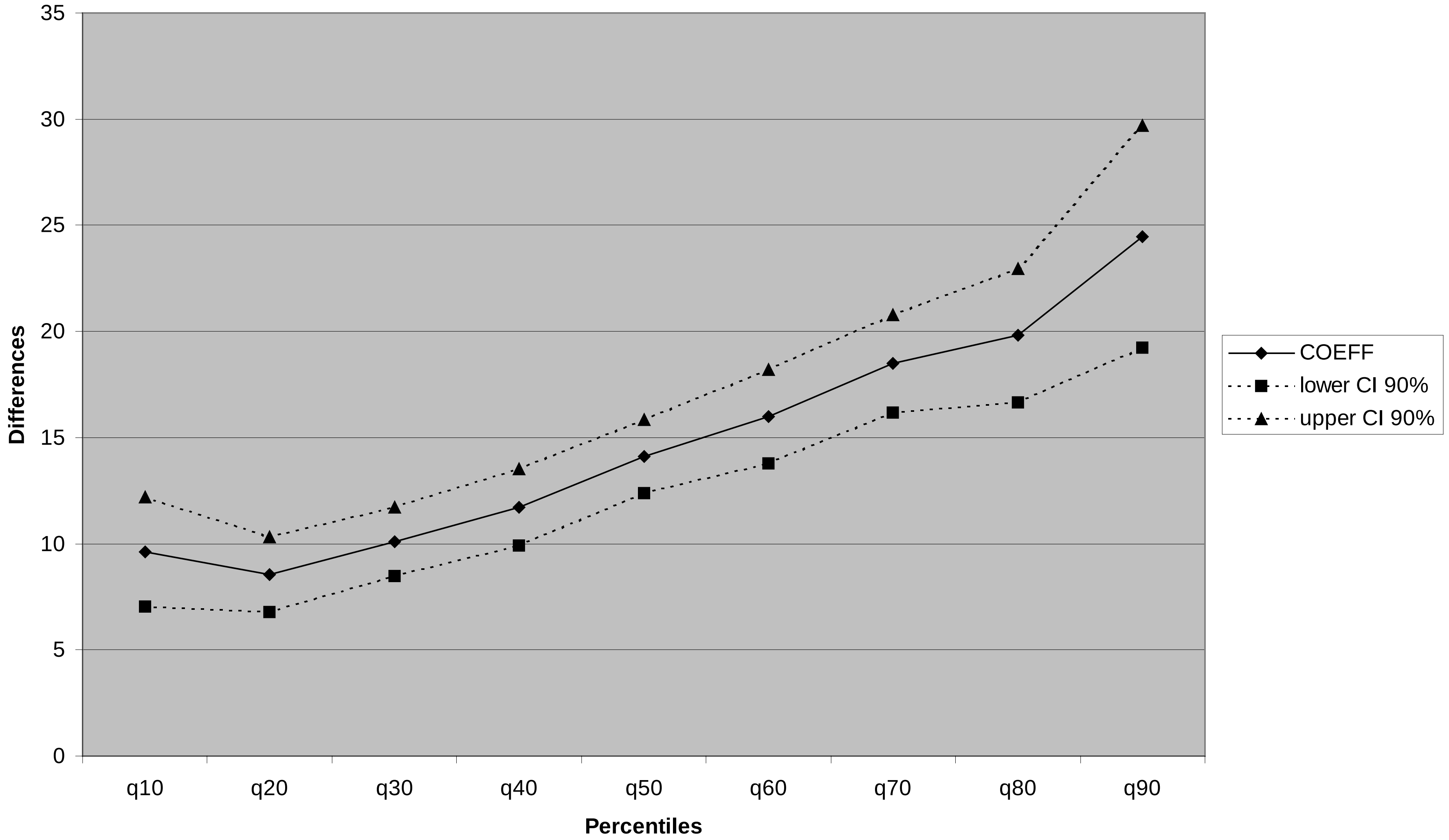


Figure 7: Difference in per capita of food consumption in treatment and control groups in November 1999 (in Nov. 98 pesos) controlling for covariates

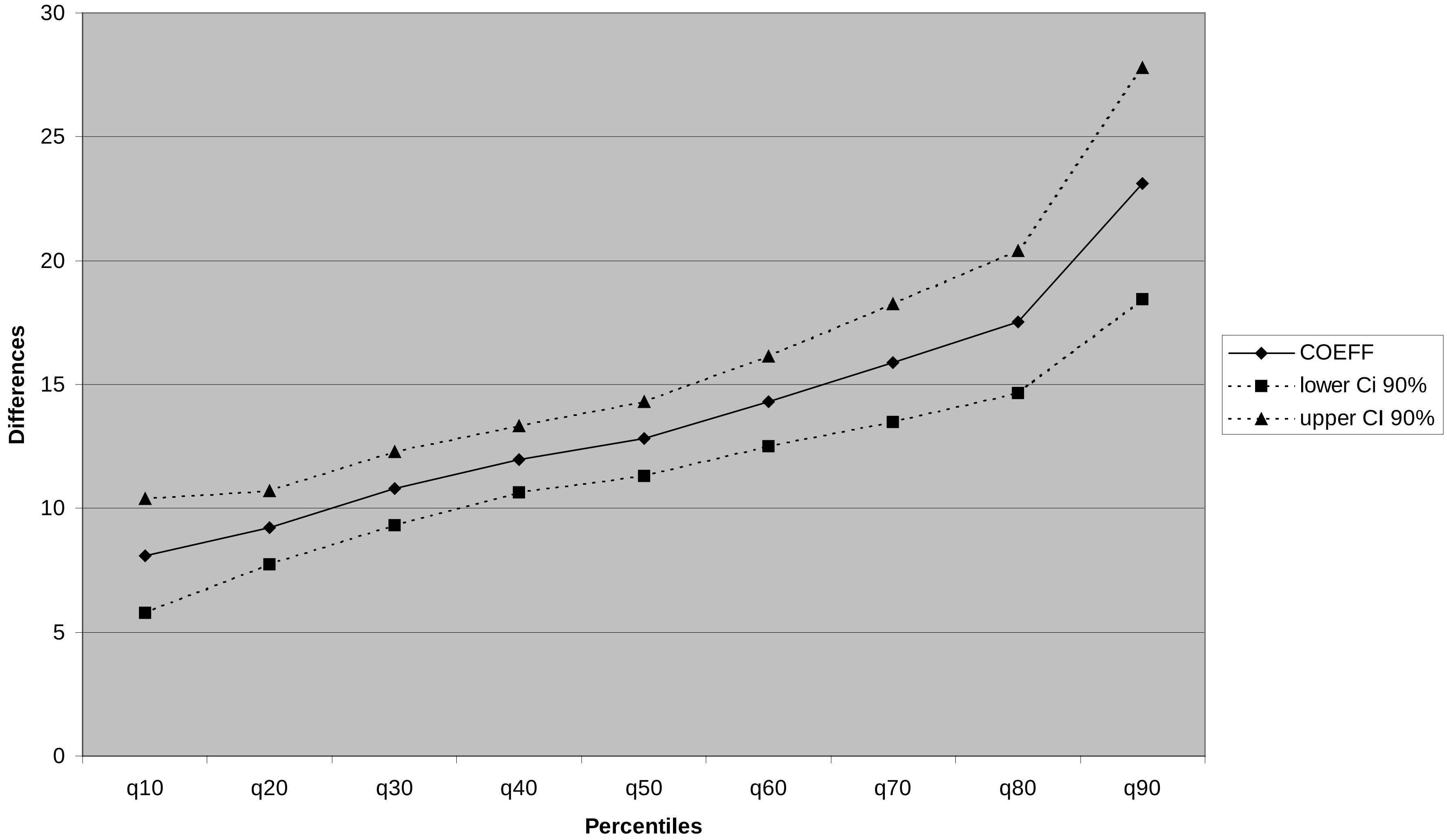


Figure 8: Distribution of impacts on total household consumption under the normality assumption

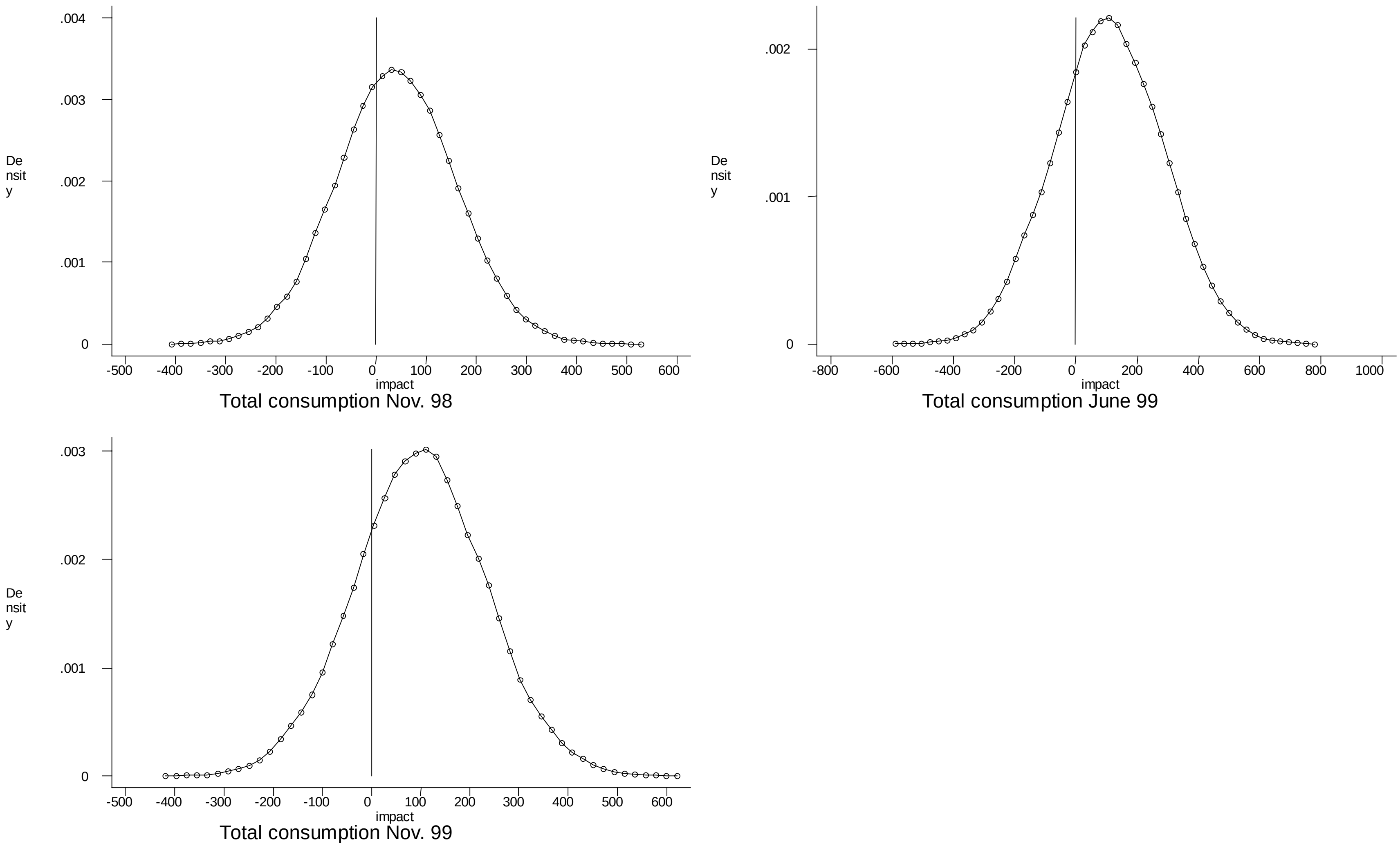


Figure 9: Distribution of impacts on food consumption under the normality assumption

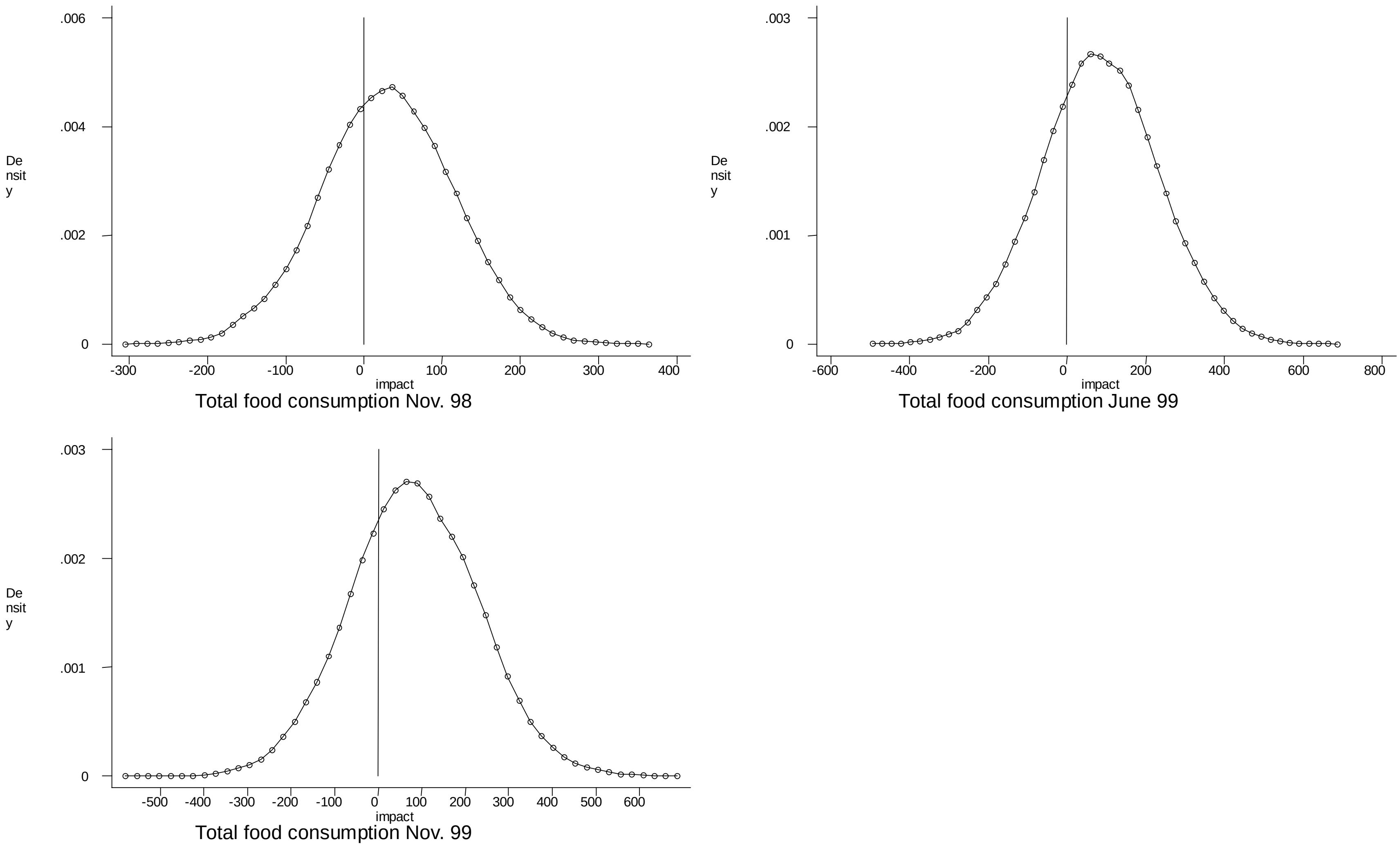
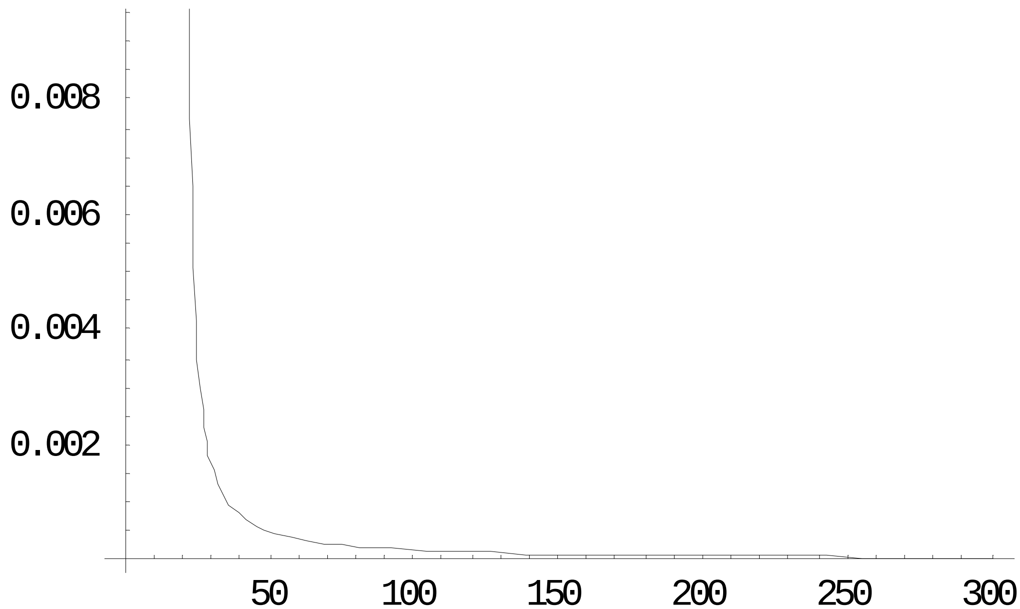
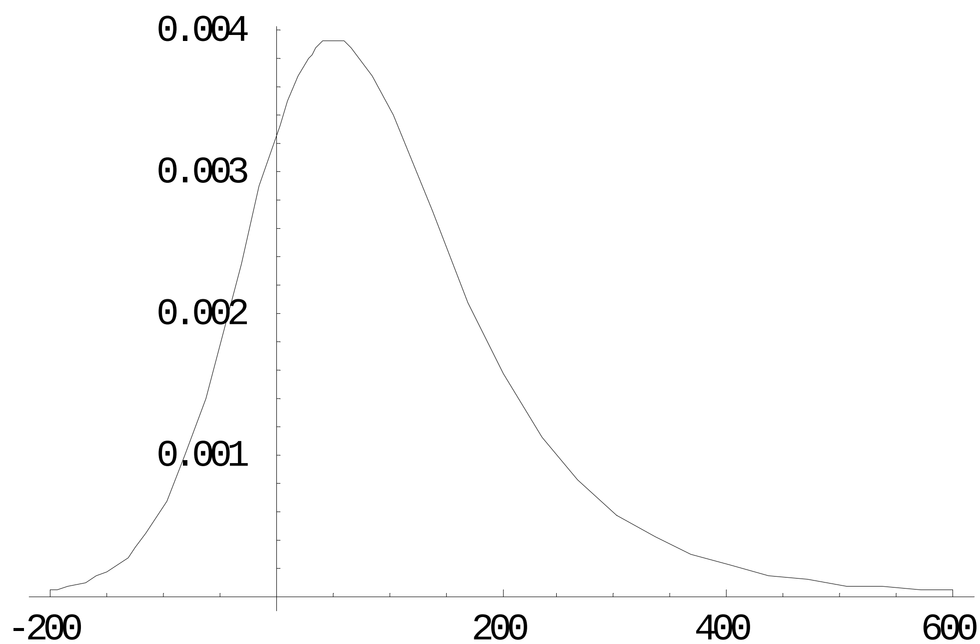


Figure 10: Distribution of impacts on November 1998 household consumption – Gamma distribution.



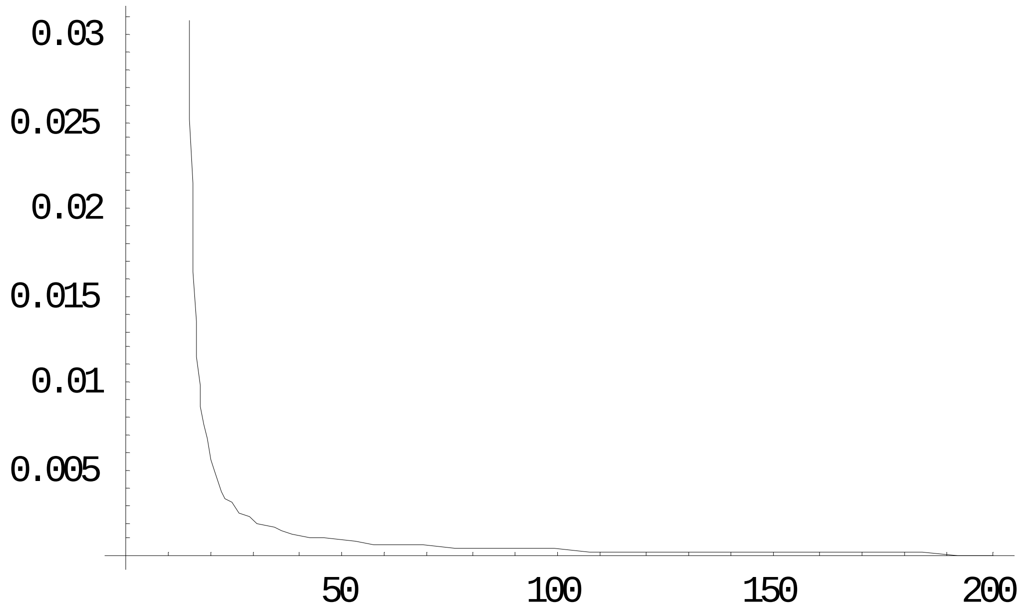
On the x-axis are program impacts in pesos, and on the y-axis the associated frequencies.

Figure 11: Distribution of impacts on November 1999 household consumption – Type IV Pearson distribution.



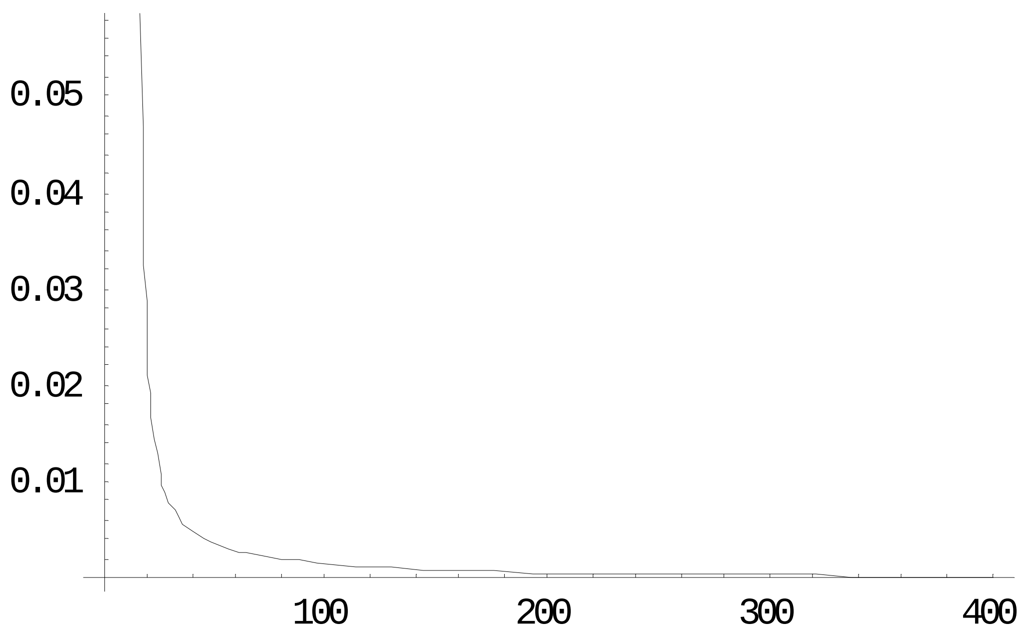
On the x-axis are program impacts in pesos, and on the y-axis the associated frequencies.

Figure 12: Distribution of impacts on June 1999 food expenditures – Gamma distribution.



On the x-axis are program impacts in pesos, and on the y-axis the associated frequencies.

Figure 13: Distribution of impacts on November 1999 food consumption – Gamma distribution.



On the x-axis are program impacts in pesos, and on the y-axis the associated frequencies.

Appendix: Deriving Moments of a Distribution from Deconvolution

Let X and Y be two independent random variables and let Z be the convolution of X and Y. Suppose the first four moments of Z and Y are known. The objective is to derive the first four moments of X as a function of the first four moments of Z and Y.

Let μ_X, μ_Y and μ_Z be the first moment of X, Y and Z, such that:

$$\begin{aligned}
\mu_X &= E(X),\\
\mu_Y &= E(Y),\\
\mu_Z &= E(Z).
\end{aligned}$$

Let μ_{2X}, μ_{2Y} and μ_{2Z} be the second moments of X, Y and Z about their means, such that:

$$\begin{aligned}
\mu_{2X} &= E((X-\mu_X)^2),\\
\mu_{2Y} &= E((Y-\mu_Y)^2),\\
\mu_{2Z} &= E((Z-\mu_z)^2).
\end{aligned}$$

Let μ_{3X}, μ_{3Y} and μ_{3Z} be the third moments of X, Y and Z about their means, such that:

$$\begin{aligned}
\mu_{3X} &= E((X-\mu_X)^3),\\
\mu_{3Y} &= E((Y-\mu_Y)^3),\\
\mu_{3Z} &= E((Z-\mu_z)^3).
\end{aligned}$$

Let μ_{4X}, μ_{4Y} and μ_{4Z} be the fourth moments of X, Y and Z about their means, such that:

$$\begin{aligned}
\mu_{4X} &= E((X-\mu_X)^4),\\
\mu_{4Y} &= E((Y-\mu_Y)^4),\\
\mu_{4Z} &= E((Z-\mu_z)^4).
\end{aligned}$$

<u>Mean of *X*.</u>

$$\begin{aligned}
&\text{If } Z = X+Y, \quad X \perp Y, \text{ then :}\\
&\mu_Z = E(Z) = E(X+Y) = E(X) + E(Y).
\end{aligned}$$

Thus, $\mu_X = E(X) = \mu_Z - \mu_Y$.

Variance of *X*.

If $Z = X + Y,\quad X \perp Y,$ then :
$$\mu_{2Z} = E((Z-\mu_Z)^2) = E((X+Y-\mu_X-\mu_Y)^2) = E(((X-\mu_X)+(Y-\mu_Y))^2),$$
$$\mu_{2Z} = \mu_{2X} + \mu_{2Y} + 2\,E(X-\mu_X)\,E(Y-\mu_Y) = \mu_{2X} + \mu_{2Y}.$$

Thus, $\mu_{2X} = \mu_{2Z} - \mu_{2Y}.$

Third moment of *X* about the mean.

If $Z = X + Y,\quad X \perp Y,$ then :
$$\mu_{3Z} = E((Z-\mu_Z)^3) = E((X+Y-\mu_X-\mu_Y)^3) = E(((X-\mu_X)+(Y-\mu_Y))^3),$$
$$\mu_{3Z} = \mu_{3X} + \mu_{3Y} + 3\,E((X-\mu_X)^2)\,E(Y-\mu_Y) + 3E((Y-\mu_Y)^2)\,E(X-\mu_X),$$
$$\mu_{3Z} = \mu_{3X} + \mu_{3Y}.$$

Thus, $\mu_{3X} = \mu_{3Z} - \mu_{3Y}.$

Fourth moment of *X* about the mean.

If $Z = X + Y,\quad X \perp Y,$ then :
$$\mu_{4Z} = E((Z-\mu_Z)^4) = E((X+Y-\mu_X-\mu_Y)^4) = E(((X-\mu_X)+(Y-\mu_Y))^4),$$
$$\mu_{4Z} = \mu_{4X} + \mu_{4Y} + 4\,E((X-\mu_X)^3)\,E(Y-\mu_Y) + 4\,E((Y-\mu_Y)^3)\,E(X-\mu_X)$$
$$+6\,E((X-\mu_X)^2)\,E((Y-\mu_Y)^2),$$
$$\mu_{4Z} = \mu_{4X} + \mu_{4Y} + 6\mu_{2X}\,\mu_{2Y}.$$

Thus, $\mu_{4X} = \mu_{4Z} - \mu_{4Y} - 6\mu_{2X}\,\mu_{2Y}.$