DET SAMFUNDSVIDENSKABELIGE FAKULTET
KØBENHAVNS UNIVERSITET

Advanced Microeconometrics

Goutham Jørgen Surendran & Nick Fabrin Nielsen

Weighted Quantile Decomposition

- with application to changes in South African living standards

Daniel le Maire

December 11, 2014

Abstract. This paper examines the effects of the abolition of apartheid on the distribution of household expenditures for colored South Africans. First, we use an unbalanced panel of colored South African households to replicate the results in Vahid & Maitra (2006). This implies estimating the conditional distribution of expenditures by IPW quantile regression. Second, we extend the analysis of Vahid & Maitra (2006) by decomposing the change in the unconditional expenditure distribution into changes in coefficients, covariates and residuals. To do this we use the method suggested by Melly (2005) extended for use with IPW, this approach is novel. We find that the decrease in the living standards for the lowest decile of colored South Africans can be attributed mainly to changes in characteristics while the increase in the living standards for the top deciles can be attributed mainly to changes in coefficients. Finally, we investigate the IPW-decomposition in a Monte Carlo experiment with focus on the decomposition when attrition is present. We find that decomposition in the presence of attrition on observables can be very misleading but that it can be successfully corrected with IPW.

We are grateful to Daniel le Maire for valuable supervision and to Blaise Melly for helpful comments.

Contents

1. Introduction

Imagine this: We are in South Africa in the year of 1993, apartheid has not yet been abolished. The world bank and co. gathers information through household surveys on living standards in developing countries, including South Africa. 5 years later, with apartheid gone and Nelson Mandela president, colored households in the Kwazulu-Natal province of South Africa are re-interviewed. Did the majority of colored South African households experience an increase in living standards from 1993 to 1998, and why?

In this paper we investigate the 1993-1998 changes in the distribution of expenditures for colored South African households using quantile regression. Quantile regression allows us to examine different parts of the distribution (as opposed to only the mean as we know from e.g. OLS). We then turn to decomposition of the differences in the distribution of expenditures into three factors: coefficients, characteristics and residuals. Throughout the analysis we take attrition into account by IPW. This has not, to our knowledge, previously been applied in a quantile decomposition framework.

Following the empirical analysis, we use a simple Monte Carlo exercise to examine IPW in the quantile regression decomposition framework. Since IPW decomposition has not been applied before, our main question is whether it actually works (in a perfect world it would be unbiased and consistent).

Our empirical results shows that the household expenditures has decreased for the majority of households between 1993 and 1998. Only the households in the very top of the distribution has experienced an increase in living standards following the abolishment of apartheid. Our decomposition shows that the drop in expenditures in the lower end of the expenditure distribution can be attributed mainly to changes in characteristics. The increase in living standards for the households in the top of the expenditure distribution, on the other hand, can be attributed mainly to increases in coefficients. Our Monte Carlo study shows that unweighted decomposition in the presence of attrition yields severely biased estimates. Luckily we also obtain some evidence that IPW decomposition may be unbiased and consistent (under some assumptions). The estimates from the IPW decomposition are, however, quite noisy. This limits the usefulness of IPW decomposition in small samples.

2. Methodology

In applied work, economists are often interested in averages (e.g. investigating the average effect of a labor-market reform on hours worked) but increasingly focus has been broadened to examine several parts of the distribution. This subsection will give a brief introduction to parametric (linear) Quantile Regression - a method that goes beyond averages and examines distributions.

Quantiles. We start with a short recap on quantiles. We are interested in the distribution of a well behaved[1] random variable y_i (think of it as standard normal) with distribution

[1]Continuously distributed and everywhere differentiable

function $F(\cdot)$. We can calculate the value of y (call it $Q_\tau(y)$) for which the probability mass below is τ. In the standard normal case we could, for instance, be interested in the value of y for which 97.5% of the probability mass is below ($\tau = 0.975$) and calculate $Q_{0.975}(y)$ to 1.96[2]. We call $Q_{0.975}(y) = 1.96$ the 0.975'th quantile of y and generally $Q_\tau(\cdot)$ the quantile function. We define the quantile function as the inverse distribution function:

$$\begin{aligned} Q_\tau(y_i) &= F^{-1}(\tau) \\ F\left(Q_\tau(y_i)\right) &= \tau \quad \text{for } \tau \in [0,1] \end{aligned} \tag{2.1}$$

The 0.50'th quantile ($Q_{0.50}(y_i)$) of y is the median, the 0.10'th quantile is the 10th percentile and so on. We simply use the quantile function, $Q_\tau(\cdot)$, to 'pick out' values that will leave τ probability mass below that value. Analogously to 2.1 we can define the conditional quantile function of y_i, conditioning on some other random variable x_i. We define the conditional quantile function as

$$\begin{aligned} Q_\tau(y_i|x_i) &= F^{-1}(\tau|x_i) \\ F\left(Q_\tau(y_i|x_i)\right) &= \tau \quad \text{for } \tau \in [0,1] \end{aligned} \tag{2.2}$$

Why is the conditional quantile of $y_i|x_i$ interesting? Think of y_i as income and x_i as the gender. We could try to estimate $E(y_i|\mathbf{x}_i)$ with OLS and look at the average returns to being male. If we believe that the returns are homogenous across the distribution[3] (high and low income earners have same return to gender) estimating $E(y_i|\mathbf{x}_i)$ may be sufficient. If, however, the returns are heterogeneous across the distribution, estimating $E(y_i|\mathbf{x}_i)$ may not be very relevant. To see this, imagine a situation where the 10% quantile of men (the man with only 10% of other men being poorer) is a lot richer than the corresponding woman, i.e. $Q_{0.10}(y_i|male) - Q_{0.10}(y_i|female) =$ "Large" and there is no difference for the 90% quantile, i.e. $Q_{0.90}(y_i|male) - Q_{0.90}(y_i|female) = 0$. With OLS we may end up estimating $E(y_i|male, x_i) - E(y_i|female, x_i) =$ "Small" and concluding that the income difference between men and women is positive but small on average. Thereby we would be neglecting the large negative consequence of being female in the bottom of the distribution.

Quantile regression. We now turn to the problem of estimating the conditional quantiles by quantile regression. We start by assuming that all conditional quantiles of y are linear in x, i.e.

$$Q_\tau(y_i|x_i) = x_i\beta_\tau \tag{2.3}$$

Notice that β_τ is allowed to vary over the quantiles. We gave an example of varying β_τ's in the gender/income story in the previous subsection. The linearity assumption is not necessary for quantile regression to work[4], but is commonly assumed and will be assumed throughout this paper. Naturally, we are interested in how we can estimate the conditional

[2]This doesn't appear out of nowhere, we calculate it as the inverse of the distribution function, $\Phi^{-1}(0.975) = 1.96$

[3]I.e. a pure location effect

[4]Intuition can be found inAngrist & Pischke (2008)

quantile function (eq. 2.3). It turns out that the β_τ of the population conditional quantile function minimizes

$$\beta_\tau = \arg\min_{\beta} E\left[\rho_\tau\left(y_i - x_i\beta\right)\right] \tag{2.4}$$

where $\rho_\tau = \tau \cdot \mathbf{1}\left(u_i > 0\right) - (1-\tau) \cdot 1\left(u_i < 0\right)$ is the asymmetric absolute loss function and $(u_i = y_i - x_i\beta)$. The loss function varies over τ, and $1 > \tau > 0$. The analogous sample problem is

$$\widehat{\beta_\tau} = \arg\min_{\beta} \frac{1}{N}\sum_{i=1}^{N}\left[\rho_\tau\left(y_i - x_i\beta\right)\right] \tag{2.5}$$

With ρ_τ defined as before. In the appendix we show more rigorously why β_τ is the solution to 2.4, here we present some intuition. Consider the case where we want to estimate the 0.25th conditional quantile, $\widehat{\beta_{0.25}}$. The weighting of each u_i in the sample minimization problem is then $\rho_{0.25} = 0.25$ for $u_i > 0$ and $\rho_{0.25} = -0.75$ for $u_i < 0$. We want to pick a $\beta_{0.25}$ that minimizes this asymmetric loss function. From the expression for $\rho_{0.25}$ we see that we always want to avoid $|u_i| \neq 0$ but situations where $y_i < x_i\beta$ are more important than $y_i > x_i\beta$ (to be exact it is 0.75/0.25=3 times more important). This means that we intuitively want to pick a $\beta_{0.25}$ that leaves $y_i > x_i\beta$, 3 times as often as $y_i < x_i\beta$. This is exactly how we interpret the 25th percentile![5]

If we assume identification, i.e. assuming β_τ to be the unique minimizer of equation 2.4 the estimator, 2.5, is consistent and $\sqrt{N}$-asymptotically normal under very weak conditions[6]. The minimization problem 2.5 has no known closed form solution but can be set up as a linear programming problem and solved numerically[7]

Adjusting for attrition. Later in this paper we will study the differences between quantiles in two periods, with attrition. This subsection introduces the problem with attrition and how it can be solved by Inverse Probability Weighting. Suppose that we have two time periods and an unbalanced panel. We will assume that we have a random sample of the underlying population in the first period (without any sample selection issues) but we don't observe all individuals in period two. If we let $s_i = 1$ when an observation is available in period 2 and $s_i = 0$ when an observation is not available, we could try to estimate the second period β_τ's using only the period two observations we actually observe. Then we would solve

$$\widehat{\beta_{2,\tau}^{UW}} = \arg\min_{\beta_2} \frac{1}{N_1}\sum_{i=1}^{N_1}\left[s_{2i}\rho_\tau\left(y_{2i} - x_{2i}\beta_2\right)\right] \tag{2.6}$$

Where N_1 is the underlying period 1 sample size (we can write it like this because $s_{2i} = 0$ for the unobserved). We call (2.6) the unweighted M-estimator. By the analogy principle

[5]For a soft introduction see Koenker & Hallock (2001)

[6]For a full specification see eg. Wooldridge (2010) p. 451 and 841 or Koenker (2005)

[7]Koenker (2005) has extensive details on this. In this paper, we use a variant of Iteratively Reweighted Least-Squares to solve the problem as suggested in ibid p. 221

we can write it as:

$$\beta_{2,\tau}^{UW} = \arg\min_{\beta_2} E\left[s_{2i}\rho_\tau\left(y_{2i} - x_{2i}\beta_2\right)\right] \tag{2.7}$$

Unfortunately, the solutions to (2.7) and (2.4) (with period 2 time-subscripts) are not generally the same. Thus, we cannot just use the observed sample to consistently estimate $\beta_{2,\tau}$. To provide some intuition, consider a case where the individuals with the highest period 1 income do not want to participate in the survey in period 2. If income in the two periods are not independent it suddenly gets very difficult to estimate Then how can we proceed? We will make an assumption on the selection rule, s_{i2}, specifically we will assume *selection on observables*

$$P(s_{i2} = 1|y_{i,2}, x_{i,2}, z_{i,1}) = P(s_{i2} = 1|z_{i,1}) \tag{2.8}$$

Here, $z_{i,1}$ contains all observed characteristics of individual i in period 1 and $P(\cdot)$ denotes probability of observing $s_{i,2}$. With selection on observables we assume that we can use the period 1 observables to predict attrition so well, that given period 1 observables, $s_{i,2}$ does not depend on $y_{i,2}, x_{i,2}$. This is a strong assumption and it excludes cases such as selection on the period 2 error-term. If we let $\pi_i^{true}(z_{i,1})$ denote the true probability of observing $s_{i,2}$ (i.e. $P(s_{i2} = 1|z_{i,1}) = \pi_i^{true}(z_{i,1})$) we can write the weighted population M-estimator as:

$$\beta_{2,\tau}^{W} = \arg\min_{\beta_2} E\left[\frac{s_{2,i}}{\pi_i^{true}(z_{i,1})}\rho_\tau\left(y_{2i} - x_{2i}\beta_2\right)\right] \tag{2.9}$$

This time the solution to (2.9) is the same as the solution to (period 2) (2.4). To see this we use the law of iterated expectations to write (2.9) as

$$\begin{aligned} \beta_{2,\tau}^{W} &= \arg\min_{\beta_2} E\left(E\left[\frac{s_{2,i}}{\pi_i^{true}(z_{i,1})}\rho_\tau\left(y_{2i} - x_{2i}\beta_2\right)|z_{i,1}, y_{i,2}, x_{i,2}\right]\right) \\ &= \arg\min_{\beta_2} E\left(E\left[\frac{(s_{2,i}|z_{i,1}, y_{i,2}, x_{i,2})}{\pi_i^{true}(z_{i,1})}\rho_\tau\left(y_{2i} - x_{2i}\beta_2\right)\right]\right) \\ &= \arg\min_{\beta_2} E\left[\rho_\tau\left(y_{2,i} - x_{2,i}\beta_2\right)\right] \end{aligned} \tag{2.10}$$

Where we use the attrition on observables assumption 2.8 to get to the last line. Naturally we don't know the true value of π_i^{true} but if we replace it with a consistent estimate, $\widehat{\pi}_i$ we can use the analogy principle to arrive at the weighted M-estimator

$$\beta_{2,\tau}^{W} = \arg\min_{\beta_2} \sum_{i=1}^{N_1}\left[\frac{s_{2,i}}{\widehat{\pi}_i(z_{i,1})}\rho_\tau\left(y_{2i} - x_{2i}\beta_2\right)\right] \tag{2.11}$$

Which is a consistent and $\sqrt{N}$-asymptotically normal estimator given weak regularity assumptions and the identification assumption. We see that all we have to do is weigh each period 2 observation by the inverse of the probability that they are still in the sample.

Up until this point there is a key point we haven't discussed: How to generally deal with panel data in a quantile regression setting. We may for instance be interested in methods to 'eliminate' a household fixed effect when we estimate the conditional quantile function like we can do in the linear panel data models. There are methods to handle panel data in a quantile regression setting[8]. However, in this paper we focus on decomposing the *difference* between living standards and the *difference* in the dependence on covariates in 1993 and 1998. Because we expect the regime change to have caused $\beta_\tau^{1993} \neq \beta_\tau^{1998}$ (and we only have T=2), we can't use the 'standard' panel data quantile regression methods. The decomposition laid out in the next subsection was developed for repeated cross-sectional purposes. In the Monte Carlo exercise, we will investigate whether the decomposition still works with serially correlated error terms.

BM-Decomposition. This subsection will introduce a method to decompose differences in unconditional distributions into differences in coefficients, characteristics and residuals. We start with an example. Think of our example with income and gender from before and expand the story with education. Suppose we observe differences in the unconditional distribution of income between men and women, perhaps $Q_{low}^{men}(Y_i) > Q_{low}^{women}(Y_i)$ such that the lowest decile of men have substantially higher income than the lowest decile of women. We want to know what is behind this difference; is the difference between income for (low decile) men and women due to different levels of education (characteristics), returns to education (coefficients) or perhaps innate unobservable ability (residuals)? From a policy point of view there is a big difference in whether it is due to characteristics (solution: increase education levels for women) or to coefficients (solution: decrease job-market gender discrimination).

With quantile regression we have already seen how to estimate a conditional quantile such as $\hat{Q}_{low}(Y_i|male, x_i)$. Unfortunately we cant simply calculate the unconditional quantiles for men by multiplying the average man ($\bar{x}_{male}$) with the coefficient estimate β_τ^{male} (at a particular τ) like we can with normal OLS for the unconditional means:

$$\begin{aligned}\hat{Q}_\tau(Y_i|x_i = male) &\neq \bar{x}'_{male}\hat{\beta}_\tau^{male}\\ \overline{Y_{male}} &= \bar{x}'_{male}\hat{\beta}^{male}\end{aligned}$$

It is a obviously difficult to decompose changes in the unconditional distribution using the conditional distribution if we can't get from the latter to the former. To work around this problem, Melly (2005) proposes a semiparametric estimator in which the conditional distribution is estimated and then integrated to find an estimate of the unconditional distribution. Thus enabling us to decompose the differences in distributions into the factors. To estimate the unconditional distribution we use the definition of the quantile function as the inverse distribution function, Eq. (2.1). Then we can write τ as

$$\tau = Prob(y_i \leq Q_\tau(y_i)) = \int 1(y_i \leq Q_\tau(y_i))\, dF(y_i)$$

We can rewrite this by first conditioning the distribution of y on x, then integrating over x (think along the lines of the Law of Iterated Expectations mechanics). Let $G(x_i)$ be

[8]Such as the correlated random effects quantile regression laid out in Abrevaya & Dahl (2008)

the distribution function of x_i.

$$\begin{aligned}\tau &= \int \left(\int \mathbf{1} \left(y_i \leq Q_\tau(y_i) \right) f\left(y|x \right) dy \right) dG\left(x \right) \\ &= \int \left(\int_0^1 \mathbf{1} \left(F^{-1}\left(\tau|x \right) \leq Q_\tau(y_i) \right) f_\tau(\tau) d\tau \right) dG\left(x \right) \\ &= \int \left(\int_0^1 \mathbf{1} \left(F^{-1}\left(\tau|x \right) \leq Q_\tau(y_i) \right) d\tau \right) dG\left(x \right)\end{aligned}$$

From the first to the second line we change the (inner) variable of integration to τ and use 2.2 . From the second to the third line we use the fact that $f_\tau(\tau) = 1$ for all τ (remember, τ is between 0 and 1). Finally, If we replace the $F^{-1}(\tau|x)$ by its consistent estimate, $x_i\hat{\beta}(\tau)$ from a quantile regression we can write the sample analog of τ as:

$$\widehat{Q}_\tau \left(\hat{\beta}_\tau, x \right) = inf \left\{ Q_\tau\left(y_i \right) : \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \left(\tau_j - \tau_{j-1} \right) \mathbf{1} \left(x_i \hat{\beta}_\tau \leq q \right) \geq \theta \right\} \tag{2.12}$$

Melly (2006) proves the estimator (2.12) to be consistent and asymptotically normal if the usual quantile regression works (i.e. if the (weak) quantile regression conditions are satisfied). For practical purposes the estimator (2.12) is simpler than it looks. Here we present the grid estimation procedure laid out in Melly (2005)[9].

(1) Estimate the conditional quantiles of y along a grid of τ's, thereby obtaining a grid of β_τ estimates (we estimate β_τ for 200 values of τ between 0 and 1) .
(2) Using the 200 β_τ-estimates from step 1, we calculate the 'predicted' quantiles of y by multiplying the estimated betas with the $x_i's$. We do this for each i leaving us with 200*N predicted quantiles of y. Together, these predicted quantiles constitute the predicted unconditional CDF of y. From this we just have to pick the quantile we want to look at[10]

Estimating the unconditional distribution is not of interest in itself, because we could just go directly to the sample distribution and plot a histogram. But it gives us the possibility of simulating counterfactual distributions. With this we can examine what happens to the unconditional distribution when we change the x's and keep everything else constant. For now, we will only give a simple example of the decomposition. Interested readers are referred to the appendix for further details and for a full specification readers are referred to Melly (2005).

Consider the example from earlier with income, gender and education. We want to decompose the differences in the unconditional income distributions between men and

[9]There is a continuous version as well, see Melly (2005)

[10]

(a) Notice that the length of our original dataset is N, but we construct a predicted y of length 200*N. Technically we should weigh each observation by $\frac{1}{200}$ to get the prediction and the true value to be of same length. This, however, is not necessary when all the weights are the same because we end up computing quantiles women into differences in characteristics. To examine the difference caused by characteristics, we calculate a counterfactual distribution of y. We calculate this by using the $\beta_\tau' s$ estimated for women but instead of using the $x_i' s$ for women we use the $x_i' s$ for men. This amounts to asking: "What would the unconditional distribution for women be, if returns to education where the female-returns but women now have same education levels as men?". The difference between this counterfactual distribution and the predicted distribution for women must then be due to differences in characteristics.

Decomposition with attrition. We will now discuss decomposition when attrition is present - and how we may handle it. The decomposition suggested in Melly (2005) is designed for repeated cross-sectional data. The key requirement for the decomposition to behave, is the consistency of the estimator of the unconditional distribution (using estimates of the conditional quantiles). Earlier we showed that M-estimators are not generally consistent when attrition is present. Fortunately, we also showed how IPW could correct this, if we were willing to assume attrition on observables. We therefore propose to use IPW in combination with the Melly (2005) decomposition. This approach is novel. At this stage we are not able to prove that decomposition will be consistent under IPW, so we will investigate this in the Monte Carlo exercise. Now, we will shortly describe how we apply the decomposition procedure with IPW:

(1) Under the attrition on observables assumption we can consistently estimate β_τ when we use IPW. Thus we can get consistent estimates of 200 β_τ's as in 1. So we just apply IPW and estimate step 1 as in the standard decomposition
(2) We proceed as described in step 2 from the normal decomposition. The only difference is that when we calculate the 'predicted' quantiles of y, we multiply the estimated betas and the $x_i' s$ (the $x_i' s$ unweighted) and make sure to weigh each predicted y with the IPW for the specific household. In this way we take the varying weight of each household into account when we re-construct the unconditional quantiles of y.

Later we will show that this IPW quantile decomposition actually works quite well to decompose when attrition is present.

3. Replication of Vahid & Maitra (2006)

The purpose of this section is to examine the distribution of the expenditures of colored South African households before and after the abolishment of apartheid. First we will present the data and then replicate the results in Vahid & Maitra (2006). This will set the scene for the decomposition of the changes we will present in the next section.

Data background. We utilize the dataset from Vahid & Maitra (2006) who assemble data from expenditure surveys of colored (Black and Indian) South African households in 1993 and 1998. The 1993 survey was conducted by the World Bank as part of the Living Standard Measurement Survey and the 1998 survey was part of the Kwazulu-Natal Income Dynamics Study (KIDS). Households surveyed in 1993 were re-interviewed in the 1998 (KIDS) study with an attrition rate of about 16%. The unbalanced panel covers 1354 households in 1993 and 1132 households in 1998. The variables include per capita expenditures, household characteristics and a few measures of survey quality (see appendix table ?? for a full description). We follow Vahid & Maitra (2006) and use the (log) per capita household expenditure as a measure of standard of living[11]. Before we start the analysis, we must stress that because the survey only covers colored South African households, we cannot generalize our conclusion to South African households in general.

FIGURE 3.1.

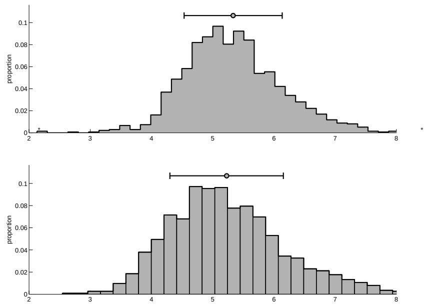

TABLE 1. Descriptive statistics, expenditure per capita

	1993	1998 (*Unweigthed*)	1998 (*Weigthed*)
Mean	295.24*** (9.81)	305.0	314.1178*** (12.98)
10th Quantile	81.71*** (2.52)	63.62	63.91*** (2.09)
25th Quantile	121.20*** (2.96)	99.14	99.93*** (3.15)
50th Quantile	198.04*** (5.26)	167.07	169.71*** (5.72)
75th Quantile	328.79*** (12.45)	321.43	330.33*** (13.51)
90th Quantile	593.12*** (24.66)	667.74	699.37*** (51.89)

Exploring the data. Figure 3.1 shows a histogram of the (log) expenditure of the households in our sample in 1993 and 1998 and table 1 shows the some descriptive figures for the expenditure per capita. The figure and table reveal something interesting: The bottom half of the 1998 households (measured by the expenditure level) had lower levels of expenditure than the bottom half in 1993; the opposite was the case for the top decile. In other words, we see increasing inequality and decreasing living standards for the poorest part of the households. We must be careful in any direct interpretation of such figures based on an unbalanced panel. If, for instance, a large part of the households with high levels of education and high living standards choose not to participate in the 1998-survey (maybe they don't have time, maybe they moved to another country etc.) we may see a difference between the 1993 and 1998 distributions simply due to attrition of that specific

[11]We refer to Vahid & Maitra (2006) for a discussion of the arguments behind this

type of household. Put differently: In 1998 we may not observe a representative sample of colored households. Vahid & Maitra (2006) discuss whether we should be worried about the attrition or not and compare the households that drop out from 1993 to 1998 (attritor households) with the households that are successfully re-interviewed (non-attritor households). They conclude that attritor households are quite different compared to non-attritors. Attritor households are smaller, less frequently headed by females, have higher levels of education and have higher per capita expenditure across the expenditure distribution. This suggests that attrition could be reason for (parts of) the difference we see in figure 3.1 and table 1 . To overcome this issue, we take account of the attrition by IPW.

Estimation. We are interested in whether the conditional distribution of household expenditures have changed from 1993 to 1998 (following the abolishment of apartheid). We may, for instance, think that returns to education have increased for the HHs in the top of the income distribution because colored South Africans had easier access to well paid jobs in 1998 compared to 1993 (like being able to seek employment as a lawyer or GP). Therefore we would like to test whether the β_τ's have changed significantly from 1993 to 1998. The simplest way to do this is with a pooled quantile regression with dummy interaction terms for year=1998. As mentioned previously, we are worried about attrition and assume attrition on observables (eq.2.8) such that we can weigh each observation in period two by the inverse of the probability of being in the sample. This assumption is unfortunately untestable and the validity of our results will depend on this assumption. We model the probability of attrition as standard logit[12]

$$P\left((1-s_i)=1\right)=\frac{exp(x\beta)}{1+exp(x\beta)}$$

Which we estimate by ML. We use the same logit specification as Vahid & Maitra (2006). The explanatory variables include (log) HH expenditure in 1993, some household characteristics and characteristics that we don't include in our quantile regression (A full specification can be found in the appendix). Assuming that our logit model is consistent for the true $P\left((1-s_i)=1\right)$, we solve the following optimization problem to estimate β_τ consistently (assuming identification and the other (weak) quantile regression assumptions):

$$\beta_\tau=\arg\min_\beta\sum_{i=1}^{N}\sum_{t=1}^{2}\left[\frac{s_{t,i}}{\widehat{\pi_{ti}}(z_{i,1})}\rho_\tau\left(y_{it}-x_{it}\beta_t\right)\right]$$

Where $s_{1,i}=1$ and $\widehat{\pi_{1,i}}(z_{i,1})=1$ for all households in period 1, and $s_{2,i}=1$ only for the observed households in period 2 and $\widehat{\pi_{2,i}}(z_{i,1})$ is the predicted probability of observing household i in period 2. Notice that Vahid & Maitra (2006) use the slightly different

[12]Thus we implicitly assume that the error term of the underlying latent variable model, $y*=x\beta+\epsilon$ with $y=\mathbf{1}(y*>0)$ has a logistic distribution. The literature suggest using either a logit or a probit model for IPW weighting (see e.g. Wooldridge (2002)). We choose logit because the logistic distribution has fatter tails than the probit and for reasons of comparability with Vahid & Maitra (2006).

weighting scheme suggested in Lipsitz et al. (1997) in their pooled quantile regression (essentially they weigh observations in both periods). We do not use results from this weighting scheme as it does not fit well into the decomposition framework we will use later. As a robustness check we have estimated the pooled quantile regression with both weighting schemes (results are shown in appendix) with very similar results.

We compute standard errors with the double-bootstrap procedure suggested in Vahid & Maitra (2006). This procedure takes the estimation uncertainty of the first-step logit regression into account. The procedure works as follows:

(1) Draw a bootstrap sample of households in period 1 to estimate the logit model.
(2) Using the estimates from the logit model, we compute the weights attributed to each household in period 2. Then we draw bootstrap samples 100 times from the households across periods and estimate the quantile regression model.
(3) We repeat step 1 and 2, 100 times, for a total of 100^2 quantile regressions for each of the 5 quantiles we investigate.

Notice that whenever we have panel data, we usually draw a panel block-bootstrap where we resample on the household level and include the drawn households in both periods. We don't do this here because we essentially ignore the panel structure of the data (we run a pooled quantile regression but don't take the possibility of serial correlation in the error terms into account). Later on, we also want to avoid implicitly changing the attrition rate as we would with the panel bootstrap. Therefore we use the two-step procedure suggested in Vahid & Maitra (2006).

Results. Table 2 shows the main results of our weighted pooled quantile regression. The top half of the table shows the estimated coefficients of the explanatory variables without any time interactions and the bottom shows the coefficients on the interacted terms. We briefly discuss some of the key points. We see from the significant coefficients on the FHH variable, that having a female household head lowers expenditure across all quantiles in 1993. From the tFHH we see that this coefficient does not change significantly from 1993 to 1998. We see from the coefficient on 1993 education that education increases expenditure across quantiles. We also see that education seems to increase expenditures more for the low quantiles than for the high quantiles. Thus education has an inequality decreasing effect. We interpret it as a sign of a lack of opportunities to take up certain types of well-paid jobs for black South Africans. If we go to the 1998 interactions, we see that the coefficients to education has increased significantly for the top quantiles, but not for the lower quantiles. This suggests that some jobs that require high education are suddenly (more) open for black South Africans. In the female household head coefficients we don't see any significant change from 1993 to 1998. When we look at the 1998 coefficients in the lower end of the distribution in table 2 and compare it with the lower end of the unconditional distribution (see table 1) we see that even though the expenditure for the households in the lowest decile has decreased unconditionally we don't see a general decrease in the coefficients in the lowest decile. So if it isn't the returns to characteristics that has lowered the expenditure for the lowest decile, what has? Vahid & Maitra (2006) suggest that "[...] some of the aspects of the distribution of household characteristics must have changed[...]" but they aren't able to examine it further. Fortunately, we are able to decompose the change in the unconditional distribution into changes in characteristics, coefficients and residuals. So we will pick up where they left.

TABLE 2. Weighted pooled quantile regression

	Quantile				
	0.10	0.25	0.50	0.75	0.90
CONSTANT	6.171*	6.1232*	6.8813*	7.6122*	8.0309*
	(0.1963)	(0.1429)	(0.156)	(0.175)	(0.2048)
AGEHD	0.0142*	0.0173*	−0.0017	−0.0199*	−0.0211*
	(0.0059)	(0.005)	(0.0059)	(0.0063)	(0.0082)
AGEHD2	−0.0001*	−0.0002*	0.0000	0.0002*	0.0002*
	(0)	(0)	(0.0001)	(0.0001)	(0.0001)
FHH	−0.1734*	−0.135*	−0.1497*	−0.1316*	−0.1645*
	(0.0328)	(0.0275)	(0.0244)	(0.0291)	(0.0367)
HDEDUC1	0.1493*	0.1294*	0.1386*	0.0983*	0.0931*
	(0.037)	(0.027)	(0.0313)	(0.0332)	(0.0447)
HDEDUC2	0.3351*	0.3545*	0.2727*	0.1857*	0.2328*
	(0.0544)	(0.0409)	(0.0351)	(0.0405)	(0.0633)
HDEDUC3	1.0275*	0.9657*	0.9471*	0.78*	0.6391*
	(0.1059)	(0.0722)	(0.0644)	(0.0761)	(0.0983)
TOTCHILD	−0.0866*	−0.0831*	−0.091*	−0.093*	−0.1073*
	(0.0092)	(0.0063)	(0.0058)	(0.0066)	(0.0093)
TOTADULT	−0.0684*	−0.0755*	−0.052*	−0.0598*	−0.0588*
	(0.0108)	(0.0087)	(0.0074)	(0.0078)	(0.0106)
TOTELDER	−0.0057	0.009	−0.0029	−0.0893*	−0.0353
	(0.0356)	(0.0303)	(0.0247)	(0.0304)	(0.0565)
BLACK	−1.3155*	−0.9953*	−0.931*	−0.8434*	−0.8511*
	(0.0855)	(0.0528)	(0.0434)	(0.0712)	(0.093)
NATAL	−0.8103*	−0.5412*	−0.5148*	−0.4093*	−0.39*
	(0.0735)	(0.0463)	(0.0439)	(0.0561)	(0.048)
RURAL	−0.3264*	−0.3126*	−0.3525*	−0.3278*	−0.4098*
	(0.0367)	(0.0325)	(0.032)	(0.0382)	(0.0452)
Year=98	−0.7519*	−0.5522*	−1.3597*	−1.2881*	−1.4922*
	(0.3043)	(0.2291)	(0.1993)	(0.2404)	(0.2761)
tAGEHD	−0.0093	−0.0031	0.0317*	0.0311*	0.0352*
	(0.0088)	(0.007)	(0.007)	(0.0076)	(0.0101)
tAGEHD2	0.0002*	0.0002*	−0.0002*	−0.0002*	−0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
tFHH	0.0512	−0.034	−0.0071	0.0146	0.0872
	(0.0544)	(0.0484)	(0.0419)	(0.0444)	(0.0608)
tHDEDUC1	0.0168	0.056	0.0812	0.1549*	0.178*
	(0.0682)	(0.0523)	(0.0551)	(0.0547)	(0.0763)
tHDEDUC2	0.1071	0.258*	0.3609*	0.428*	0.3518*
	(0.0907)	(0.0706)	(0.0595)	(0.0616)	(0.0955)
tHDEDUC3	0.1196	0.3472*	0.2731*	0.3756*	0.4123*
	(0.1951)	(0.115)	(0.0952)	(0.1084)	(0.1447)
tTOTCHIL	0.0399*	0.0351*	0.0343*	0.0328*	0.0591*
	(0.0137)	(0.0099)	(0.0106)	(0.0097)	(0.0133)
tTOTADUL	0.0138	0.0083	−0.0143	0.012	0.0045
	(0.0165)	(0.0125)	(0.0109)	(0.011)	(0.0146)
tTOTELDE	−0.1526*	−0.1432*	−0.1475*	−0.0273	−0.0716
	(0.0537)	(0.0448)	(0.0367)	(0.04)	(0.0715)
tBLACK	0.3589*	0.0196	−0.0256	−0.1276	−0.3107*
	(0.1168)	(0.0859)	(0.0867)	(0.0981)	(0.1258)
tNATAL	0.7315*	0.3537*	0.3612*	0.2265*	0.2831*
	(0.0902)	(0.0617)	(0.0749)	(0.0708)	(0.0944)
tRURAL	−0.0659	−0.0858	−0.0119	−0.0538	0.1243
	(0.0711)	(0.0582)	(0.0512)	(0.0554)	(0.0818)

4. Applying the decomposition

In this section we decompose the change in the unconditional distribution of household expenditures from 1993 to 1998 into changes in coefficients, characteristics and residuals.

Figure 4.1. Decomposition of changes

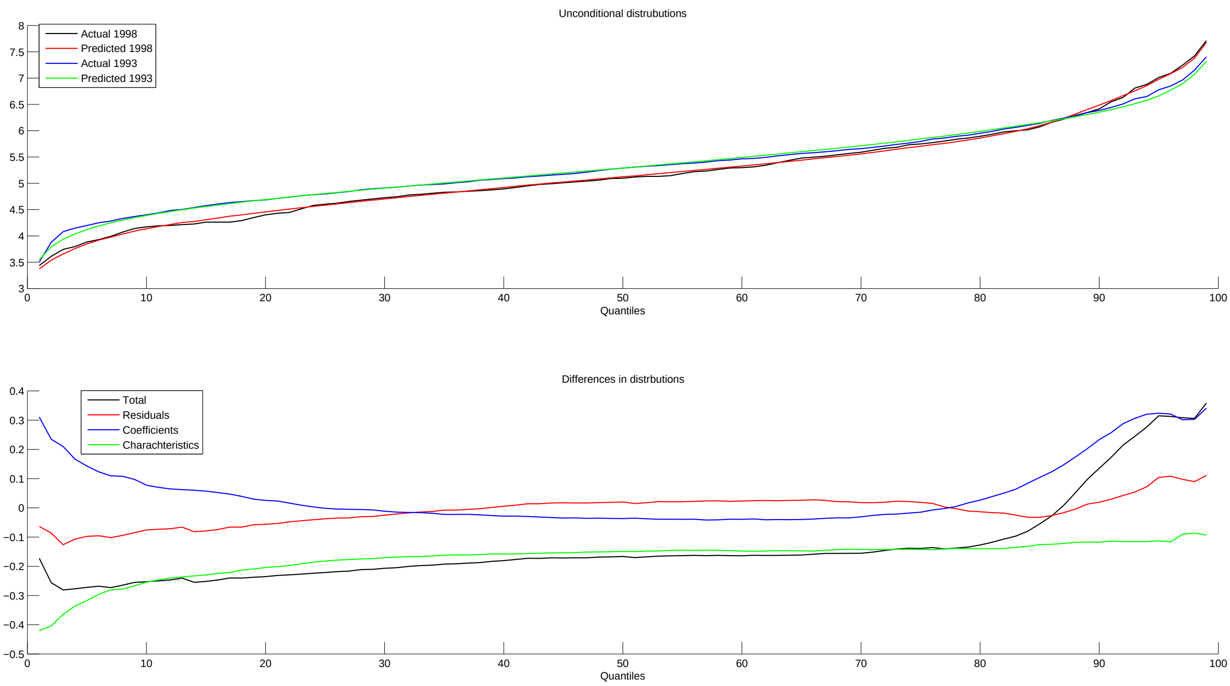

Figure 4.1 and 4.2show the main results from our decomposition. The top panel in figure 4.1 shows the actual and predicted distributions of households expenditures. We show them to show how close the predicted distribution of y is to the true sample distribution of y. The lower panel in figure 4.1 shows the decomposition. Figure 4.2 shows the decomposition with 95% confidence bands. These were computed with the bootstrap-in-bootstrap procedure described earlier, but only for 20^2 replications (This procedure is computationally very intensive).

The black/grey figure shows the total change in the unconditional distribution. We notice that it is below zero for the majority of the distribution. To examine the reasons behind this change, we look at the tree factors coefficients, characteristics and residuals. The green figure shows the change in the unconditional distribution that can be attributed to changes in characteristics. We see that characteristics have a significant effect almost everywhere on the distribution, but the effect is strongest at the lower deciles. We believe this to be driven by an increase in the number of households with a female head.

The blue figure shows the change attributed to changes in coefficients. We see that the difference due to coefficients is never negative, but significantly positive in the very top of the income distribution. This suggests that returns to education have increased for the richest part of the black South Africans. We believe this to be driven by increases in well-paid job opportunities for colored South Africans.

Vahid & Maitra (2006) suggest that the change in characteristics may be the driving force behind the decrease in the living standards for the majority of the households. Our decomposition support that hypotheses. We also see that coefficients have had a positive effect in the highest decile of South African households.

FIGURE 4.2. Decomposition of changes with confidence bounds

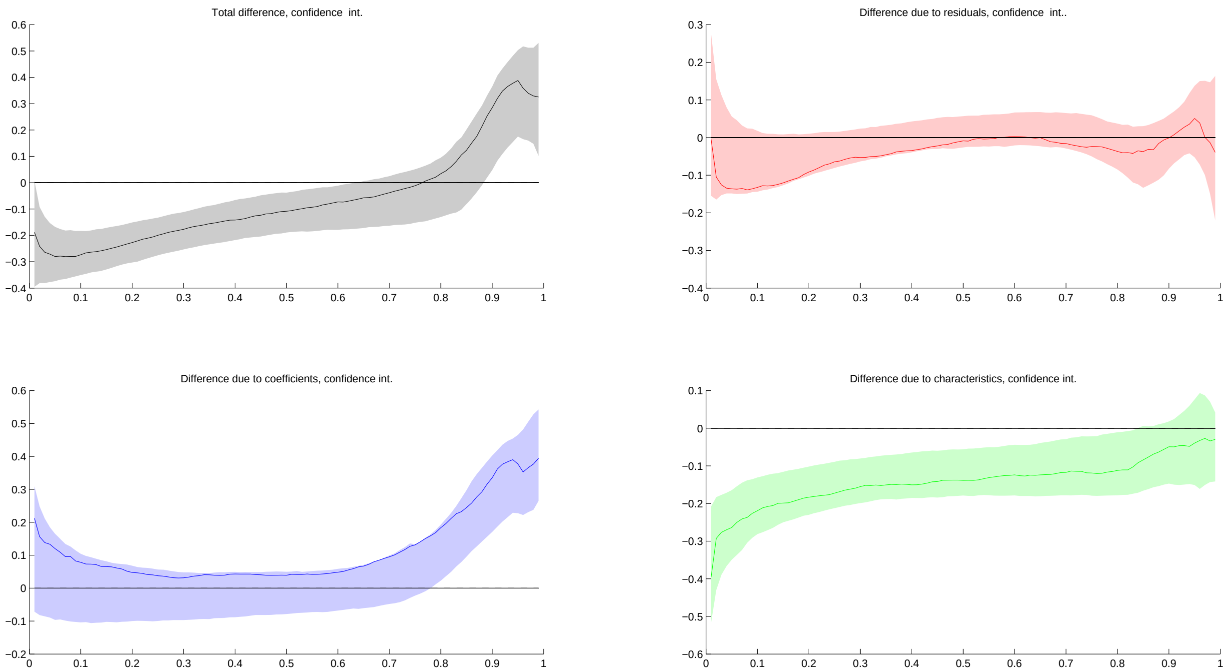

TABLE 3. Decomposition of changes in quantiles

	Total	Residual	Coefficient	Characteristics
10%	−0.2735***	−0.1326**	0.0783	−0.2192***
	(0.0425)	(0.0413)	(0.0539)	(0.0374)
25%	−0.1995***	−0.0645**	0.0376	−0.1726***
	(0.0373)	(0.0245)	(0.0405)	(0.0308)
50%	−0.1085**	−0.0089	0.0391	−0.1387***
	(0.0382)	(0.0212)	(0.0328)	(0.0308)
75%	−0.0114	−0.0239	0.1314***	−0.1190**
	(0.0489)	(0.0287)	(0.0366)	(0.0377)
90%	0.2848***	−0.0012	0.3353***	−0.0493
	(0.0752)	(0.0357)	(0.0577)	(0.0404)

Bootstrap-in-bootstrap (20 repx20 rep) std. err. in parentheses

5. Monte-Carlo Simulations

In the previous sections we have used IPW quantile decomposition, a method that we haven't proved to be consistent (or even unbiased). In this MC exercise we examine the properties of the quantile decomposition, with special emphasis on the IPW quantile decomposition. We will consider quantile decomposition in each of the following four scenarios relating to the data available for the researcher:

(1) Two-period repeated cross-sections in two periods
(2) Two-period panel data without attrition

(3) Two-period panel data with attrition on observables
(4) Two-period panel data with attrition on observables which is taken into account by IPW weighting

DGP: The linear location-scale model. For our Monte Carlo exercise we construct a linear location-scale model. The model we construct is

$$y_{i,ref} = 1 + x_{i,ref} + (0.3 + x_{i,ref})(u_{i,ref} + a_{i,})$$

Where $x_{ref} \sim N(4,1)$, $u_{ref} \sim N(0,1)$ and $a \sim N(0,10)$. An important aspect of this model is that it has heteroschedastic errors [13]. This has the effect that the β_τ's will vary over the quantiles. For example the slope coefficient (β_1) on the median will be $\beta_{,median} = 1$ while the coefficient for the first decile will be $\beta_{0.10} = -11.88$[14]. Our main focus is whether we can consistently and unbiasedly decompose the changes in the unconditional distribution of y, when attrition is present using an IPW weighted version of the Melly (2005) decomposition. We therefore need a two-period model. The second period is modeled as model:

$$y_{i,new} = 1 + 2x_{i,new} + (0.3 + x_{i,new})(u_{i,new} + a_{i,})$$

Where $x_{new} \sim N(5,1)$, $u_{new} \sim N(0,1)$. The second period is very similar to the first period, except for a change in the slope coefficient (from 1 to 2) and an increase in the mean of the x (from 4 to 5). The $a_{i,}$ is what makes this a panel data model because it is the same in the two periods (we draw new $a_i^{\prime}s$ in period 2 for the figures for repeated cross-section). The individual specific effect is independent of the x_i's, which is analogous to the assumption we know from the linear random effects panel data model. Notice that the standard deviation in $a_{i,}$ is quite large, which induces strong serial correlation in the combined error term ($u_{i,new} + a_{i,}$). This essentially helps us build a model where our IPW weighting will work reasonably well (specifics will follow). We are willing to do this, because we want to know whether decomposition with attrition and IPW will work, **given** that IPW would work in a standard quantile regression. We are quite convinced that our decomposition will fail miserably if IPW is not appropriate in our context (as quantile regression would break down), but we don't actually test this.

Since we are interested in modeling attrition, we must have an attrition rule. We construct the following rule:

$$attrite = \mathbf{1}\left(y_{i,ref} + \nu_i - 10 > 0\right)$$

[13] For a brief introduction readers are referred to le Maire (2011). For a detailed exposition readers are referred to Koenker (2005)

[14] How do we see this? $E[y_i|\mathbf{x}_i] = 1 + x_i$ and $var(y_i|\mathbf{x}_i) = (0.3+1)^2(10^2+1^2)$. $y_i|\mathbf{x}_i$ is normally distributed because the combined error term is the sum of two independent normal random variables. Thus we can write $F(y_i|x_i) = \Phi\left(\frac{y_i - (1+x_i)}{(0.3+1)^2(10^2+1^2)}\right)$. We can then write the conditional quantile as $Q_\tau(y_i|x_i) = 1 + x_{i,ref} + (0.3 + x_{i,ref})\Phi^{-1}_{0,101}(\tau)$ where $\Phi^{-1}_{0,101}(\tau)$ is the quantile of the normal distribution with mean 0 and std. error $\sqrt{101}$. Knowing this we can calculate $\beta_{1,0.10} = 1 + \Phi^{-1}_{0,101}(0.10) = -11.88$

Where $\nu_i \sim logistic(0, 30)$. With this attrition rule we have attrition on observables, as attrition only depends on period 1 data. The logistic distribution may seem a a bit too convenient since we will estimate the attrition probability with a logit model. Again we must stress that we do not want to test whether IPW is generally appropriate. Instead we want to test whether IPW can be used in the decomposition context **given** that IPW would work in a standard quantile regression. We choose a substantial variance for the error term, this is again because we want IPW to work. Essentially the large variance of the error term ensures that we have enough households that were 'supposed to drop out' but didn't. Therefore we can give them large weight in our estimation, enabling us to recover the true underlying unconditional distribution. This is the same sort of mechanics as those we need for propensity score matching to work. We subtract a constant to avoid too many households exiting after period 1. With this specification approximately 50% of households drop out.

The last step before we can look at some graphs, is the estimation of the probability of attrition we use in IPW. As in the empirical application, we estimate parameters of a logit model with maximum likelihood. We estimate the binary variable attrition including only $y_{i,ref}$ and a constant term as explanatory variables. Having obtained this estimate we follow the decomposition procedure described earlier.

Results. Table 4 and 5 show the key points of our Monte Carlo study[15]. We have estimated the decomposition for the three different types of available data (repeated cross sections, balanced panel data and unbalanced panel data). For the unbalanced panel data we decompose with no attrition adjustment (Unbalanced/-) and with IPW adjustment (Unbalanced/IPW). The underlying DGP is the same for all models (described in the previous section) and the true values are simulated[16]. The tables show decomposition at the median and the decomposition of the change in the interdecile range. From the decomposition at the median β we see that the quantile decomposition for repeated cross section and panel data are able to decompose the changes quite well. For the interdecile range, the standard deviations are quite large except for the estimation of changes due to characteristics which they both seem to capture well. For the unbalanced panel data, we see that the unweighted estimator is severely biased for the decomposition into coefficients and residuals for the median and interdecile range respectively. We see that the IPW decomposition does not suffer from this severe bias and is actually able to capture the changes at the median quite well. For the interdecile range there is more noise in the IPW decomposition compared to the estimates from data without attrition problems.

[15]For reasons of computational time we could only do 100 Monte Carlo replications, we realize that this is not sufficient

[16]We exploit that our DGP gives us the unconditional distribution directly and draw the largest sample size possible from our DGP. Then we calculate the true unconditional distribution (from which we can calculate true counterfactuals). We do this 100 times and take the mean.

Table 4. Monte Carlo Simulation with 4 different DGP's

$\beta_{0.5}$	Residuals		Coefficients		Characteristics	
Data	Mean	std	Mean	std	Mean	std
Repeated	0.02	0.57	5.12	3.32	1.03	1.89
Panel	-0.07	0.55	4.89	1.82	1.04	1.57
Unbalanced/-	0.11	0.71	-18.85	2.79	0.93	1.28
Unbalanced/IPW	-0.23	0.73	5.29	2.73	0.81	1.40
True	0,00		4,80		1,06	

N=1000,Rep.=100

Table 5. Monte Carlo Simulation with 4 different DGP's

$\beta_{0.9} - \beta_{0.1}$	Residuals		Coefficients		Characteristics	
Data	Mean	std	Mean	std	Mean	std
Repeated	0.23	6.73	0.20	0.41	25.78	2.84
Panel	0.07	3.52	0.15	0.36	25.38	3.04
Unbalanced/-	-18.88	5.52	0.23	0.67	25.28	2.54
Unbalanced/IPW	1.57	11.52	0.46	0.94	25.39	3.13
True	0,00		0,07		25,65	

N=1000,Rep.=100

Table 6 and 7 is basically the same tables as the two previous ones, but with N=10,000 instead of N=1,000. The standard errors are uniformly lower for the N=10,000 tables compared to the N=1,000 tables. This suggest consistency of the decomposition estimators. Again the unweighted decomposition is biased when attrition is present.

Table 6. Monte Carlo Simulation with 4 different DGP's

$\beta_{0.5}$	Residuals		Coefficients		Characteristics	
Data	Mean	std	Mean	std	Mean	std
Panel	0.01	0.16	4.71	0.55	1.13	0.52
Unbalanced/-	-0.01	0.23	-18.87	0.93	1.03	0.44
Unbalanced/IPW	-0.05	0.20	4.98	0.88	0.97	0.53
True	0.01		4.80		1.06	

N=10.000,Rep.=100

Table 7. Monte Carlo Simulation with 4 different DGP's

$\beta_{0.9} - \beta_{0.1}$	Residuals		Coefficients		Characteristics	
Data	Mean	std	Mean	std	Mean	std
Panel	0.09	1.26	0.07	0.07	25.57	0.86
Unbalanced/-	-18.61	1.95	0.14	0.15	25.55	0.93
Unbalanced/IPW	0.39	3.52	0.13	0.18	25.63	0.92
True	-0.01		0.07		25.65	

N=10.000,Rep.=100

The tables presented here provide some evidence that the estimates from the IPW decomposition are unbiased and consistent (under our assumptions). In the appendix we provide tables and graphs shedding additional light on this issue and we strongly encourage readers to take a look at the awesome graphs. We repeat: awesome graphs! Generally we see decreasing standard deviations when the sample size increases and we don't see signs of bias. Despite of this, we believe that there are still many untold stories on the IPW quantile decomposition which must be studied further. At the present, our Monte Carlo study provides us with confidence that IPW decomposition is appropriate under attrition on observables.

6. Conclusion

In this paper we investigate the differences in the expenditure distribution for colored South African households between 1993 and 1998. We find that the expenditures for the poorest majority of households have decreased while the expenditures have increased only for the very top of the income distribution. We decompose the differences into changes in coefficients, characteristics and residuals using a modified version of the Melly (2005) quantile regression decomposition. This IPW quantile regression decomposition adjusts for attrition bias. We find that the changes in the expenditures for the lower deciles are mainly due to changes in characteristics while the changes in the upper deciles are due to changes in coefficients.

We do a small Monte Carlo exercise to test the properties of IPW quantile regression decomposition under the attrition on observables assumption. We find that the IPW decomposition is able to successfully correct the attrition bias that otherwise have grave consequences for the properties on the regular quantile regression decomposition (it is severely biased). We don't observe bias in any particular direction in the IPW decomposition and we see decreasing standard errors when the sample size increases. This suggests consistency and unbiasedness of the IPW decomposition estimator.

The Monte Carlo study leaves us optimistic about the properties of the IPW decomposition, but a lot of work remains for the future. Future work will have to dig much further into the IPW decomposition, testing a variety of DGP's and assumptions and increasing the sheer number of Monte Carlo replications.

References

Abrevaya, J. & Dahl, C. M. (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, 26(4), 379–397.

Angrist, J. D. & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.

Koenker, R. & Hallock, K. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4), 43–56.

le Maire, D. (2011). Extra lecture: Quantile regression and bootstrapping.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., & Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: Application to cd4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(4), 463–476.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4), 577–590.

Melly, B. (2006). *Applied Quantile Regression*. PhD thesis, University of St. Gallen.

Vahid, F. & Maitra, P. (2006). The effect of household characteristics on living standards in South Africa 1993-1998: a quantile regression analysis with sample attrition. *Journal of Applied Econometrics*, 21(7), 999–1018.

Wooldridge, J. M. (2002). Inverse probability weighted m-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1(2), 117–139.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Number 0262232588 in MIT Press Books. The MIT Press, 2 edition.

Appendix A. Appendix

Quantile Regression. We will now show that the β_τ of the population conditional quantile function is:

$$\beta_\tau = \arg\min_\beta E\left[\rho_\tau\left(y_i - x_i\beta\right)\right]$$

where $\rho_\tau = \tau \cdot \mathbf{1}\left(u_i > 0\right) - \left(1-\tau\right) \cdot 1\left(u_i < 0\right)$ $1 > \tau > 0$ is the asymmetric absolute loss function and $\mathbf{1}\left(\cdot\right)$ is (binary) indicator function. We can rewrite the minimization problem using the fact that:

$$\begin{aligned} E\left[\rho_\tau\left(y_i - x_i\beta\right)\right] &= \int_{-\infty}^{\infty} \rho_\tau \underbrace{\left(y_i - x_i\beta\right)}_{u_i} dF_{x|y}\left(y\right) \\ &= \int_{-\infty}^{\infty} \left[\tau \cdot 1\left(u_i > 0\right) - \left(1-\tau\right) \cdot 1\left(u_i < 0\right)\right] u_i dF_{x|y}\left(y\right) \end{aligned}$$

By construction we have that

$$u_i \begin{cases} > 0 & y_i > x_i\beta \\ < 0 & y_i < x_i\beta \end{cases}$$

We use this and split the integrals to further reduce the equation

$$\begin{aligned} E\left[\rho_\tau\left(y_i - x_i\beta\right)\right] &= \int_{-\infty}^{x_i\beta_\tau} \left[\tau \cdot \underbrace{1\left(u_i > 0\right)}_{=0} - \left(1-\tau\right) \cdot \underbrace{1\left(u_i < 0\right)}_{=1}\right] u_i dF_{x|y}\left(y\right) \\ &\quad + \int_{x_i\beta_\tau}^{\infty} \left[\tau \cdot \underbrace{1\left(u_i > 0\right)}_{=1} - \left(1-\tau\right) \underbrace{1\left(u_i < 0\right)}_{=0}\right] u_i dF_{x|y}\left(y\right) \\ &= \int_{-\infty}^{x_i\beta_\tau} -\left(1-\tau\right) \cdot u_i dF_{x|y}\left(y\right) + \int_{x_i\beta_\tau}^{\infty} \tau \cdot u_i dF_{x|y}\left(y\right) \\ &= -\left(1-\tau\right) \cdot \int_{-\infty}^{x_i\beta_\tau} \left[y_i - x_i\beta\right] dF_{x|y}\left(y\right) + \tau \cdot \int_{x_i\beta_\tau}^{\infty} \left[y_i - x_i\beta\right] dF_{x|y}\left(y\right) \end{aligned}$$

Thereby we now obtain the following minimization problem:

$$\min_{\beta_\tau} \; -\left(1-\tau\right) \cdot \int_{-\infty}^{x_i\beta_\tau} \left[y_i - x_i\beta\right] dF_{x|y}\left(y\right) + \tau \cdot \int_{x_i\beta_\tau}^{\infty} \left[y_i - x_i\beta\right] dF_{x|y}\left(y\right)$$

Differentiating with respect to β we obtain the FOC

$$\begin{aligned} \frac{\partial}{\partial \beta_\tau} = 0 &= -\left(1-\tau\right) \cdot \int_{-\infty}^{x_i\beta_\tau} \left[-x_i\right] dF_{x|y}\left(y\right) + \tau \cdot \int_{x_i\beta_\tau}^{\infty} \left[-x_i\right] dF_{x|y}\left(y\right) \\ &= \left(1-\tau\right) \cdot \int_{-\infty}^{x_i\beta_\tau} x_i dF_{x|y}\left(y\right) - \tau \cdot \int_{x_i\beta_\tau}^{\infty} x_i dF_{x|y}\left(y\right) \\ &= \left(1-\tau\right) \cdot F\left(x_i\beta_\tau | x_i\right) - \tau \cdot \left[1 - F\left(x_i\beta_\tau | x_i\right)\right] \\ \Leftrightarrow \tau &= F\left(x_i\beta_\tau | x_i\right) \end{aligned}$$

and using Eq. 2.2 we obtain

$$\begin{aligned} F^{-1}(\tau|x_i) &= F^{-1}(F(x_i\beta_\tau|x_i)|x_i) \\ Q_\tau(y_i|x_i) &= x_i\beta_\tau \quad \blacksquare \end{aligned}$$

Decomposition. Melly (2006) argues that the central tendency of the distribution is the median for which we can consider following simple equation:

$$\begin{aligned} y_i^i &= \mathbf{x}_t^i \beta_{0.5}^t + u_i^t, \\ & \text{where } t \in \{93, 98\} \text{ and } \tau \in [0;1] \end{aligned}$$

which describes the median regression in year t. Using the 2-step method described in the main text, we could ask how the unconditional distribution in 98 would have looked, if characteristics were the only thing changing from 93 to 98? We would estimate the counter factual $\widehat{Q}_\tau\left(\hat{\beta}_\tau^{93}, x^{98}\right)$, to answer such a question. Any difference in $\widehat{Q}_\tau\left(\hat{\beta}_\tau^{93}, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}_\tau^{93}, x^{93}\right)$ must be due to changes in characteristics.

Melly (2005) further proposes a method to separate the effects from coefficients and residuals by assuming that the residuals of theτ'th quantile of the distribution can be estimated by $x\left(\hat{\beta}_\tau - \hat{\beta}_{0.5}\right)$. Again, simulating a counter factual distribution, but now keeping characteristics in year 98 constant, difference in distribution due to changes in residuals can be estimated as

$$\widehat{Q}_\tau\left(\hat{\beta}_\tau^{98} - \hat{\beta}_{0.5}^{98} - \left(\hat{\beta}_\tau^{93} - \hat{\beta}_{0.5}^{93}\right), x^{98}\right)$$

To estimate changes due to coefficients, we will turn to the idea of the median being the central tendency. By comparing the median regression estimate with a counter factual distribution, where characteristics and residuals in year 98 are constant, difference in distribution due to coefficients are estimated as

$$\widehat{Q}_\tau\left(\hat{\beta}_{0.5}^{98} - \hat{\beta}_{0.5}^{93}, x^{98}\right)$$

By summing up the 3 differences gives us the total difference in distrubutions

$$\begin{aligned} \widehat{Q}_\tau\left(\hat{\beta}_\tau^{98}, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}_\tau^{93}, x^{93}\right) &= \widehat{Q}_\tau\left(\hat{\beta}_\tau^{98} - \hat{\beta}_{0.5}^{98} - \left(\hat{\beta}_\tau^{93} - \hat{\beta}_{0.5}^{93}\right), x^{98}\right) \\ &\quad + \widehat{Q}_\tau\left(\hat{\beta}_{0.5}^{98} - \hat{\beta}_{0.5}^{93}, x^{98}\right) \\ &\quad + \widehat{Q}_\tau\left(\hat{\beta}_\tau^{93}, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}_\tau^{93}, x^{93}\right) \end{aligned} \tag{A.1}$$

Since we are able to estimate the counter factual distributions we are also able to decompose all relevant statics. Notice that Eq. A.1 is a version of Melly (2005)'s decomposition:

$$
\begin{aligned}
\widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) &= \widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{m98,r93}, x^{89}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{m98,r93}, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{98}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) \\
& \quad where\ \hat{Q}_\tau\left(\hat{\beta}^{m98,r93}, x^{89}\right) \equiv \hat{\beta}^{98}_{0.5} + \hat{\beta}^{93}_\tau - \hat{\beta}^{93}_{0.5}
\end{aligned}
$$

which is shown below

$$
\begin{aligned}
\widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) &= \widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{98}_{0.5} + \hat{\beta}^{93}_\tau - \hat{\beta}^{93}_{0.5}, x^{89}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{98}_{0.5} + \hat{\beta}^{93}_\tau - \hat{\beta}^{93}_{0.5}, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{98}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) \\
&= \widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau - \hat{\beta}^{98}_{0.5} - \hat{\beta}^{93}_\tau + \hat{\beta}^{93}_{0.5}, x^{89}, x^{98}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{98}_{0.5} + \hat{\beta}^{93}_\tau - \hat{\beta}^{93}_{0.5} - \hat{\beta}^{93}_\tau, x^{98}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) \\
\widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) &= \widehat{Q}_\tau\left(\hat{\beta}^{98}_\tau - \hat{\beta}^{98}_{0.5} - \left(\hat{\beta}^{93}_\tau - \hat{\beta}^{93}_{0.5}\right), x^{89}, x^{98}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{98}_{0.5} - \hat{\beta}^{93}_{0.5}, x^{98}\right) \\
&+ \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{98}\right) - \hat{Q}_\tau\left(\hat{\beta}^{93}_\tau, x^{93}\right) \ \blacksquare
\end{aligned}
$$

Table 8. Variable Definition, Table VII from Vahid & Maitra (2006)

Variable	Description
PCEXP	Per capita household expenditure
LPCEXP	Log per capita household expenditure
LPCINC	Log per capita household income
AGEHD	Age of household head
AGEHD2	Age of household head squared
FHH	=1 if household head is female
HDEDUC1	=1 if highest education attained by household head is primary school
HDEDUC2	=1 if highest education attained by household head is middle school
HDEDUC3	=1 if highest education attained by household head is secondary school or higher
TOTCHILD	Total number of children in the household (individuals aged less than 18)
TOTADULT	Total number of working age adults in the household (males aged 18–64,females aged 18–59)
TOTELDER	Total number of elderly in the household (males aged 65 and higher, females aged 60 and higher)
BLACK	=1if household is black
NATAL	=1if household is resident of former Natal
RURAL	=1if the household resides in a rural area
YEAR=1998	=1 if 1998
ATTRITE	=1if the household was not re-interviewed in 1998
LPCEXP93	Log per capita household expenditure in 1993
HHSIZE93	Household size in 1993
VERIFY93	=1 if questionnaire was verified by a supervisor in 1993
TARROAD93	=1 if there is a tarred road in the cluster in 1993
CLINIC93	=1 if there is a clinic in the cluster in 1993
DOCTOR93	=1 if there is a doctor in the cluster in 1993
HDEDUC1-93	=1 if highest education attained by household head in 1993 is primary school
HDEDUC2-93	=1 if highest education attained by household head in 1993 is middle school or higher
HDEDUC3-93	=1 if highest education attained by household head in 1993 is secondary school or higher
TOTCHILD93	Total number of children in the household in 1993 (individuals aged less than 18)

Replications results. Table 9 is the estimated *coefficient* for the logit model used to estimate the IPW

Table 10 is the replicated estimates for Vahid & Maitra (2006) weighted pooled quantile regression

Table 9. Binomal logit estimates of attritor households

	Coefficients
CONSTANT	2.4003** (0.7586)
LN(PCEXP93)	−0.2877* (0.1292)
HHSIZE93	−0.0940* (0.0454)
HDEDUC2	0.4310* (0.1938)
HDEDUC3	0.7628 (0.4067)
TOTCHILD	−0.1361* (0.0668)
TARROAD9	−0.7257 (0.2013)
CLINIC93	−0.5798 (0.1743)
DOCTOR93	0.3909 (0.2041)
VERIFY93	−0.7938 (0.1774)
Observed probability	0.164
Predicted probability($:\equiv e^{\bar{x}\beta}/(1+e^{\bar{x}\beta})$)	0.137

TABLE 10. Replication of Vahid & Maitra (2006)'s weighted quantile regression

	Quantile				
	0.10	0.25	0.50	0.75	0.90
CONSTANT	−0.0085 (0.1289)	−0.2281* (0.1076)	−0.1157 (0.0692)	−0.0367 (0.0735)	0.0601 (0.0955)
AGEHD	0.0108 (0.0203)	0.0197 (0.0152)	0.0136 (0.0136)	−0.0075 (0.0145)	−0.0164 (0.0191)
AGEHD2	−0.0001 (0.0002)	−0.0002* (0.0001)	−0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0002)
FHH	−0.1959* (0.0890)	−0.1743* (0.0793)	−0.1509** (0.0567)	−0.1744** (0.0652)	−0.1468 (0.0816)
HDEDUC1	0.2034* (0.0964)	0.1501 (0.0786)	0.1581 (0.0835)	0.0632 (0.0908)	0.0019 (0.1047)
HDEDUC2	0.3346* (0.1370)	0.2439* (0.1026)	0.2606** (0.0974)	0.1315 (0.1130)	0.0585 (0.1284)
HDEDUC3	1.0785*** (0.2764)	0.7623*** (0.1908)	0.9281*** (0.1642)	0.6796*** (0.1470)	0.5378** (0.2000)
TOTCHILD	−0.0623* (0.0287)	−0.0424 (0.0258)	−0.0598*** (0.0139)	−0.0846*** (0.0125)	−0.1095*** (0.0154)
TOTADULT	−0.0687* (0.0286)	−0.0649** (0.0210)	−0.0639** (0.0207)	−0.0415 (0.0247)	−0.0556* (0.0253)
TOTELDER	−0.0869 (0.0773)	−0.0829 (0.0815)	−0.0812 (0.0727)	−0.1139 (0.0680)	−0.0826 (0.0787)
BLACK	−1.2413*** (0.1885)	−0.9545*** (0.1103)	−0.8461*** (0.1035)	−0.6676*** (0.1580)	−0.7600*** (0.1287)
NATAL	−0.7333*** (0.1488)	−0.5767*** (0.0767)	−0.4871*** (0.0880)	−0.3260** (0.1248)	−0.3432*** (0.1012)
RURAL	−0.3749*** (0.1076)	−0.4271*** (0.0804)	−0.4026*** (0.0833)	−0.4593*** (0.0912)	−0.4786*** (0.0884)
Year=98	−0.6850 (0.7946)	−0.4289 (0.5454)	−0.8979 (0.4783)	−0.8781 (0.5954)	−1.4910* (0.6866)
IPW	6.1073*** (0.6727)	6.1920*** (0.4355)	6.4541*** (0.3764)	7.2569*** (0.4319)	7.9920*** (0.5028)
tAGEHD	−0.0057 (0.0240)	−0.0065 (0.0185)	0.0193 (0.0173)	0.0180 (0.0198)	0.0300 (0.0261)
tAGEHD2	0.0001 (0.0002)	0.0002 (0.0002)	−0.0001 (0.0002)	0.0000 (0.0002)	−0.0001 (0.0002)
tFHH	0.0738 (0.1089)	0.0058 (0.0968)	0.0025 (0.0734)	0.0518 (0.0806)	0.0641 (0.1051)
tHDEDUC1	−0.0366 (0.1290)	0.0415 (0.1024)	0.0580 (0.1037)	0.1871 (0.1109)	0.2761 (0.1415)
tHDEDUC2	0.1082 (0.1827)	0.3916** (0.1303)	0.3617** (0.1176)	0.4752*** (0.1364)	0.5268** (0.1718)
tHDEDUC3	0.0689 (0.3605)	0.5846* (0.2287)	0.2660 (0.1917)	0.4797** (0.1838)	0.5167* (0.2563)
tTOTCHIL	0.0157 (0.0309)	−0.0005 (0.0271)	0.0043 (0.0176)	0.0240 (0.0153)	0.0612** (0.0206)
tTOTADUL	0.0145 (0.0337)	−0.0018 (0.0240)	−0.0012 (0.0234)	−0.0056 (0.0267)	0.0001 (0.0293)
tTOTELDE	−0.0712 (0.0968)	−0.0446 (0.0945)	−0.0665 (0.0819)	−0.0024 (0.0772)	−0.0239 (0.1010)
tBLACK	0.2849 (0.2445)	−0.0298 (0.1539)	−0.1325 (0.1434)	−0.3043 (0.1922)	−0.3938* (0.1928)
tNATAL	0.6543*** (0.1941)	0.3579*** (0.0981)	0.3018* (0.1202)	0.1402 (0.1512)	0.2423 (0.1548)
tRURAL	−0.0176 (0.1353)	0.0259 (0.1050)	0.0339 (0.0973)	0.0781 (0.1069)	0.1893 (0.1269)

Monte Carlo Simulations, Decomposition. See description of what lies underneath these Monte Carlo Simulations and DGP in the Monte Carlo section.

The figures below show decomposition when we have differences in data availability. First we show all figures for simulation with N=1000 then we show the figures for N=10,000. The bands in the figures are 95% MC confidence bands. To get an overview, take a look at figure A.5. This figure was estimated for panel data (no issues with attrition) for N=10,000 and is very close to the true decomposition values. Notice how far the unweighted/regular decomposition is from the most ideal figure, figure A.5. Especially for coefficients.

Figure A.1. MC Simulation of decomposition, Repeated Cross-Sectional Data, $N = 1000$, $MCRep = 100$

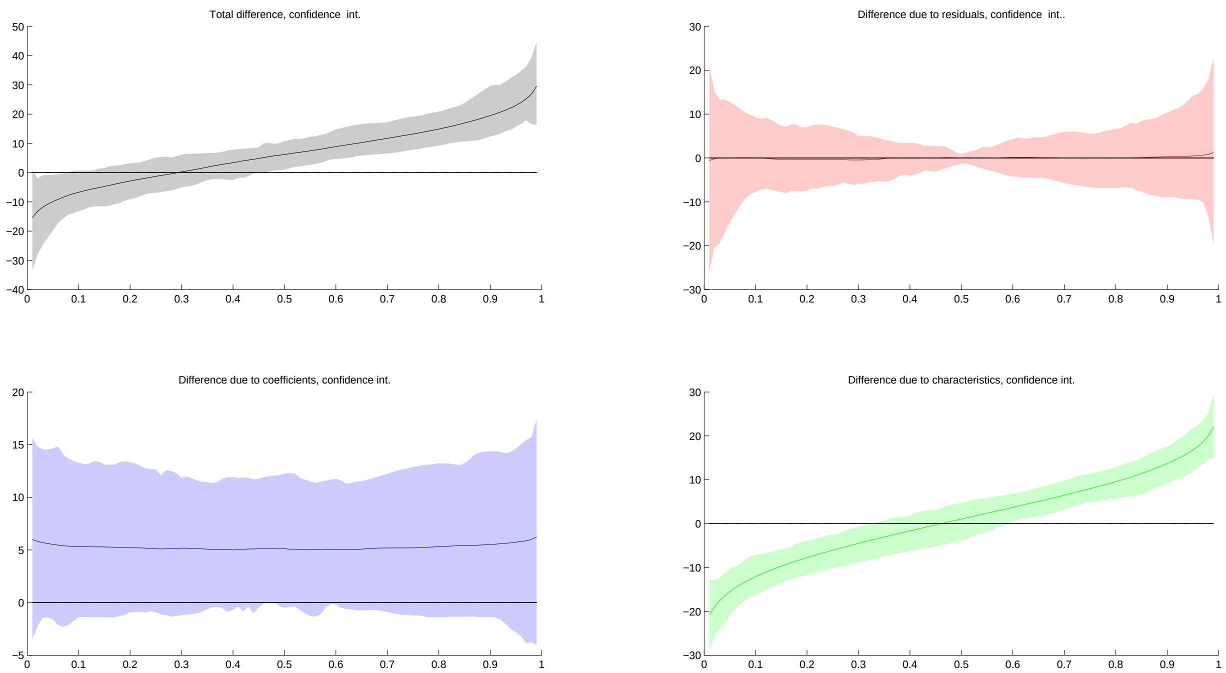

Figure A.2. MC Simulation of decomposition, Panel Data $N = 1000$, $MCRep = 100$

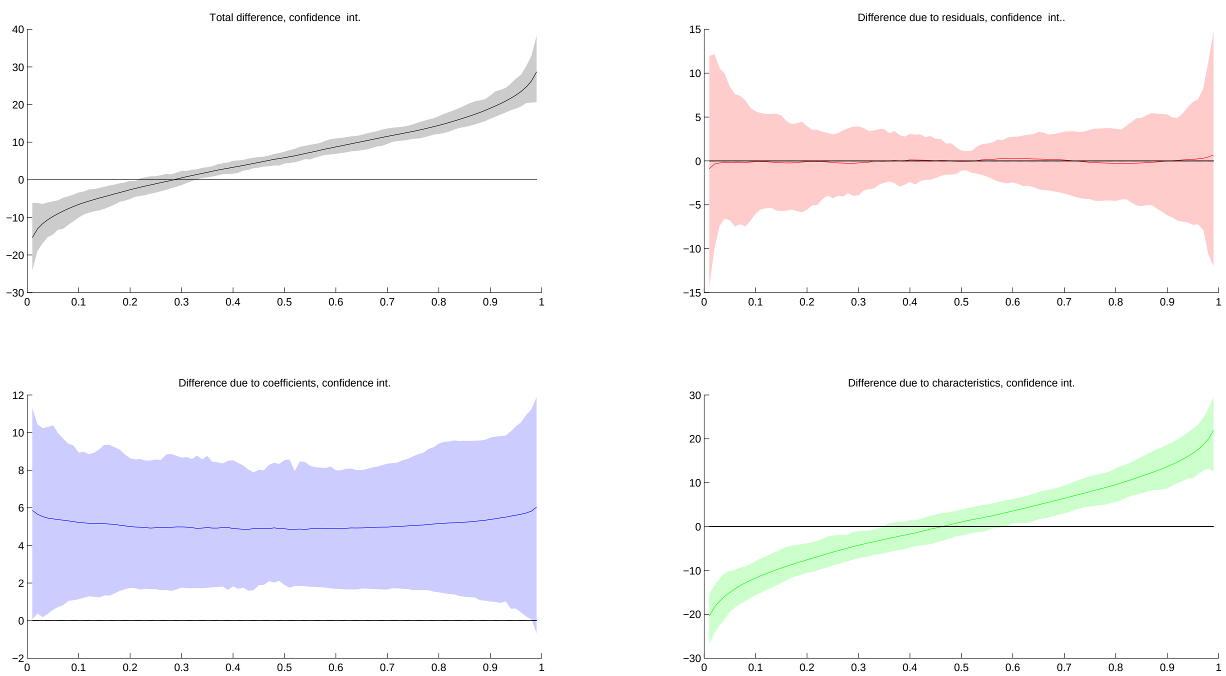

FIGURE A.3. MC Simulation of decomposition, Unbalanced Panel Data/-
$N = 1000,\ MCRep = 100$

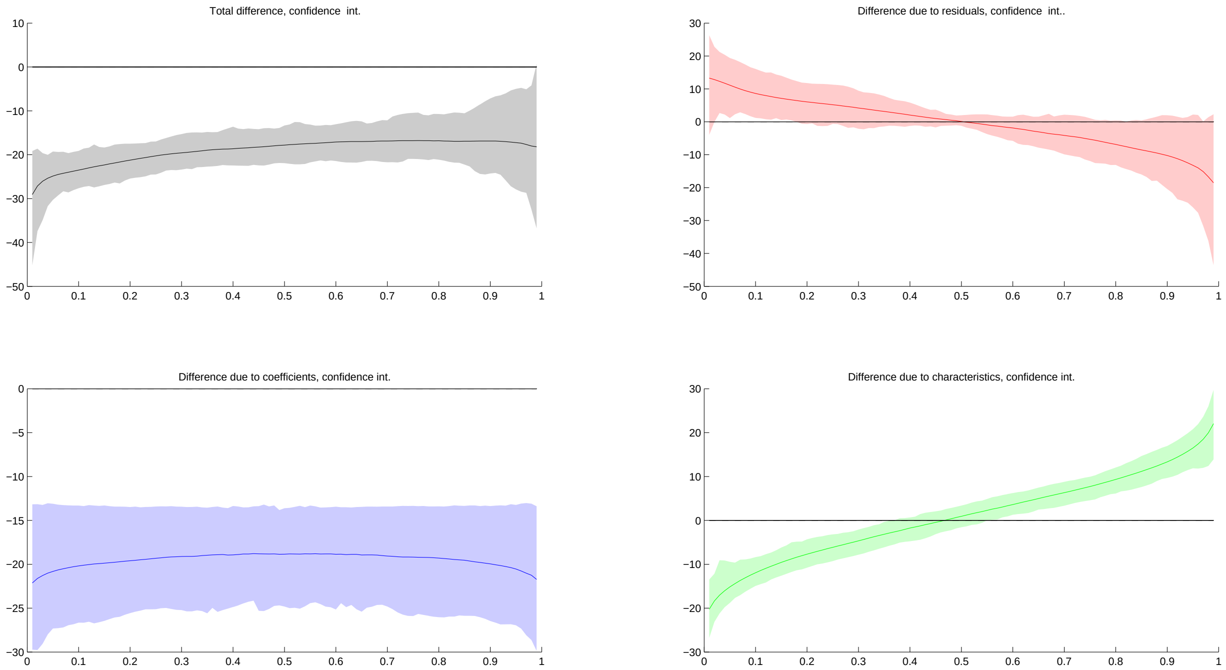

FIGURE A.4. MC Simulation of IPW decomposition, Unbalanced Panel Data/IPW $N = 1000,\ MCRep = 100$

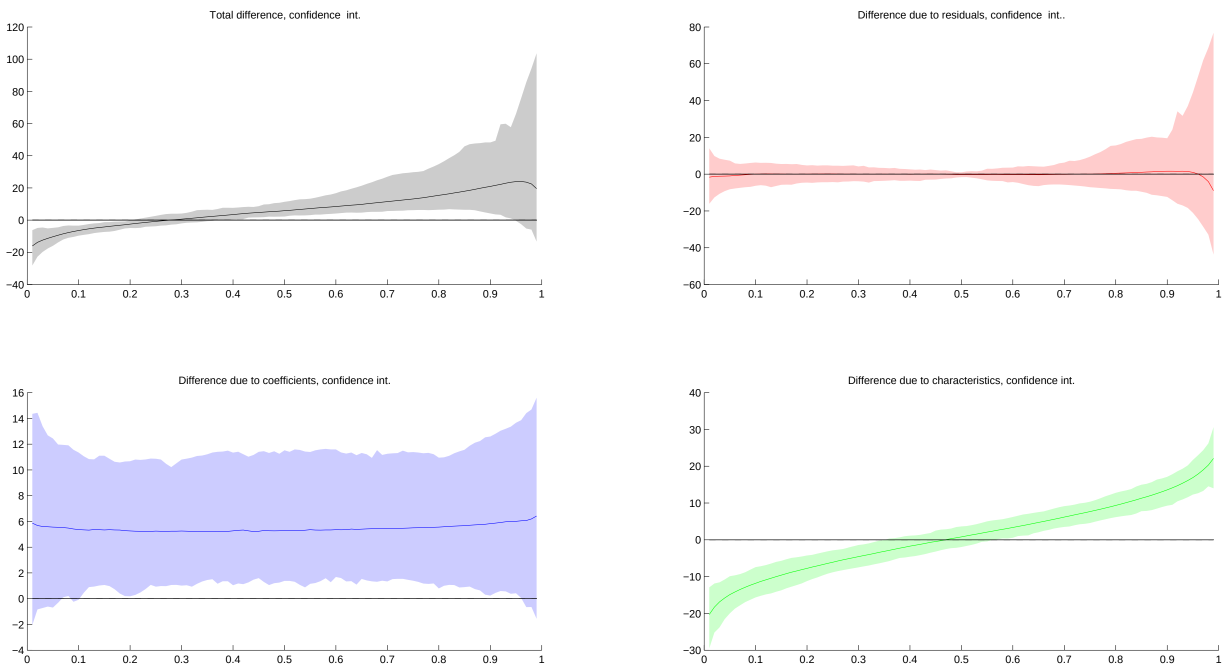

FIGURE A.5. MC Simulation of decomposition, Panel Data, $N = 10,000$, $MCRep = 100$

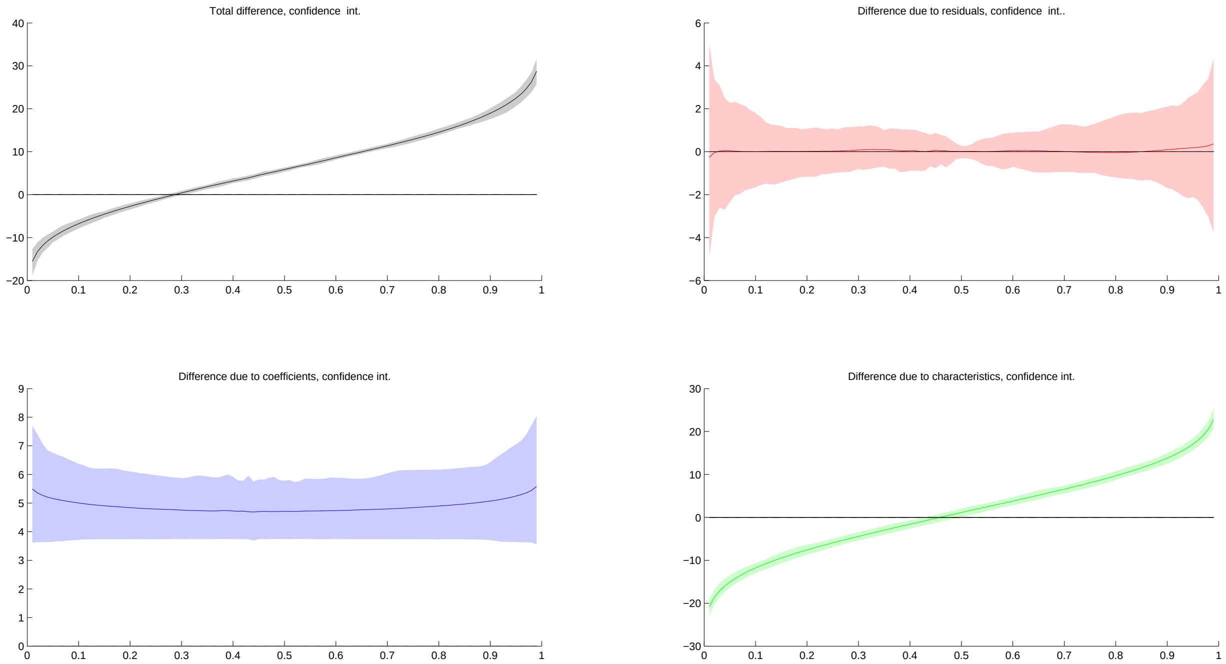

FIGURE A.6. MC Simulation of decomposition, Unbalanced Panel Data, $N = 10,000$, $MCRep = 100$

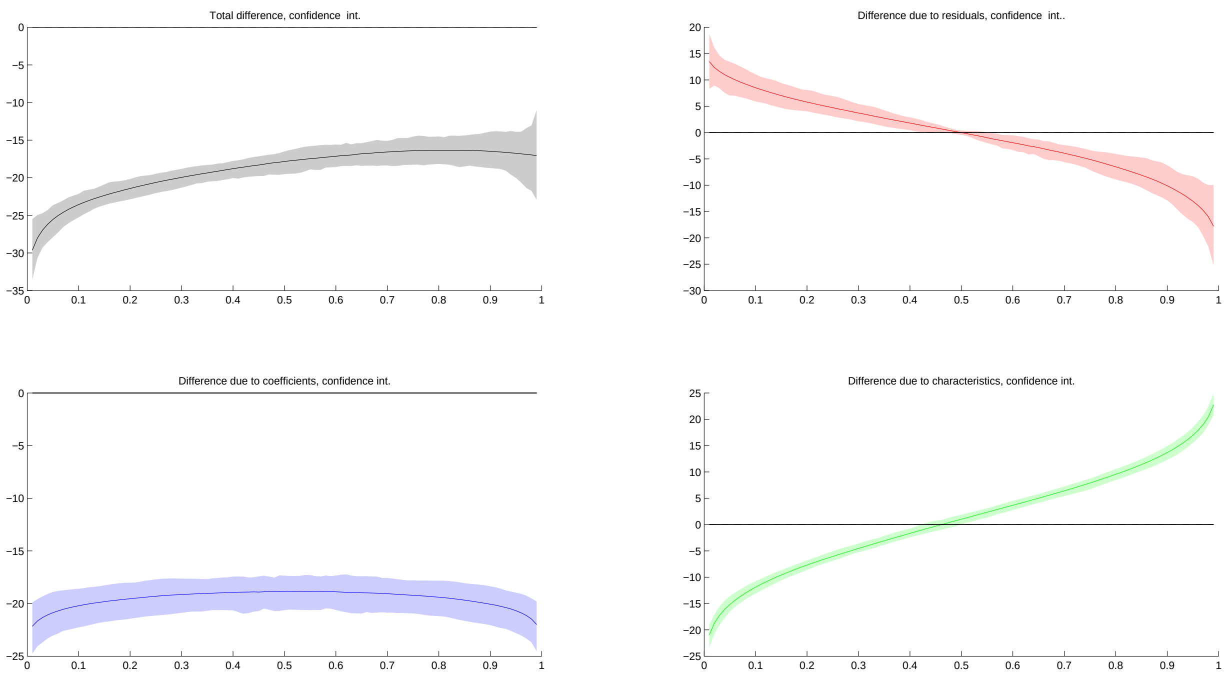

Figure A.7. MC Simulation of IPW decomposition, Unbalanced Panel Data , $N = 10,000$, $MCRep = 100$