# Estimating Treatment Effects at the Quantile

Jesse Bruhn and Juan Carvajal*

*Department of Economics, Boston University*

E-mail: jessebr@bu.edu

## Abstract

Suppose you have access to data from a randomized control trial or a convincing natural experiment, and you would like to answer a question of the following form: "What is the impact of treatment conditional on the subject occupying a specific quantile of the untreated distribution of outcomes?" Questions like this are typically answered by explicitly (or implicitly) making a rank invariance assumption and calculating Quantile Treatment Effects. But rank invariance assumptions are strong and rule out interesting economic behavior. For example, rank invariance rules out an important prediction from poverty trap theory: that individuals in the bottom of the income distribution who experience an exogenous wealth shock will come to occupy a higher place in the income distribution over the long run. In this paper, we provide an alternative set of assumptions to identify treatment effects at the quantile and prove that the resulting estimator is consistent and root-n asymptotically normal. We also provide evidence from simulation that indicates the estimator performs well in finite sample. Last, we provide applications to a micro-finance RCT in India and to a job training program evaluation in the United States.

## Introduction

Suppose you have access to data from a Randomized Control Trial or a convincing natural experiment, and you would like to answer a question of the following form: "What is the impact of treatment conditional on the subject occupying specific quantiles of the untreated distribution of outcomes?" This type of question arises naturally when conducting cost benefit analysis of program implementation and when testing predictions of economic theory that vary based on an agent's place in the distribution. For example, poverty trap theory predicts a cutoff level of counterfactual wealth such that those below the cutoff should respond more strongly to an exogenous wealth shock than those above the cutoff. As a result, if the entire population were treated, the theory predicts that in the long run some mass of individuals in the bottom of the counterfactual distribution of wealth would come to occupy positions in the middle and top of the treatment distribution.

Currently, there are two statistical methods that allow a researcher to provide evidence regarding this type of question. The first method involves making some form of a rank invariance assumption with respect to the distribution of potential outcomes and then computing standard quantile treatment effects. However, rank invariance assumptions are restrictive and may or may not be plausible depending upon the context of the question. In poverty trap theory, for example, an assumption of rank invariance effectively rules out some of the main implications of the theory we would like to test: that some swapping of ranks occurs across distributions. The second method involves stratifying the sample on some observable characteristics and then estimating the average treatment effects over subsamples. This requires the researcher to make arbitrary decisions with respect to the specific subsamples to consider and only provides an average treatment effect over some bin of the predicted value of the counterfactual outcome (and not treatment effects at the counterfactual quantile of interest).

In this paper we develop an alternative set of conditions which identify treatment effects at the quantile (TATQ). In particular, we will use predetermined characteristics (for example, anything known about the observation at baseline) to pin down the appropriate treatment group comparison even in the presence of rank invariance. However, this comes at some cost. In place of rank invariance we will require conditional orthogonality of the errors across potential outcomes. Fortunately, this assumptions is testable whereas rank invariance is not. Under these conditions, we derive an estimator that is consistent and root-n normal with standard errors that are consistently estimated via bootstrap. Evidence from simulation suggests that when the conditions for consistency are satisfied the estimation procedure is unbiassed in finite sample.

This paper is organized into five sections. Section one discusses related literature. Section two provides a statistical framework and discusses our main results. Section three presents evidence from simulation. Section four provides two empirical applications. The first is an application to the evaluation of a micro finance randomized control trial in India. The second is an application to a job training program evaluation in the United States. Section five concludes.

## Related Literature

The estimation procedure developed in this paper further develops the econometric literature on treatment effects, quantile regression, and endogenous stratification. I will summarize the relationship of this paper to each of these literatures in turn.

The treatment effect literature is large and well summarized in standard econometric references. The most important paper related to the present work is that of Heckman, Smith, and Clements (1997) which discusses the importance of response heterogeneity in the context of program evaluation. While the primary focus of Heckman, Smith, and Clements (1997) is bounding the distribution of treatment effects $F(Y_1 - Y_0)$, they do consider $F(Y_1 - Y_0 | Y_0 = y_0^u)$ as a potential quantity of interest and discuss its importance and potential use for program evaluation purposes. Recent scholarship has focused on generalizing the treatment effects literature by weakening assumptions such as monotonicity (see de Chaisemartin, 2014) or developing methods to recover local average treatment effects from other experimental designs (see Hahn, 2001 and Hirano et al, 2003). This paper extends this literature by allowing researchers to estimate quantities that have an interpretation that is similar in spirit to the average treatment effect (in the sense that the TATQ communicates information that is an average over individuals as opposed to a shift in distributions) but in a quantile setting.

The quantile regression literature is summarized well in Koenker (2005). This literature has seen much growth in recent years, with much of the focus on enabling the researcher to recover quantile treatment effects from different experimental designs. Many of these techniques, however, involve making assumptions of rank invariance even when identifying distributional shifts. For example, Chernozhukov and Hansen (2005) develop an instrumental variable method for estimating quantile treatment effects. But in their environment, even identification of distributional shifts require a rank invariance assumption. In the context of deriving efficiency bounds for QTE estimates in the presence of selection, Firpo (2004) provides a nice discussion of the role of rank invariance and weak rank invariance assumptions in QTE estimates. Although we do not explore it in this paper, it is possible that the techniques developed here may also have applications in other settings where rank invariance is important. For example, we suspect that the machinery developed here may make it possible to identify distributional shifts via quantile instrumental variable techniques without making a rank invariance assumption.

Abadie and Chingos (2014) provide theory for endogenous stratification methods which recover average treatment effects over bins of predicted control outcomes. The estimation procedure we propose here shares many features with endogenous stratification. For example, both use predetermined characteristics to construct an appropriate comparison based on predicted values. Thus, one interpretation of the technique developed in this paper is that it brings endogenous stratification methods to quantile estimation.

## Statistical Model

Let $Y_i \sim F_i(y_i)$ for $i \in \{0, 1\}$ denote the distribution of potential outcomes, with 0 indicating the counterfactual outcome and 1 indicating the treatment outcome. We will assume that the treatment indicator $\tau$ is orthogonal to both potential outcomes. Note that we observe only $Y = \tau Y_1 + (1 - \tau)Y_0$.

We will denote a given quantile of the outcome distribution by $y_i^u = F_i^{-1}(u)$ where $u$ denotes the quantile of interest and is a realization from $U \sim \mathbb{U}(0, 1)$, and $F_i^{-1}(u) = \inf_y\{y|F(y) \geq u\}$. Last, we will assume the existence of a vector of covariates $Z \sim G(z)$ which is orthogonal to treatment status (for example, any predetermined characteristics). From this we define three quantities of interest.

Recall that the average treatment affect (ATE) is defined as:

$$\delta_{ATE} = \mathbb{E}[Y_1 - Y_0] \tag{1}$$

Next, we have the standard quantile treatment effect (QTE):

$$\delta(u) = y_1^u - y_0^u \tag{2}$$

Last, we will define the treatment effect at the quantile (TATQ) as:

$$\delta^*(u) = \mathbb{E}[Y_1 - Y_0|Y_0 = y_0^u] \tag{3}$$

One way to identify quantity (3) would be to make a rank invariance assumption. For example, if $Y_1$ and $Y_0$ are comonotonic, we have that for $(Y_0, Y_1) = (F_0^{-1}(U), F_1^{-1}(\tilde{U}))$, it must be that $U = \tilde{U}$. This amounts to assuming that any observation at the $u$ quantile of the counterfactual distribution would occupy the $u$ quantile of the treatment distribution. This situation is illustrated in figure 1. Under this assumption, expression (3) simplifies to:

$$\delta^*(u) = y_1^u - y_0^u = \delta(u) \tag{4}$$

A weaker assumption would be that $\mathbb{E}[Y_1|Y_0 = y_0^u] = y_1^u$. This assumptions is interpreted as saying that to the extent rank invariance occurs, it is the result of sampling error or some other source of purely random variation. Hence, on average, an observation occupying the $u$ quantile of the treatment distribution would occupy the $u$ quantile of the control distribution and equation (4) would still be true.

To see how we may identify quantity (3) without an assumption of rank invariance, let $\varepsilon_i \equiv Y_i - \mathbb{E}[Y_i|z]$. To further simplify notation, denote the conditional expectation function by $\gamma_i(z)$, so that $\mathbb{E}[Y_i|z] = \gamma_i(z)$. Then we may rewrite equation (3) as:

$$\delta^*(u) = \mathbb{E}[\gamma_1(z)|y_0^u] + \mathbb{E}[\varepsilon_1|y_0^u] - y_0^u \tag{5}$$

If we assume that errors across potential outcomes are orthogonal conditional on the predetermined characteristic (i.e. that $\varepsilon_0 \perp \varepsilon_1$). Then equation (5) becomes:

$$\delta^*(u) = \mathbb{E}[\gamma_1(z)|y_0^u] - y_0^u = \int_Z \mathbb{E}(Y_1|z) f(z|y_0^u) dz - y_0^u \tag{6}$$

Figure (2) illustrates this situation in the population. Conditioning on $y_0^u$ induces a distribution over $z$ given by $f(z|y_0^u)$ that provides the appropriate "weight" we should give to each observation when averaging over predicted values from the conditional expectation function $\mathbb{E}(Y_1|z)$.

In sample, since $y_0^u$ is consistently estimated via quantile regression, then provided we have some reasonable estimate of the conditional expectation function $\mathbb{E}(Y_1|z)$ (for example via local linear regression), all that is left is to find the appropriate weighting. We propose to do this by averaging over the predicted values of observations that are "close" to $y_0^u$. This intuition is presented in figure (3).

Let $h_N(\alpha)$ be a function that varies with the sample size $N$ and a free parameter $\alpha$ which affects the size of the estimation window independently from the sample size. Further, let $\hat{y}_0^u$ denote the estimate of $y_0^u$ given by the $u$ sample quantile. From this, we define the interval:

$$A_N = A_N(\alpha) = [\hat{y}_0^u - h_N(\alpha), \hat{y}_0^u + h_N(\alpha)] \tag{7}$$

Let $\hat{\gamma}_i(z)$ denote a consistent estimate of the conditional expectation function. Then a natural sample analogue to equation (6) is given by:

$$\hat{\delta}^*(u) = \frac{\sum_{i=1}^N \hat{\gamma}_1(z_i)\mathbb{1}(y_{i0} \in A_N)}{\sum_{i=1}^N \mathbb{1}(y_{i0} \in A_N)} - \hat{y}_0^u \tag{8}$$

Note that introducing the window function $\mathbb{1}(y_{i0} \in A_N)$ when $\hat{y}_0^u$ is an estimated quantity poses a significant technical challenge, since the indicator is not continuous and hence we cannot apply the continuous mapping theorem. To simplify, we will use a continuous approximation of the indicator function: $\chi(y_{i0} \in A_N) \approx \mathbb{1}(y_{i0} \in A_N)$[1]. We will postpone a discussion of how to choose $\chi$ (and the associated function $h_N$) until the results from simulation.

What is important is that we assume $h_N(\alpha)$ converges in probability to $h(\alpha) > 0$. This implies that $A_N$ converges in probability to the set $A = \{y_0^u - h(\alpha), y_0^u + h(\alpha)\}$, so that by the continuous mapping theorem we have that $\chi(y_{i0} \in A_N)$ converges to $\chi(y_{i0} \in A)$. Since we can approximate the indicator function to within an arbitrary degree of error, for large N we have that $\chi(y_{i0} \in A_N) \approx \mathbb{1}(y_{i0} \in A)$ up to some non-stochastic approximation error.

We now prove the main result of this paper.

**Theorem 1** *Suppose we have $N$ i.i.d observations of the random variable $Y = \tau Y_1 + (1-\tau)Y_0$ such that $\tau \perp (Y_1, Y_0)$. Consider the following estimator of $\mathbb{E}[Y_1 - Y_0 | Y_0 = y_0^u]$:*

$$\hat{\delta}^*(u) = \frac{\sum_{i=1}^N \hat{\gamma}_1(z_i)\chi(y_{i0} \in A_N)}{\sum_{i=1}^N \chi(y_{i0} \in A_N)} - \hat{y}_0^u \tag{9}$$

[1]We believe the result will still hold with the indicator function, but have not proved it at this time

*Where $\hat{\gamma}_1(z)$ is a consistent estimator of $\gamma_1(z) = \mathbb{E}[Y_1|z]$, $\hat{y}_0^u$ is the $u$ sample quantile of the observed control outcomes, $\chi(y_{i0} \in A_N)$ is a continuous approximation to the indicator function $\mathbb{1}(y_{i0} \in A_N)$, and the sequence of window sets $A_N$ is defined to be the interval $A_N = A_N(\alpha) \equiv [\hat{y}_0^u - h_N(\alpha), \hat{y}_0^u + h_N(\alpha)]$ with the length of the window determined by the function $h_N(\alpha)$ which varies with the sample size and a free parameter $\alpha$. Suppose further that the following conditions are met:*

**a** $\varepsilon_0 \perp \varepsilon_1$

**b** *$\hat{y}_0^u$ is a consistent estimate of $y_0^u$.*

**c** *$h_N(\alpha)$ converges in probability to $h(\alpha) > 0$*

**d** *For any $h(\alpha) > 0$, $Y_0$ has support on the interval $A = [y_0^u - h(\alpha), y_0^u + h(\alpha)]$ .*

**e** *$\hat{y}_0^u$ is root-n asymptotically normal.*

**f** *$Var(X)$ is finite where $X = \gamma_1(Z)\mathbb{1}(Y_0 \in A)\mathbb{E}[\mathbb{1}(Y_0 \in A)] - \mathbb{E}[\gamma_1(Z)\mathbb{1}(Y_0 \in A)]\chi(Y_0 \in A)$*

*Then the estimator $\hat{\delta}^*(u)$ is consistent and root-n asymptotically normal.*

**Proof** First, note that condition (a) allows us to write the quantity of interest in the population as given in equation (6). Now we consider the sample analogue given by equation (9). Note that condition (b) ensures that $\hat{y}_0^u$ is consistent for $y_0^u$. Hence, we need only show that the first term on the right hand side of equation (9) is consistent for $\mathbb{E}[Y_1|y_0^u]$ and consistency of $\delta^*(u)$ follows immediately.

To simplify notation, let $\omega(u) \equiv \frac{\sum_{i=1}^N \hat{\gamma}_1(z_i)\chi(y_{i0} \in A_N)}{\sum_{i=1}^N \chi(y_{i0} \in A_N)}$. Note that $\hat{\gamma}(z_i) = \gamma(z_i) + op(1)$. Observe that by condition (c) we have that $\text{plim}_{N\to\infty} A_N = A$. Hence, by the continuous mapping theorem we get that $\text{plim}_{N\to\infty} \chi(y_{i0} \in A_N) = \chi(y_{i0} \in A)$ so that we may write $\chi(y_{i0} \in A_N) = \chi(y_{i0} \in A) + op(1)$. It follows that $\hat{\gamma}(z_i)\chi(y_{i0} \in A_N) = \gamma(z_i)\chi(y_{i0} \in A) + op(1)$. Thus we may write:

$$\omega(u) = \frac{\sum_{i=1}^{N} \gamma_1(z_i)\chi(y_{i0} \in A)/N + op(1)}{\sum_{i=1}^{N} \chi(y_{i0} \in A)/N + op(1)} \tag{10}$$

Taking the probability limit of equation (10) gives:

$$\underset{N\to\infty}{\text{plim}}\,\omega(u) = \frac{\mathbb{E}[\gamma_1(z)\chi(Y_0 \in A)]}{\mathbb{E}[\chi(Y_0 \in A)]} \approx \frac{\mathbb{E}[\gamma_1(z)\mathbb{1}(Y_0 \in A)]}{\mathbb{E}[\mathbb{1}(Y_0 \in A)]} = \mathbb{E}[\gamma_1(z)|y_0 \in A] \tag{11}$$

Where the first equality in equation (11) follows from the law of large numbers, Slutsky's theorem, and condition (d). The second (approximate) equality follows from the fact that $\chi(Y_0 \in A)$ can approximate the indicator to within an arbitrary degree of error. And the third equality follows from the definition of a conditional expectation.

Observe that $\lim_{h(\alpha)\to 0} A = \lim_{h(\alpha)\to 0} [y_0^u - h(\alpha), y_0^u + h(\alpha)] = \{y_0^u\}$. Hence, provided $h(\alpha)$ is small, we have that:

$$\mathbb{E}[\gamma_1(z)|y_0 \in A] \approx \mathbb{E}[\gamma_1(z)|y_0 = y_0^u] \tag{12}$$

And we see that $\omega(u)$ is consistent for $\mathbb{E}[Y_1|y_0^u]$ up to some non-stochastic approximation error (coming from $h(\alpha)$ and $\chi(y_{i0} \in A_N)$) which can be made arbitrarily small. Consistency of $\hat{\delta}^*(u)$ follows immediately.

To see that $\hat{\delta}^*(u)$ is asymptotically normal, observe that by condition (e) $\hat{y}_0^u$ is asymptotically normal. Hence, if we can show that $\omega(u)$ is asymptotically normal, the result follows immediately. Now consider the following quantity:

$$\sqrt{N}(\omega(u) - \mathbb{E}[\gamma_1(z)|Y_0 \in A]) = \sqrt{N}\left(\frac{\sum_{i=1}^{N} \hat{\gamma}_1(z_i)\chi(y_{i0} \in A_N)}{\sum_{i=1}^{N} \chi(y_{i0} \in A_N)} - \frac{\mathbb{E}[\gamma_1(z)\mathbb{1}(Y_0 \in A)]}{\mathbb{E}[\mathbb{1}(Y_0 \in A)]}\right) \tag{13}$$

We can write the right hand side of equation (13) as follows:

$$\frac{(\sum_{i=1}^{N}\hat{\gamma}_1(z_i)\chi(y_{i0}\in A_N)\mathbb{E}[\mathbb{1}(Y_0\in A)] - \mathbb{E}[\gamma_1(z)\mathbb{1}(Y_0\in A)]\sum_{i=1}^{N}\chi(y_{i0}\in A_N))/\sqrt{N}}{\mathbb{E}[\mathbb{1}(Y_0\in A)]\sum_{i=1}^{N}\chi(y_{i0}\in A_N)/N} \tag{14}$$

First, consider the denominator of equation (14), and observe that because we can approximate the indicator function to within an arbitrary degree of error, we have:

$$\mathbb{E}[\mathbb{1}(Y_0\in A)]\frac{\sum_{i=1}^{N}\chi(y_{i0}\in A_N)}{N} \approx \mathbb{E}[\mathbb{1}(Y_0\in A)]\frac{\sum_{i=1}^{N}\mathbb{1}(y_{i0}\in A)}{N} + op(1) \tag{15}$$

Hence, the denominator of equation (14) converges in probability to the constant $\mathbb{E}[\mathbb{1}(Y_0 \in A)]^2$. Let $X_i = \gamma_1(z_i)\mathbb{1}(y_{i0}\in A)\mathbb{E}[\mathbb{1}(Y_0\in A)] - \mathbb{E}[\gamma_1(z)\mathbb{1}(Y_0\in A)]\chi(y_{i0}\in A)$. Then, consider that the numerator of equation (14) is approximately equal to:

$$\frac{\sum_{i=1}^{N}X_i}{\sqrt{N}} + op(1) \tag{16}$$

Since the $X_i$ are i.i.d., $Var(x) < \infty$ by assumption, and observing that $\mathbb{E}[X_i] = 0$, then as N gets large (and suppressing the stochastic approximation error for clarity) we can apply the standard central limit theorem to get:

$$(\frac{\sum_{i=1}^{N}X_i}{\sqrt{N}}) = \sqrt{N}(\bar{X} - 0) \sim_a N(0, Var(X)) \tag{17}$$

Thus we can apply Slutsky to equation (14) to get that:

$$\sqrt{N}(\omega(u) - \mathbb{E}[\gamma_1(z)|Y_0\in A]) \sim_a N(0, \frac{Var(X)}{\mathbb{E}[\mathbb{1}(Y_0\in A)]^4}) \tag{18}$$

And the root-n asymptotic normality of $\hat{\delta}^*(u)$ follows immediately. ∎

A few notes about the estimator. Because the test statistic is asymptotically normal, it satisfies the conditions for inference with bootstrap (Horowitz, 2001). While it should be possible to estimate the asymptotic variance directly, we have not taken that approach in this paper. It is also worth noting that the component of the asymptotic variance that comes from estimating $\omega(u)$ is influenced heavily by the probability density of $F_1(y)$ on the set $A$. Practically, this may make inference difficult in the tails of the distribution.

Another issue we have skirted so far is the proper way to collapse the estimation window. By assuming that the function $h_N()$ converges to a strictly positive limit we are potentially introducing asymptotic bias if the treatment varies non-linearly with counter factual quantiles. This can be seen by noting that the asymptotic distribution is centered on $\mathbb{E}[\gamma_1(z)|Y_0 \in A]$ (and not $\mathbb{E}[\gamma_1(z)|Y_0 = y_0^u]$). Intuitively, we want the window to collapse slowly enough that the true value of $y_0^u$ is contained in $A$ with probability one, but fast enough that $A \approx \{y_0^u\}$.

While we do not treat this issue formally in the current version of the paper, it may be possible to estimate the bias up to some asymptotic order with bootstrap. One advantage of leaving the window "open" in this fashion and manually estimating the bias it introduces is that it may be possible to subsequently estimate an optimal $\alpha$ according to some criterion function such as mean square error. This procedure may be beneficial on efficiency grounds. The other possibility for handling this issue would be to force the function $h_n()$ to converge to 0 at an "appropriate" rate which we have yet to determine.

# Results from Simulation

The statistical theory presented in the last section leaves open several question about the usefulness of the estimator in practice. First, the asymptotic theory provides no guidance as to the degree of bias that may be present in finite sample. We will provide evidence from simulation which suggests that, at least for some data generating processes and when the window is small, the estimation procedure proposed in this paper is, in fact, unbiased. Second, in most practical applications, both rank invariance and orthogonality across errors are strong assumptions. Thus, it is natural to ask the question "Which estimator performs better in the presence of misspecification." Here we provide evidence from simulation to argue that conditional on some degree of rank invariance, the estimation procedure we propose in this paper outperforms a conventional quantile regression estimate when the TATQ is the object of interest. Last, the asymptotic theory provides no guidance on how to choose the window function $\chi$ in practice. Conditional on using a window with a strictly positive size, there does not appear to be significant differences among the choice of window.

We should note, however, that all results in this section (especially with respect to choice of window) should be taken as very preliminary. For all three questions, more systematic simulation must be done before we can draw more definitive conclusions. In particular, until the asymptotic theory with respect to window collapse is more well developed, it is impossible to say with certainty that one window function will outperform another, since answering the former question may provide insight on how to optimally implement a given window.

## Finite Sample Bias

Let the Data Generating Process be given by:

$$Y_i = \alpha_i + Z\beta_i + \epsilon_i \qquad i \in \{0, 1\} \tag{19}$$

We have explored a variety of distributions for the above parameters and variables (including normal, uniform, and beta). For brevity, the results we present here are for the case where $\alpha_0 = 2$, $\alpha_1 = 2.5$, $\beta_0 = 1$, $\beta_1 = 2$, $Z \sim N(0, 1)$, and $\epsilon_i \sim N(0, 1)$, though varying these values/distributions does not change the results. The sample size for the particular graph we present below is 1,000, although the results are virtually identical for smaller sample sizes as well. Since the true TATQ is often difficult to solve analytically, we calculate the true treatment effect at the quantile via monte-carlo simulation.

In this simulation, we implement the estimation procedure in this paper with an indicator function for the window, and a window size of 0 (i.e. using only the predicted value of the observation at the u sample quantile). Figure 4 presents, at each quantile, the average bias of our estimation procedure (labeled av.bias.tatq), the average bias of quantile regression on a treatment dummy (labeled av.bias.QTE.noZ), and the average bias of quantile regression on a treatment dummy and Z (labeled av.bias.QTE.withZ) over 500 monte-carlo trials.

The results are suggestive that the estimation procedure proposed in this paper is unbiassed. Additionally, we observe that in the presence of rank invariance, quantile regression is biassed even when controlling for the predetermined characteristic. In other words, controlling for Z in the quantile regression does not control for rank variance *even* when the systematic rank variance operates entirely through Z.

## Estimation with Misspecification

To understand the consequences of misspecification, we ran the same simulation described in the preceding section with the important difference that, instead of the errors being orthogonal, we had that $\epsilon_1 = \gamma\epsilon_0 + x$. While we have explored various choices for the distribution of $x$ and the value of $\gamma$, the results of a simulation with $x \sim N(0,1)$ and $\gamma = .3$ are presented in Figure 5.

For the estimation procedure proposed in this paper, positive correlation in the errors across potential outcomes produces positive bias in the lower tail of the distribution and negative bias in the upper tail. This result should be intuitive. If there is positive correlation in the errors, those in the lower tail of the control distribution had some combination of a small draw of $Z$ and a bad shock $\epsilon_0$. Because estimation of the conditional expectation function does not account for the correlation between $\epsilon_1$ and $\epsilon_0$, this leads to on overestimate of the predicted value of $Y_1$ for those with small $Z$. A similar dynamic produces negative bias in the upper tail.

It is interesting to note, however, that the estimates from quantile regression are also biassed under this data generating process even after controlling for Z. In general, the bias is opposite in sign to the bias generated by the estimation procedure proposed in this paper.

This suggests that there may be very weak assumptions that would allow us to bound the treatment effect between the two estimates. However, we have not explored this connection further at this time.

### Window Choice

This section presents preliminary results from an investigation into the choice of optimal window. Here, we run the simulation from above but instead of using the singleton window with an indicator function, we use Uniform, Bartlett, Parzen, and Tukey windows, with window size (i.e. those observations given non-zero weight in the estimation) fixed at 10 "lags" and "leads" around the estimated $\hat{y}_0^u$. To summarize the performance of these windows, we calculated the standard deviation across point estimates for the 500 Monte Carlo trials at each quantile for each window. To condense this information into one easily interpretable statistic summarizing the performance of the window, for a given window we averaged these standard deviations across quantiles. The results are presented in Table 1, along with the same summary measure for the estimate using a singleton window.

These results are suggestive that estimation with a continuous window improves the performance of the estimator. However, the exact choice of window function does not appear to matter (at least when the "size" of the window is fixed). As mentioned earlier, this result should be seen as very preliminary, since the asymptotic theory around the exact way to handle estimation with a window is still a work in progress.

## Applications

For applied work, the TATQ is most likely to be a qunatity of interest in two contexts. The first, and most obvious, is when a prediction of economic theory involves some systematic swapping of ranks across the counter factual and treatment distributions of outcomes. These kinds of predictions are common in the development literature, where poverty trap theory predicts that individuals stuck in the bottom tail of the income distribution due to some credit market imperfection should systematically move to higher points in the distribution given the exogenous relaxing of the credit constraint. See, for example, Bleakley (2013) or Banerjee and Newman (1993). Along these lines, we provide an application of our estimation procedure to a recent randomized control trial involving for profit micro-credit in India.

The second place that the TATQ is likely to be of interest is in program evaluation. As discussed in Heckman, Smith and Clements (1997), knowing $F(Y_1 - Y_0|Y_0 = y_0^u)$ is useful for satisfying normative criteria. For concreteness, let's suppose the outcome variable is income. Then given limited resources for program implementation, it is natural to target treatment to those who would benefit most.

One way to accomplish this goal would be via endogenous stratification methods as in Abadie and Chingos (2013). This method has the attractive feature that it has a built in a model for predicting which group should receive treatment. However, there are a few flaws with this approach. First, it is not clear how finely the researcher should subset the data in order to detect the heterogeneity of interest. Because we estimate impacts at the quantile, no such consideration arises with our methods. Second, in some contexts, the endogenous stratification method may understate the true impact of treatment. For example, in the context of a job training program, we could view the error $\epsilon_0$ as some transitory shock to employability that would have occurred in absence of treatment. The endogenous stratification method does not include as part of its average treatment effect the fact that treated individual will on average draw a mean zero error upon reentry into the labor market.

Whether to include the fact that the program participant receives a new error upon assignment to treatment is not obvious since we assume that conditional on predetermined characteristics the errors are orthogonal across potential outcomes. When computing average treatment effects for the whole population, this issue never arises. But if we could observe both potential outcomes for the entire population and were tasked with implementing the program in such a way that all those who benefit from treatment in excess of cost receive it, one natural method would be to compute $F(Y_1 - Y_0|Y_0 = y_0^u)$ for the population, and provide treatment to all individuals at quantiles for whom $\mathbb{E}[Y_1|Y_0 = y_0^u] - y_0^u > c$ where $c$ is the cost of treatment. This method would include the benefit of receiving a new error. In this spirit, we provide an application of the estimation procedure on the same subset of the Job Training Partneryship Act (JTPA) data used in Abadie and Chingos (2013).

## Micro-Finance Randomized Control Trial

Banerjee et al (2015) examine the impacts that providing access to group lending micro finance for females in poor neighborhoods in Hyderabad had on the outcomes of its residents 1 and 2 years later.

The basics of the experimental setup were as follows:

- The authors worked with the for profit micro finance company Sapndana to determine 104 poor neighborhoods in Hyderabad where the lender was indifferent to opening a branch. Each neighborhood contained 46 to 450 households.

- The neighborhoods were paired based on similarity of demographic characteristics and concerns about geographical separation of control and treatment areas. Within each pair, one neighborhood was randomly assigned to treatment and the other to control. Spandana then opened a branch in the treatment neighborhoods.

- Residents of the neighborhood were surveyed prior to the opening of the branch. Unfortunately, due to time constraints imposed on the researchers by Spandana, the survey was not conducted in an ideal fashion. As a result, families surveyed at baseline were not intentionally surveyed at subsequent endlines. Additionally, the authors consider the baseline data unreliable except for use in identifying neighborhood level characteristics.

- Eighteen months after the introduction of micro finance, an average of 65 households were surveyed within each neighborhood. These same households were then surveyed again 2 years after treatment.

The headline findings of the RCT were that treatment caused an 8.4 percentage point increase in take-up of micrcredit, small business investment and profits increased, consumption did not increase, and expenditure on durable and temptation goods declined.

However, the authors argue that, especially with respect to results concerning business and borrowing, the average intent to treat effects calculated might mask significant heterogeneity of response. While underpowered, the authors provide graphs meant to illustrate this possibility using standard quantile regression estimates. But it would be no surprise that providing access to cheap credit simply shifted outcome distributions to residents in these neighborhoods. To the extent that heterogeneity in response matters in this context, we would like to isolate whether this is simply a distributional shift or true heterogeneity in the sense that the TATQ is meant to capture. In this micro finance setting, the latter is arguably more relevant for both policy and theory.

To this end, we find estimates of treatment effects at the quantile that are, in some cases, very different from the QTE estimates presented in the paper and in other cases very similar. However, in at least two cases we see effect size magnitudes that are clearly unrealistic. Further, the bias is consistent with what we see in simulation when using our estimation procedure in the presence of significant positive correlation in the errors (see figure 5).

In all cases, the TATQ estimates in this section reflect estimation with an indicator window function and a window size of 0 (i.e. using the singleton). As a result, the estimates may appear quite noisy[2]. For stratifiers, we use neighborhood dummies, household size, and

[2]Since these estimates were originally computed, the theory for the estimation procedure has made considerable progress. Since it is not clear at the moment whether this particular application will be contained in any final version of the paper, we have not invested the time in implementing the estimation procedure with a more sophisticated window function. If we did, we believe this would decrease the noise considerably.

indicators for whether the head of household was male, had no education, and had a literate spouse.

First, we examine results related to informal borrowing. A priori, it is not obvious what the impact of micro finance should be on informal borrowing. Among those households that are not credit constrained, because these micro finance loans are at below market rates, we would expect to see substitution away from current sources of credit (informal or formal) and towards the micro credit products. But among households that are credit constrained in some fashion(across all credit markets - formal and informal), it is not clear that we should expect informal borrowing to go down. If these households have access to highly productive investments that go unfunded due to credit market imperfections, it seems possible that informal borrowing could actually rise as households take on additional informal debt to supplement the funds from micro finance in pursuit of a new business venture. In other words, it possible that informal and formal credit are complemenatry goods for credit constrained households.

Figure 6 presents the results from Banerjee et al (2015). These quantile treatment effects seem largely consistent with the first story, as informal borrowing declines for nearly all quantiles.

Figure 7 presents the results from our estimation procedure. First note that the magnitudes are almost certainly wrong. As mentioned earlier, this is likely due to violations of the assumption that the errors across outcomes are orthogonal so we could take these estimates to be closer to an upper bound on the possible TATQ. Qualitatively, however, the story fits quite well with the credit market imperfection interpretation. This graph says that people who would have done very little informal borrowing in the absence of micro credit (perhaps due to a credit market imperfection) will seek to borrow more once treated. On the other hand, those who would have borrowed quite heavily in the absence of treatment borrow less once treated likely due to the substitution effect discussed previously.

Next we examine results related to business profits among households who already had a business prior to the introduction of micro credit. Figure 8 presents the results from Banerjee et al (2015). Figure 9 presents the results from the estimation procedure proposed in this paper. Note that the TATQ estimates and the QTE estimates tell essentially the same story here: micro finance has little effect on business profits conditional on the business being started prior to the introduction of micro credit.

Figures 10 and 11 give QTE and TATQ estimates of the effect of micro credit on business profits among households who started a business after the introduction of micro credit. Once again, we see unrealistic values in the TATQ estimation that are consistent with the bias that emerges when the errors are positively correlated across outcomes. Thus, we should view the the estimates here as bounds on the range of possible TATQ effects. With that said, the picture that emerges qualitatively is quite different from the one we observe with the QTE estimates.

Figures 12 and 13 present the results on business profits from the full sample of business owners at the time of the second end line survey over 2 years after the initial introduction of micro credit.

Here we see that, if we interpret the QTE estimates as treatment effects at the quantile, we would be lead to conclude that micro finance has potentially large impacts on the upper tail of the distribution. In other words, that those businesses that would have already been very profitable have simply become more profitable. The estimates from our procedure, on the other hand, tell the polar opposite story. They suggest that the overall effect is small and positive at the bottom tail of the distribution, but that profits go down among businesses that would have been profitable in the absence of treatment. Again, this could fit with a story where increased competition from previously credit constrained households now starting businesses drives down the profits of households that would have started a business anyways in the absence of treatment.

Overall, the important takeaway from this application is that the assumption made about the correlation between the errors across counterfactual outcome distributions (whether that be rank invariance or orthogonality) is consequential if we want to make statements that go beyond identifying distributional shifts.[3] If we assume rank invariance, the story that emerges is drastically different than one based on error orthogonality.

## Job Training Program

The Job Training Partnership Program (JTPA) was a program deployed by the US federal government in the late 80's that provided vocational job training to out of work youth. The details of the program are well described in Heckman, Smith, and Clements (1997) as well as Abadie and Chingos (2013). Importantly, conditional on application to the program, receipt of the vocational job training was randomized.

For estimation of the TATQ, we use a uniform window function with a window size incorporating 10 "leads" and "lags" around the estimate of $y_0^u$. For stratifiers, we follow Abadie and Chingos (2013) and use age, age squared, previous earnings, and indicators for treatment site, being married, being black, being hispanic, whether the respondent held a job in the preceding 13 weeks, and whether the respondent had a highschool diploma or GED. Again, following Abadie and Chingos, we restrict our sample to males and discard observations from three treatment sites that had a low number of observations. This leaves 849 untreated observations and 1681 treatment observations across 12 treatment sites. The outcome variable is income 30 months after treatment.

Figure 14 and 15 are two attempts to represent the intuition from figure 3 in the theoretical section of the paper in the context of the JTPA sample. The x-axis of figure 14 is variation in previous earning net of variation due to all other controls. The y-axis of figure 14 is variation in observed income for the control group net of variation due to all other stratifiers. This presentation reduces the dimensionality of the data and allows for a graphical representation. The black line is the OLS estimate of $E(Y_1|Z)$ where $Z$ is residual variation in previous earnings. In other words, the line gives expected income as it varies

[3]To be clear, Banerjee et al (2015) do not over interpret their results in the paper. I am simply using their QTE estimates to illustrate this point

with previous earnings holding constant all other stratifiers.

Figure 15 reduces the dimensionality of the data by putting the OLS estimate of $E(Y_1|Z)$ on the x-axis, where $Z$ is the entire vector of controls. The y-axis is observed income for the control group. In this case, the 45 degree line gives the appropriate predicted value for a given observation.

Figures 16 and 17 present non-parametric estimates of the density and the CDF of observable income for the treatment and control groups. Figure 18 present standard QTE estimates of the impact of the job training program on income conditional on predetermined characteristics. Figure 19 presents results from our procedure for estimating the TATQ implemented as described above.

We employ the test for error orthogonality discussed in Heckman, Smith, and Cements (1997) and fail to reject the null hypothesis that the errors are orthogonal at the 95% level. However, examining figure 19, it is clear that the effect sizes appear to suffer from the type of bias that emerges in simulation when there is significant positive correlation in the errors. The R-Square of the OLS estimate of $\mathbb{E}[Y_1|Y_0 = y_0^u]$ is low; hence, we beleieve that the failure to reject is a result of the lack of explanatory power and not solid evidence in support of the null. Thus the estimates in Figure 19 should be seen as something closer to an upper bound than an accurate point estimate.

Overall, the message from this application is that if the estimation procedure in this paper is to produce credible estimates, we need a richer set of predetermined characteristics in order for the error orthogonality assumption to be plausibly satisfied.

# Conclusion

The estimation procedure proposed in this paper still has much work that needs to be done. Priority number one is to work out the theoretical issues concerning asymptotic window collapse. Until that detail is worked out, all applications and simulations must be qualified as preliminary.

Once the asymptotic theory around window collapse is compmlete, the next priority is to do more systematic simulation around the practical application of the estimator in finite sample. In particular, we would like to know how it performs and compares to standard QTE estimates under misspecification and which window functions perform best in finite sample.

The last priority is to find data with a rich enough set of predetermined characteristics such that the error orthogonality is plausibly satisfied.

Additionally, future work should explore the possibility of finding more general conditions to bound the TATQ between standard QTE estimates and the procedure outlined in this paper.

# Bibliography

Abadie and Chingos. "Endogenous Stratification in Randomized Experiments." NBER Working Paper, (2013): No. 19742.

Banerjee, Abhijit V., and Andrew F. Newman. "Occupational choice and the process of development." Journal of political economy (1993): 274-298.

Banerjee, Duflo, Glennerster, and Kinnan. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." American Economic Journal: Applied Economics (2015), 7(1): 22-53

Banerjee, Karlan, and Zinman. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." American Economic Journal: Applied Economics (2015), 7(1): 1-21

Bleakley, Hoyt, and Joseph P. Ferrie. Up from poverty? the 1832 Cherokee land lottery and the long-run distribution of wealth. No. w19175. National Bureau of Economic Research, 2013.

Chernozhukov and Hansen. “An IV Model of Quantile Treatment Effects.” Econometrica 81: 2205-2268

de Chaisemartin. “Tolerating Defiance? Local Average Treatment Effects Without Monotonicity,” mimeo, Warwick University.

Firpo. “Efficient Semiparametric Estimation of Quantile Treatment Effects.” Econometrica (2007), 75(1): 259-276

Hahn, Todd, and Van der Klaauw. “Identification and Estimation of Treatment Effects with a Regression Discontinuity Design.” Econometrica (2001), 66(2): 315-331

Heckman, Smith, and Clements. “Making the Most Out of Program Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts.” The Review of Economic Studies (1997), 64(4): 487-535

Hirano, Imbens, and Ridder “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score.” Econometrica (2003), 71(4): 1161-1189

Horowitz, Joel L. “The bootstrap.” Handbook of econometrics 5 (2001): 3159-3228.

Imbens and Angrist. “Identification and Estimation of Local Average Treatment Effects.” Econometrica (1994) 61(2): 467-476.

Koenker. “Quantile Regression.” Cambridge University Press (2005).

# Figures and Tables

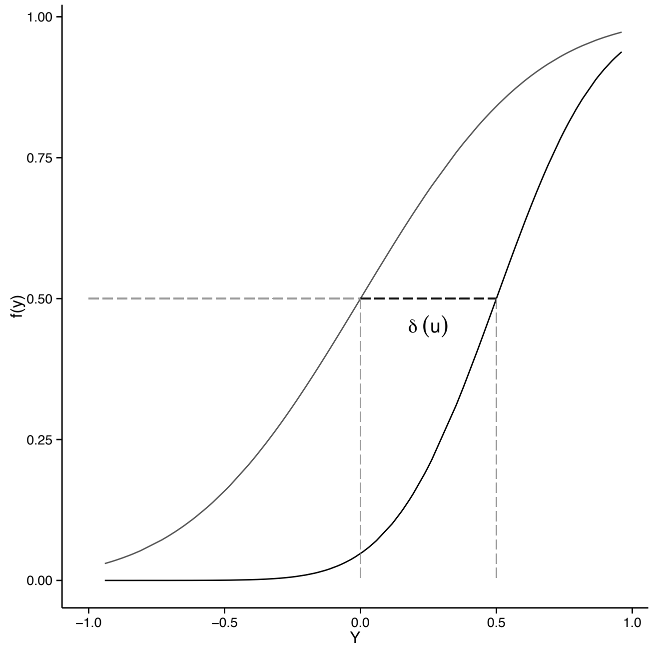


Figure 1: The QTE is equivalent to the TATQ under the assumption that $Y_1$ and $Y_0$ are comonotonic.

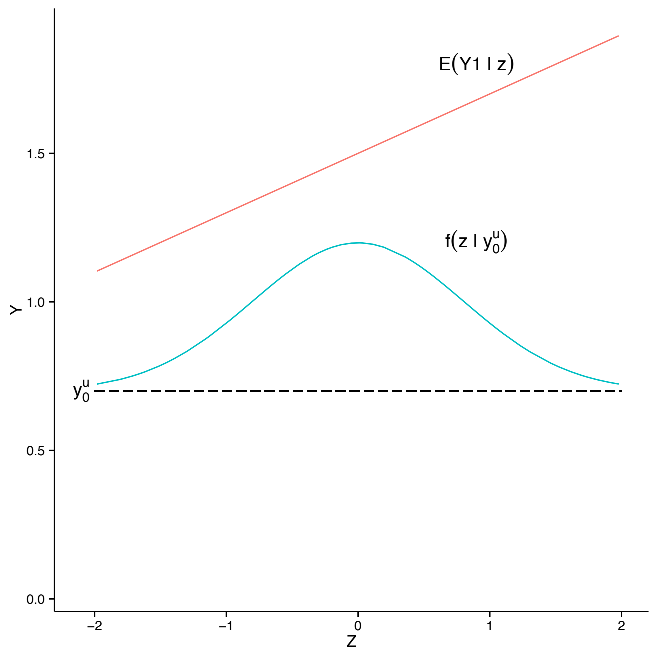


Figure 2: Calculating the TATQ in the population when errors are orthogonal across potential outcomes. After fixing a specific quantile $y_0^u$, the density $f(z|y_0^u)$ provides the appropriate weight to use when averaging over predicted outcomes $E(Y_1|z)$.

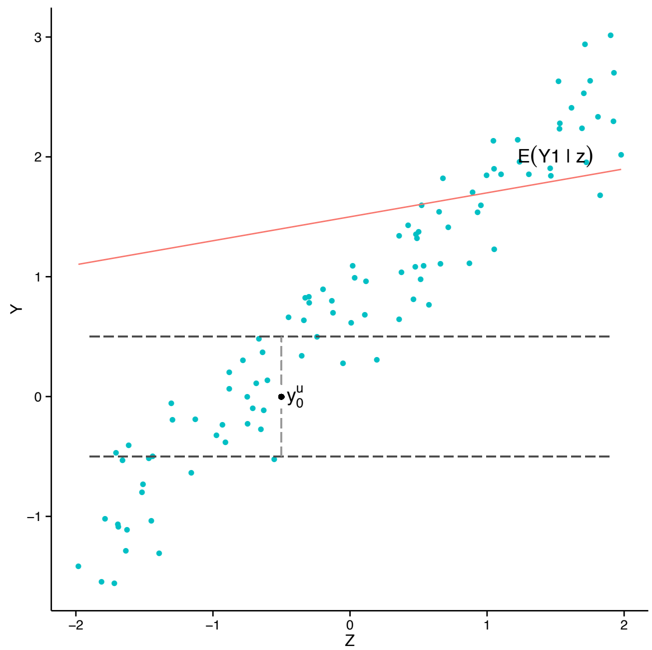


Figure 3: In sample, averaging over observations (represented by dots in the figure) that are close to the estimated $y_0^u$ is a natural empirical analogue to the situation in the population outlined in figure 2. Intuitively, as $N$ becomes large we observe more observations in an increasingly tiny window. In the limit, we recover the conditional distribution $f(z|y_0^u)$ which allows us to estimate the TATQ without a rank invariance assumption.


Figure 4: Average bias across 500 monte-carlo trials at each quantile when errors are orthogonal. The line labeled TATQ corresponds to the estimation procedure proposed in this paper. The line labeled QTE corresponds to quantile regression with a treatment dummy and no covariates. The last line corresponds to quantile regression on a treatment dummy and a set of covariates comprised of predetermined characteristics.

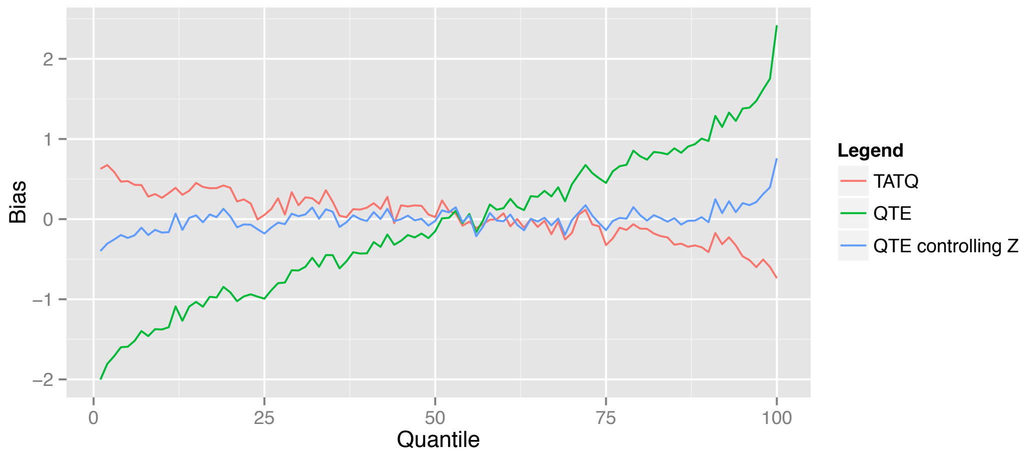


Figure 5: Average bias across 500 monte-carlo trials at each quantile when errors are positively correlated. The line labeled TATQ corresponds to the estimation procedure proposed in this paper. The line labeled QTE corresponds to quantile regression with a treatment dummy and no covariates. The last line corresponds to quantile regression on a treatment dummy and a set of covariates comprised of predetermined characteristics.

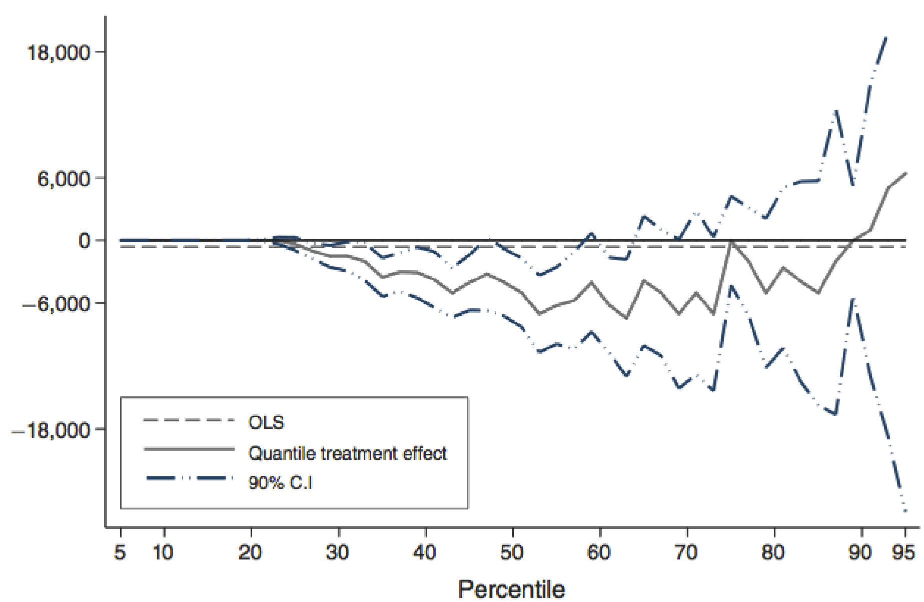


Figure 6: QTE estimates of the impact of micro finance on informal borrowing from Banerjee et al (2015).

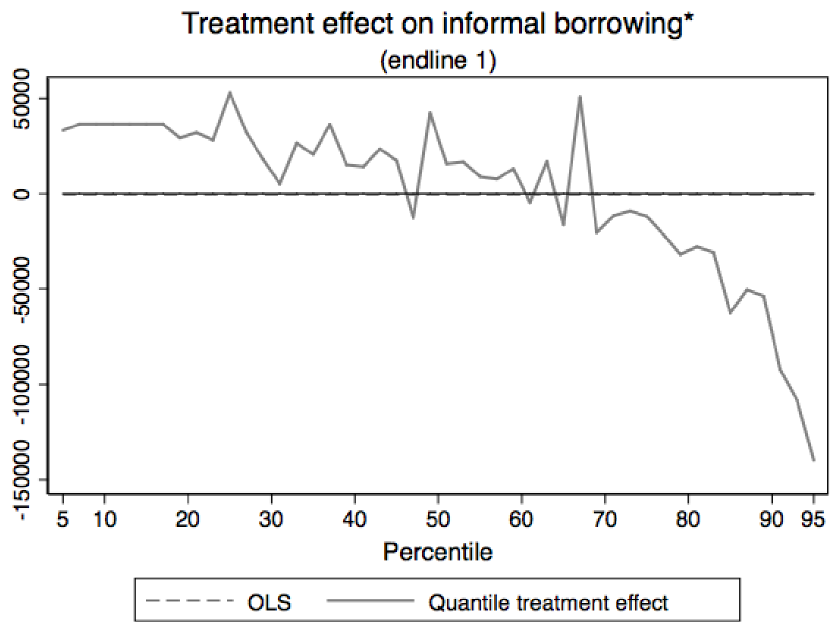


Figure 7: TATQ estimates of the impact of micro finance on informal borrowing.

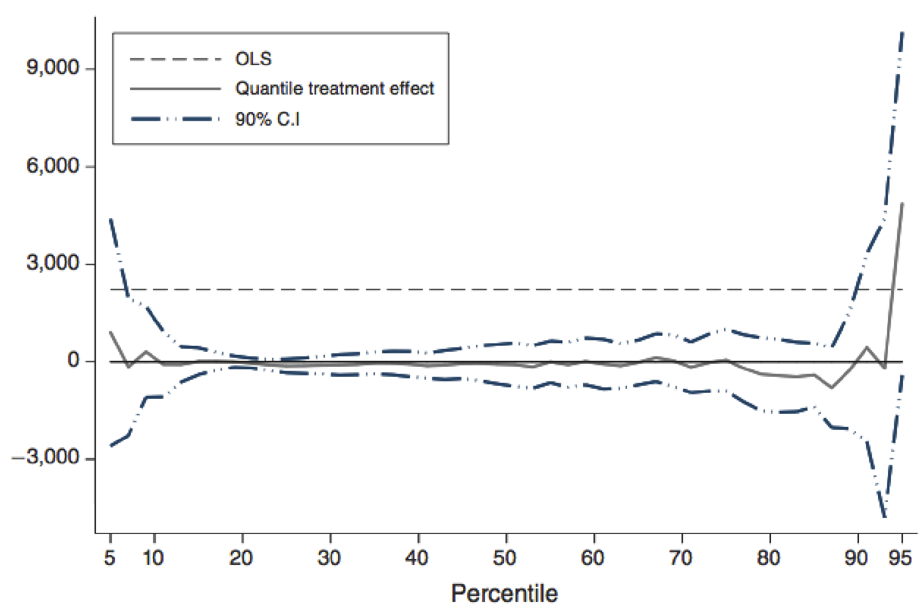


Figure 8: QTE estimates from Banerjee et al (2015) of the impact of micro finance on business profits among households who had a business prior to the introduction of micro credit.

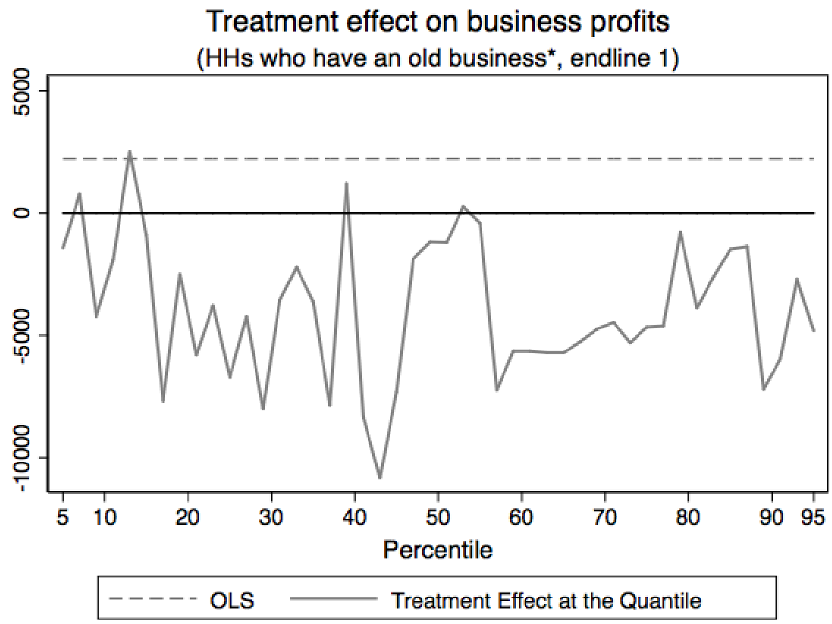


Figure 9: TATQ estimates of the impact of micro finance on business profits among households who had a business prior to the introduction of micro credit.

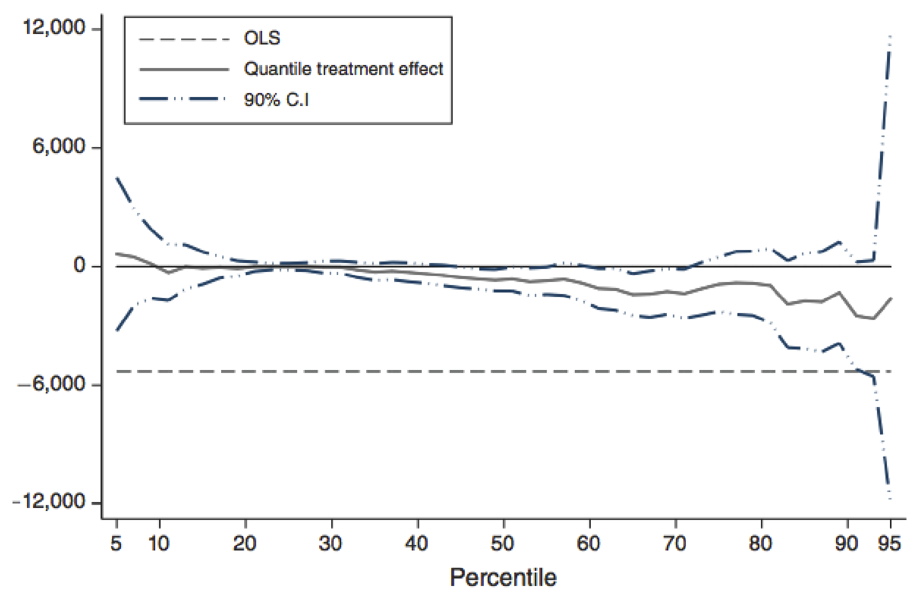


Figure 10: QTE estimates from Banerjee et al (2015) of the impact of micro finance on business profits among households who started a business after the introduction of micro credit.

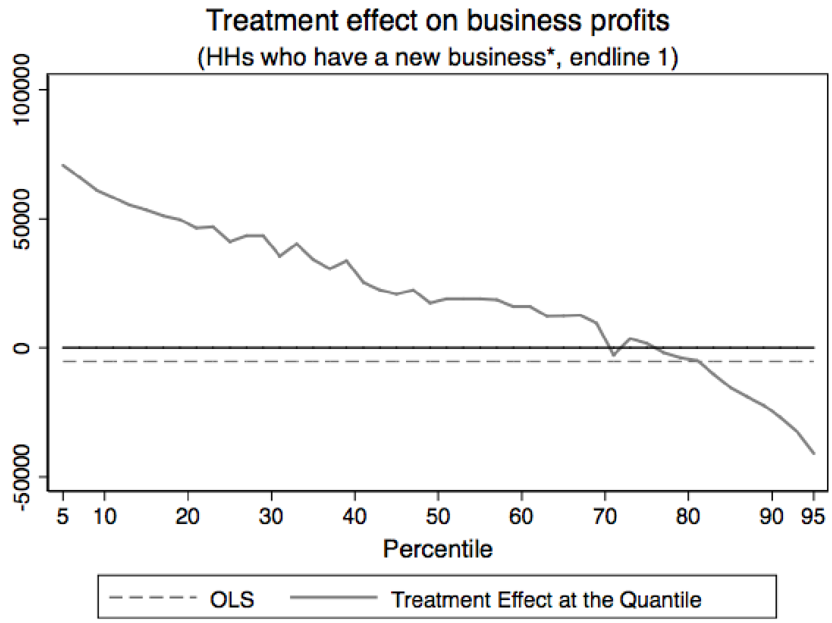


Figure 11: TATQ estimates of the impact of micro finance on business profits among households who started a business after the introduction of micro credit.

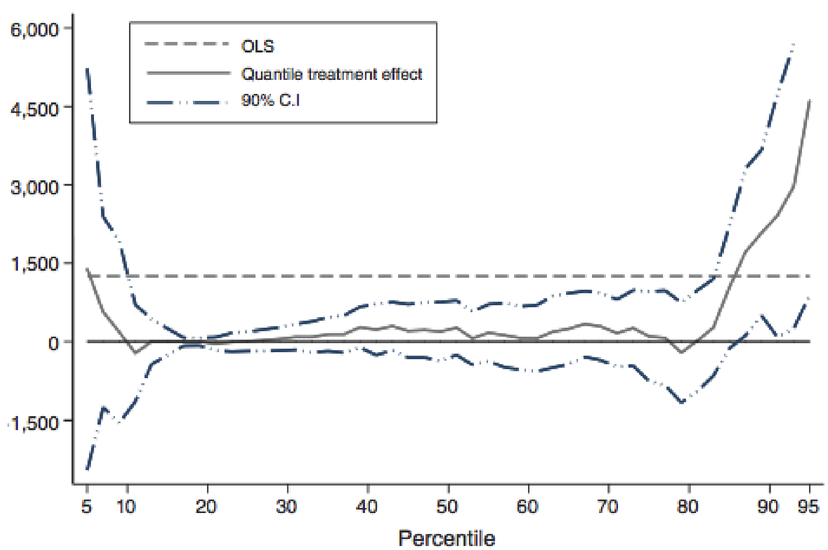


Figure 12: QTE estimates from Banerjee et al (2015) of the impact of micro finance on business profits among all households that owned a business two years after the introduction of micro credit.

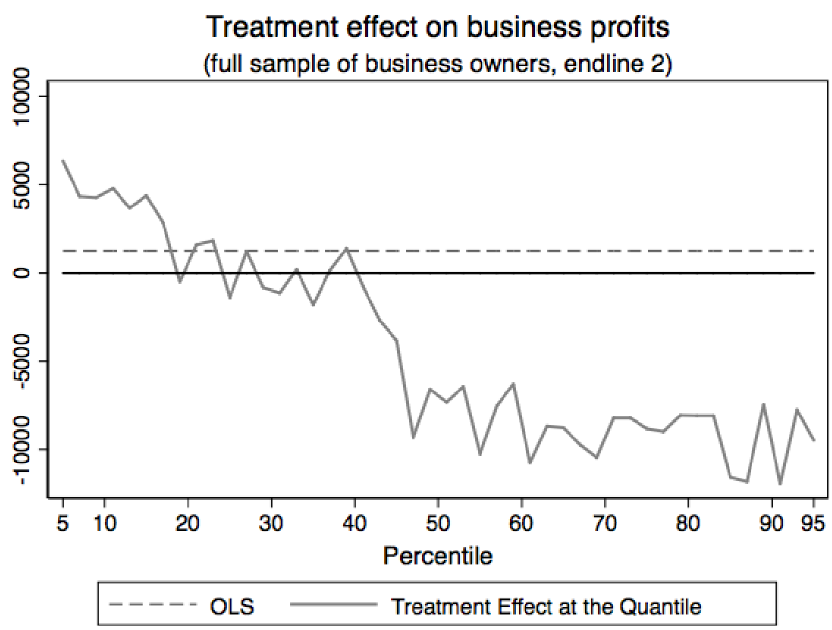


Figure 13: TATQ estimates of the impact of micro finance on business profits among all households that owned a business two years after the introduction of micro credit.

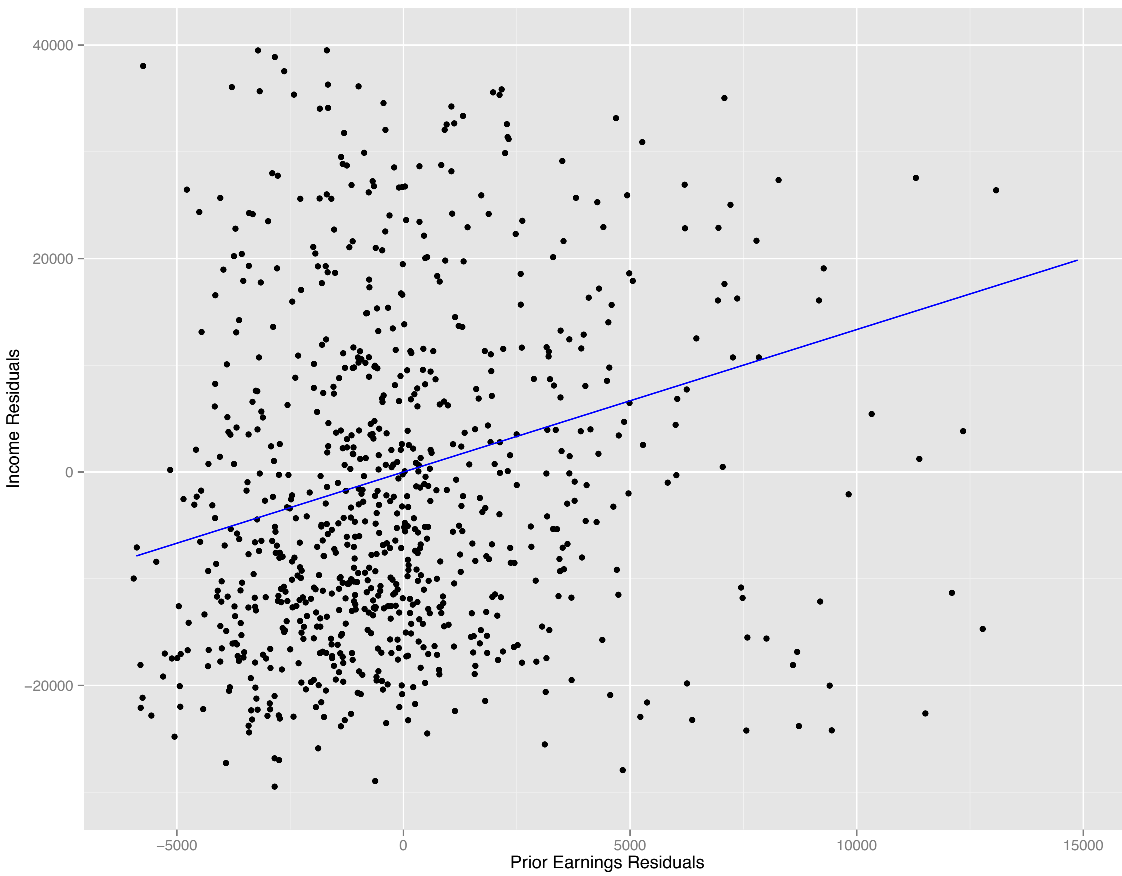


Figure 14: Residual variation in income plotted against residual variation in prior earnings for the control group. The line is the OLS estimate of the expected treated outcome for a given value of the prior earnings residual. This figure is meant to mirror figure 3 from the theoretical section in sample.

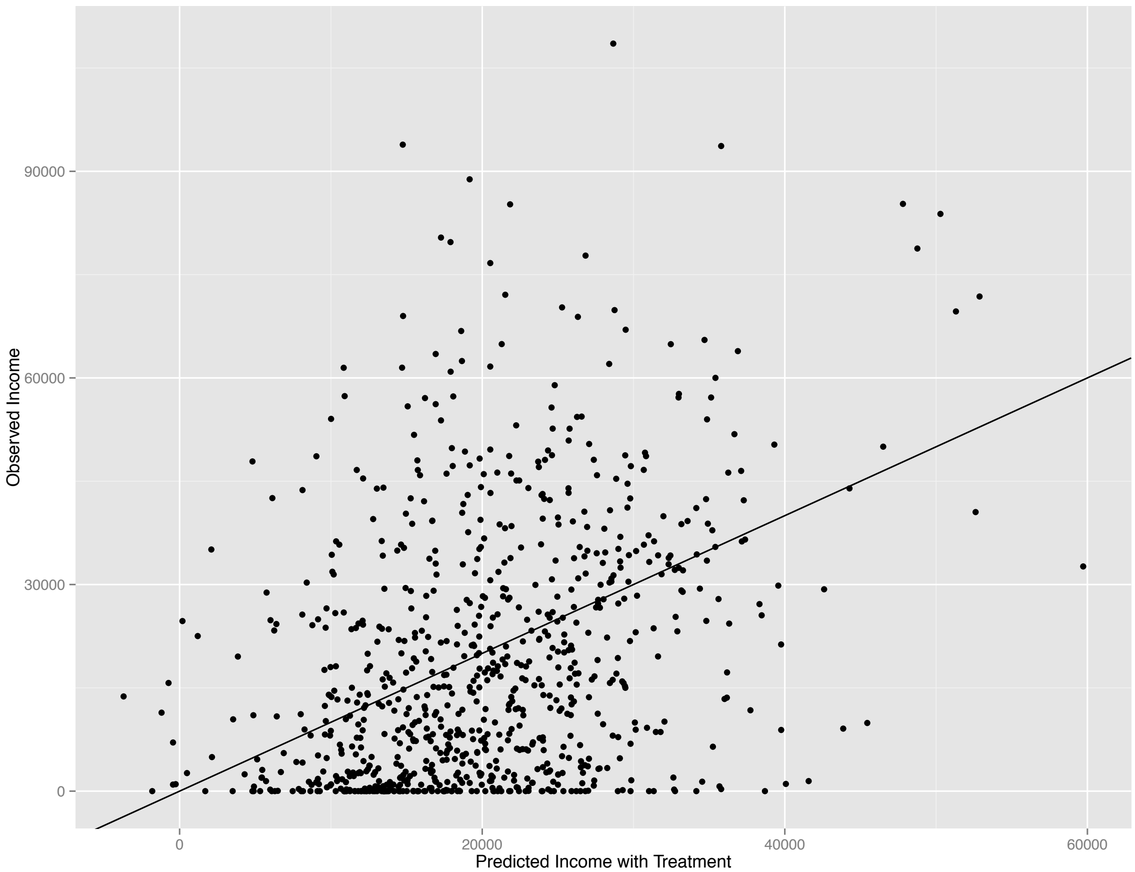


Figure 15: Observed income plotted against predicted values of the OLS estimate of the conditional expectation function $\mathbb{E}[Y_1|Z]$ for the control subsample. The 45 degree line provides the appropriate comparison value and the frequency of observations provide the appropriate weights within a window of a given quantile. This is meant to mirror figure 3 from the theoretical section in sample.

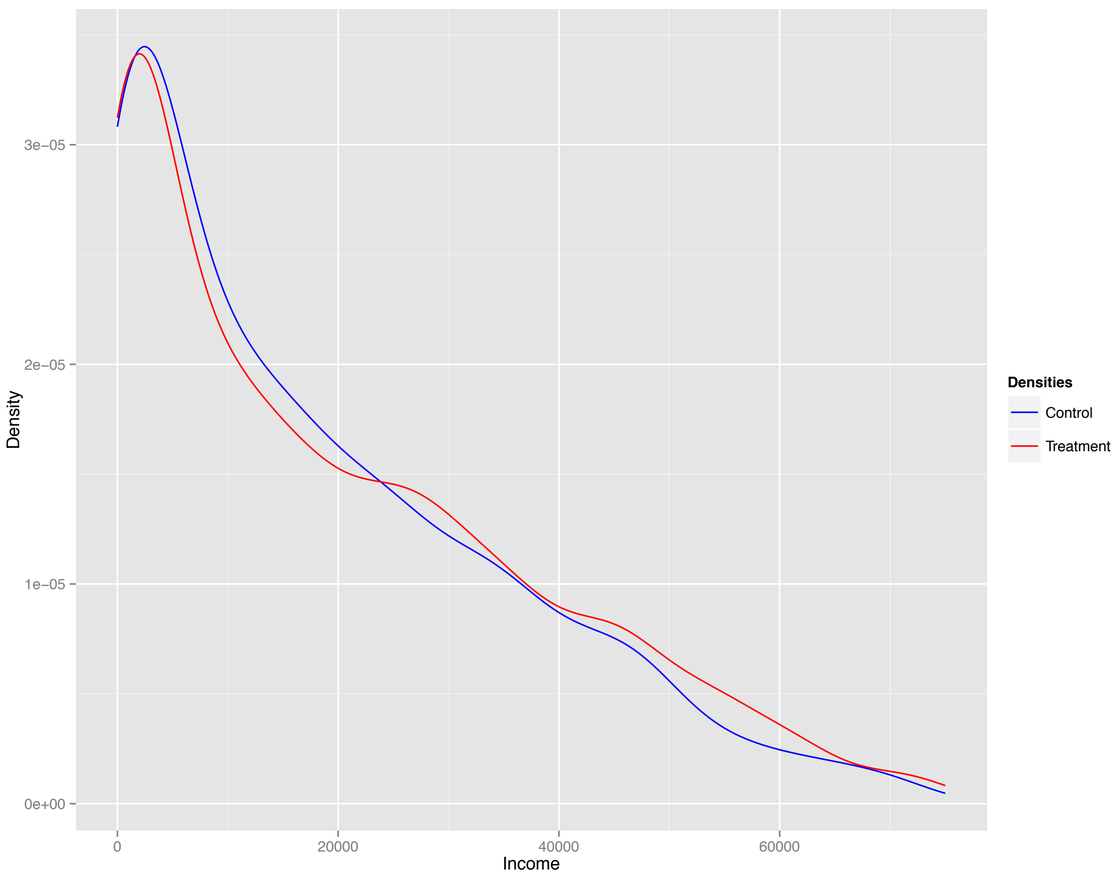


Figure 16: Kernel density estimates of the density of treatment and control outcomes.

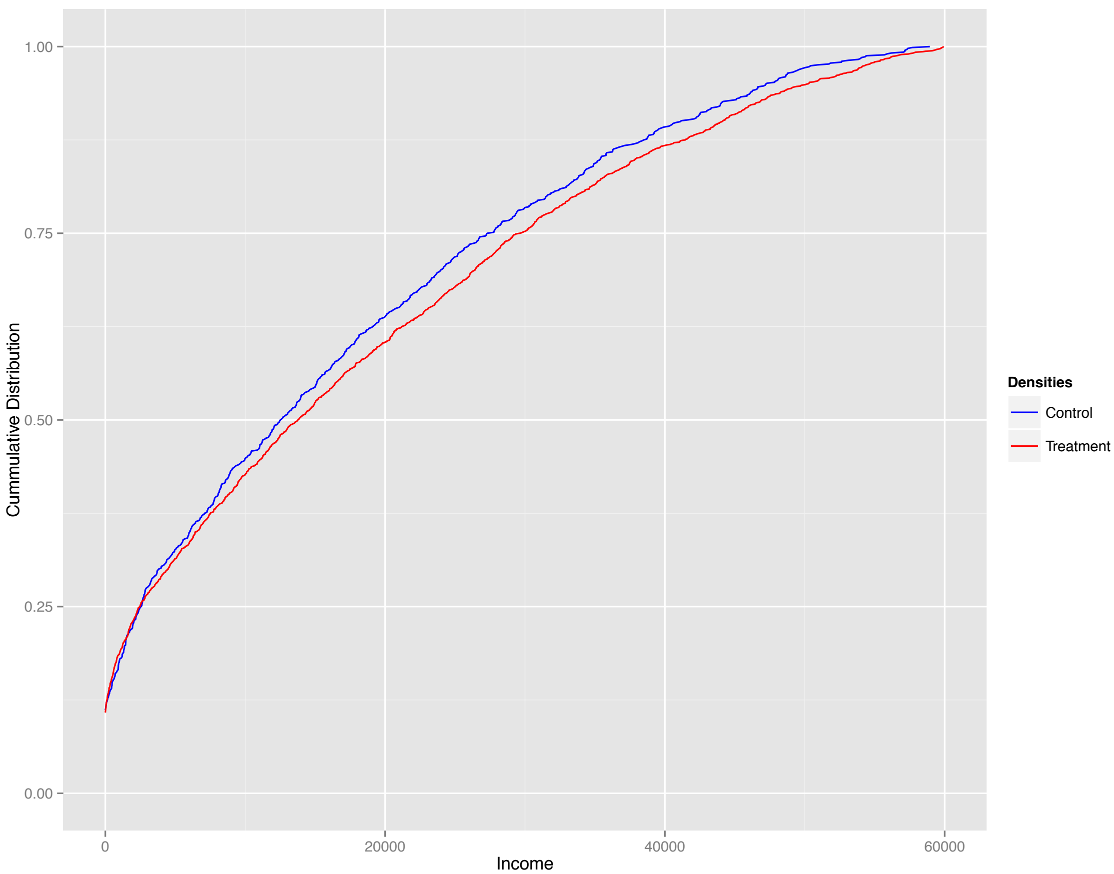


Figure 17: Estimates of the CDF of treatment and control outcomes.

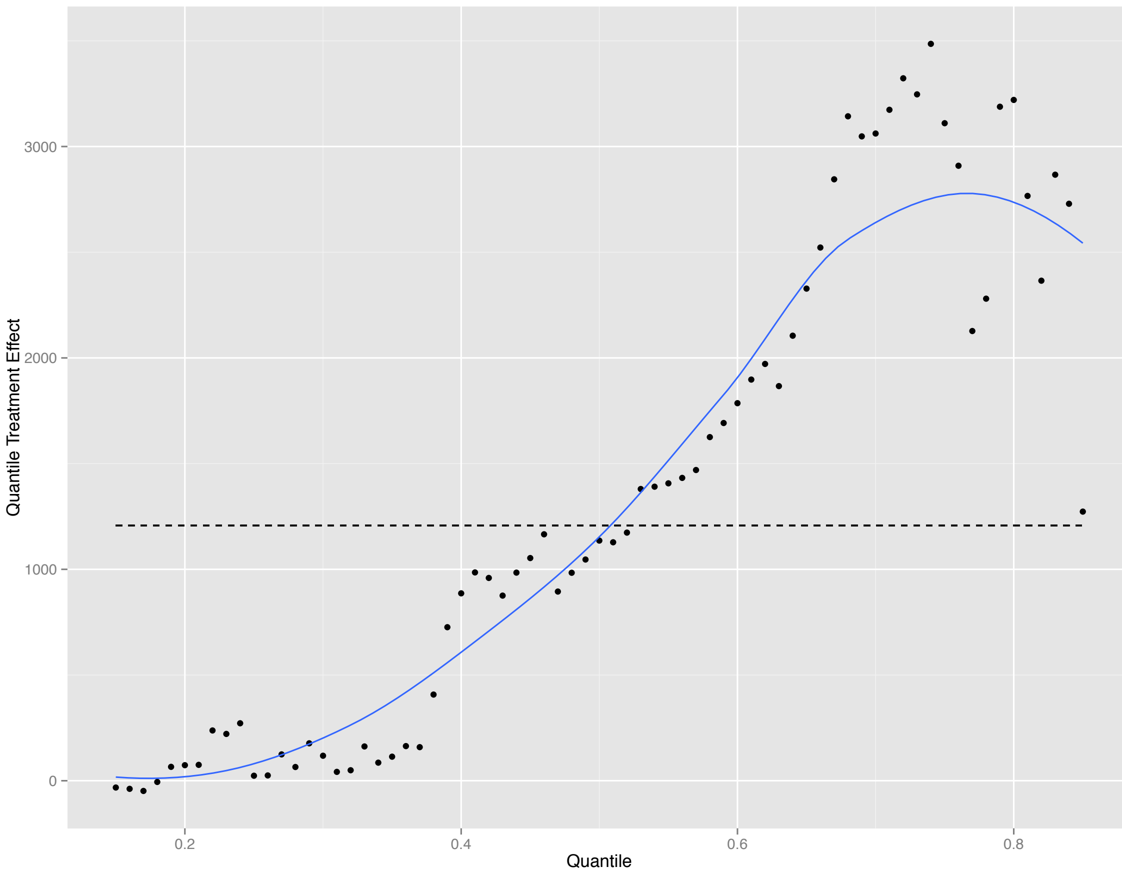


Figure 18: Quantile treatment effect estimates at each quantile controlling for predetermined characteristics. The dotted line is the OLS estimate of the average treatment effect.

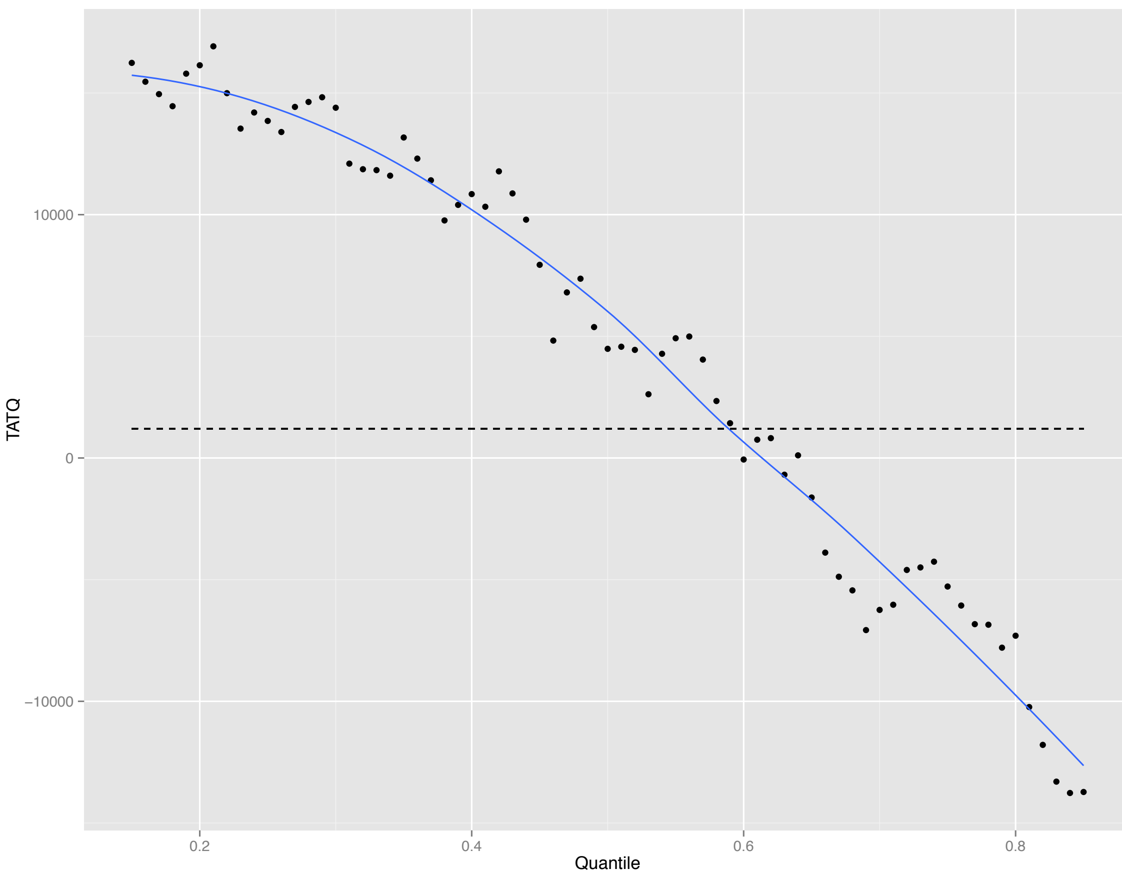


Figure 19: Estimates of the TATQ at each quantile using the estimation procedure outlined in this paper. The dotted line is the OLS estimate of the average treatment effect.