Microeconometrics I – Problem Set 03

Vinícius de Almeida Nery Ferreira

September 8th, 2024

1 Identification of the ATU with IPW

Define the Average Treatment Effect on the Untreated (ATU) as the parameter of interest:

$$\beta \coloneqq \mathbb{E}\left[Y(1) - Y(0) | D = 0\right] \tag{ATU}$$

where D is the binary treatment assignment variable and $Y(0)$ and $Y(1)$ denote potential outcomes. Let X be a vector of observed covariates which are not affected by treatment.

We claim that the ATU is point-identified by the IPW estimand

$$\frac{1}{1-\mathbb{E}[D]}\left\{\mathbb{E}\left[\frac{(1-P(X))DY}{P(X)}\right] - \mathbb{E}[(1-D)Y]\right\},$$

where $P(x) \coloneqq \mathbb{P}(D = 1|X = x) = \mathbb{E}[D|X = x]$ is the propensity score.

Proof. We will do the derivation term by term. For the first one, note that

$$\begin{aligned}
\frac{1}{1-\mathbb{E}[D]}\mathbb{E}\left[\frac{(1-P(X))DY}{P(X)}\right] &= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}_X\left\{\mathbb{E}\left[\frac{(1-P(X))DY}{P(X)}\middle| X\right]\right\} && (\text{LIE } |X) \\
&= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}_X\left\{\frac{1-P(X)}{P(X)}\mathbb{E}\left[DY|X\right]\right\} && (P(X) = f(x))
\end{aligned}$$

To continue, note that we can use the LIE and the definition of $P(x) \coloneqq \mathbb{P}(D = 1|X = x)$ to write

$$\begin{aligned}
\frac{\mathbb{E}\left[DY|X\right]}{P(X)} &= \frac{\mathbb{P}(D=0|X=x)\cdot\mathbb{E}\left[DY|X, D=0\right] + \mathbb{P}(D=1|X=x)\cdot\mathbb{E}\left[DY|X, D=1\right]}{\mathbb{P}(D=1|X=x)} && (\text{LIE}) \\
&= \mathbb{E}\left[Y|X, D=1\right] && (D = \{0,1\}) \\
&= \mathbb{E}\left[Y(1)|X, D=1\right] && (\text{Pot. Outcomes}) \\
&= \mathbb{E}\left[Y(1)|X\right] && (\text{Unconfoundness})
\end{aligned}$$

Note that we need the *Commom Support* assumption so that the first equality can actually be written. Recall that *Unconfoundness* (or conditional independence assumption (CIA)) states that

$$((Y(0), Y(1)) \perp D) \mid X \tag{Unconfoundness}$$

Using the result $\frac{\mathbb{E}[DY|X]}{P(X)} = \mathbb{E}\left[Y(1)|X\right]$, we can continue our derivation:

$$
\begin{aligned}
\frac{1}{1-\mathbb{E}[D]}\mathbb{E}\left[\frac{(1-P(X))DY}{P(X)}\right] &= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}_X\left\{(1-P(X))\mathbb{E}\left[Y(1)|X\right]\right\} && \text{(Previous Result)}\\
&= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}_X\left\{(1-\mathbb{E}[D|X])\mathbb{E}\left[Y(1)|X\right]\right\} && (P(X)=\mathbb{E}[D|X])\\
&= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}_X\left\{(\mathbb{E}[1-D|X])\mathbb{E}\left[Y(1)|X\right]\right\} && (\text{Linearity of } \mathbb{E})\\
&= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}_X\left\{\mathbb{E}\left[(1-D)Y(1)|X\right]\right\} && \text{(Unconfoundness)}\\
&= \frac{1}{1-\mathbb{E}[D]}\mathbb{E}\left[(1-D)Y(1)\right] && (\text{LIE } |X)\\
&= \frac{\mathbb{P}(D=0)\cdot\mathbb{E}\left[(1-D)Y(1)|D=0\right]}{1-\mathbb{E}[D]} && (\text{LIE } |D) \text{ and } ((1-D)|_{D=1}=0)\\
&= \frac{(1-\mathbb{P}(D=1))\cdot\mathbb{E}\left[Y(1)|D=0\right]}{1-\mathbb{E}[D]} && (D\in\{0,1\})\\
&= \frac{(1-\mathbb{E}[D])\cdot\mathbb{E}\left[Y(1)|D=0\right]}{1-\mathbb{E}[D]} && (\mathbb{P}(D=1)=\mathbb{E}[D])\\
&= \mathbb{E}\left[Y(1)|D=0\right]
\end{aligned}
$$

Now, for the second term,

$$
\begin{aligned}
\frac{\mathbb{E}[(1-D)Y]}{1-\mathbb{E}[D]} &= \frac{\mathbb{P}(D=0)\cdot\mathbb{E}[(1-D)Y|D=0]+\mathbb{P}(D=1)\cdot\mathbb{E}[(1-D)Y|D=1]}{1-\mathbb{E}[D]} && (\text{LIE } |D)\\
&= \frac{(1-\mathbb{P}(D=1))\cdot\mathbb{E}[Y|D=0]}{\mathbb{E}[1-D]} && (D\in\{0,1\})\\
&= \frac{(1-\mathbb{P}(D=1))\cdot\mathbb{E}[Y(0)|D=0]}{1-\mathbb{E}[D]} && \text{(Pot. Outcomes)}\\
&= \frac{(1-\mathbb{E}(D))\cdot\mathbb{E}[Y(0)|D=0]}{1-\mathbb{E}[D]} && (D \text{ binary})\\
&= \mathbb{E}[Y(0)|D=0]
\end{aligned}
$$

Combining both terms and using the linearity of the expected value operator,

$$
\begin{aligned}
\frac{1}{1-\mathbb{E}[D]}\left\{\mathbb{E}\left[\frac{(1-P(X))DY}{P(X)}\right]-\mathbb{E}[Y(1-D)]\right\} &= \mathbb{E}[Y(1)|D=0]-\mathbb{E}[Y(0)|D=0]\\
&= \mathbb{E}[Y(1)-Y(0)|D=0],
\end{aligned}
$$

and so the proposed estimand point-identifies the ATU. □

2 Monte Carlo Simulations of Different ATE Estimators under Different Data Generating Processes (DGPs)

Define the following random variables, which are common across all DGPs:

- $U \sim \text{Uniform}[0,1]$;
- $\varepsilon \sim \text{Uniform}[0,1]$;
- $X = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \right)$;
- $X_1 = \Phi\left(\tilde{X}_1\right)$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution;
- $X_2 = \Phi\left(\tilde{X}_2\right)$;
- $Y = D \cdot Y(1) + (1-D) \cdot Y(0)$.

What changes across DGPs is the assignment mechanism, D, and the potential outcomes $\{Y(0), Y(1)\}$. Denoting DGPs by superscripts, we have that

$$D^1 = \mathbb{1}\left\{X_1 + X_1^2 \geq 2 \cdot U\right\} \quad D^2 = \mathbb{1}\left\{X_1 + X_2 \geq 2 \cdot U\right\} \quad D^3 = \mathbb{1}\left\{X_1 + X_1^2 \geq 2 \cdot U\right\}$$
$$Y(0)^1 = 1 + X_1 + \varepsilon \quad Y(0)^2 = 1 + X_1 + X_2 + \varepsilon \quad Y(0)^3 = 1 + X_1 + X_2 + \varepsilon$$
$$Y(1)^1 = Y(0)^1 + 2 + X_1 \quad Y(1)^2 = Y(0)^2 + 2 + X_1 + X_2 \quad Y(1)^3 = Y(0)^3 + 2 + X_1 + X_2$$

For each DGP, we observe a sample $\{D, X_1, X_2, Y\}_{i=1}^N$, where $N = 10,000$. For each of the three samples, we generate new draws from X, U and ε to construct the appropriate variables.

Note that the true ATE is 2.5 in DGP 1 and 3 in DGPs 2 and 3, since $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0.5$. We now discuss the consistency of each estimator depending on the DGP.

CEF Estimator

Our CEF Estimator for the ATE does not use X_2 as a covariate in the regressions. That is, we estimate the regression

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_{1i} + \gamma_3 D_i \left(X_{1i} - \bar{X}_1\right),$$

where the OLS estimator $\hat{\gamma}_1$ will be our estimator for the ATE.

Since the estimator does not include X_2 in the conditional expectation function of Y, it will be inconsistent for the ATE in the DGPs where X_2 affects potential outcomes and when it is correlated to treatment, as then we have omitted variable bias. As a consequence, we expect that it be inconsistent for the ATE only in DGP 2.

Inverse Probability Weighting Estimator

Following Imbens and Wooldridge (2009, p.34-35), we estimate the propensity score using a logit link function involving flexible functions of the covariates X_1 and X_2. In particular, we estimate

$$P_i(X_1, X_2) := \mathbb{E}[D_i \mid X_1, X_2] = \lambda\left(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_1^2 + \alpha_3 X_2 + \alpha_4 X_2^2\right),$$

where $\lambda(\cdot)$ is the logit link function. We then estimate the ATE using the expression

$$\frac{\sum_{i=1}^N \frac{D_i \cdot Y_i}{P_i(X_1,X_2)}}{\sum_{i=1}^N \frac{D_i}{P_i(X_1,X_2)}} - \frac{\sum_{i=1}^N \frac{(1-D_i)\cdot Y_i}{(1-P_i(X_1,X_2))}}{\sum_{i=1}^N \frac{(1-D_i)}{(1-P_i(X_1,X_2))}}$$

Since the propensity score estimator is correctly specified in all DGPs, we expect that the IPW be consistent in all of them.

Doubly Robust Estimator

As the name suggests, the DR estimator is robust to misspecification in one of two dimensions: either on the CEF estimation or on the propensity score estimation (but not both).

In the first phase, we use the same estimated propensity score used to construct the IPW estimator. We then run two weighted regressions using only X_1 as a covariate:

$$\text{Control Sample:} \; \min_{\alpha_0,\beta_0} \sum_{i:D_i=0} \frac{\left(Y_i - \alpha_0 - (X_{1i} - \bar{X}_1)\beta_0\right)^2}{1 - P_i(X_1, X_2)}$$

$$\text{Treated Sample:} \; \min_{\alpha_1,\beta_1} \sum_{i:D_i=1} \frac{\left(Y_i - \alpha_1 - (X_{1i} - \bar{X}_1)\beta_1\right)^2}{P_i(X_1, X_2)}$$

Our estimator for the ATE is then given by $\hat{\alpha_1} - \hat{\alpha_0}$, as we are centering X_1 around its mean.

Since our propensity score specifications include flexible functions of both X_1 and X_2, we expect that the DR be consistent in all DGPs, even though the CEF part may be misspecified.

Naive Difference in Means

We can view this estimator as a bivariate OLS of Y on D. Since X_1 always affects treatment assignment and potential outcomes, we expect that this estimator be inconsistent in all DGPs.

Simulation Results

The results of our simulations are shown in Figures 1 and 2. The second figure does not include the Naive estimators, allowing for a better visualization of the scale of the biases of the other estimators.

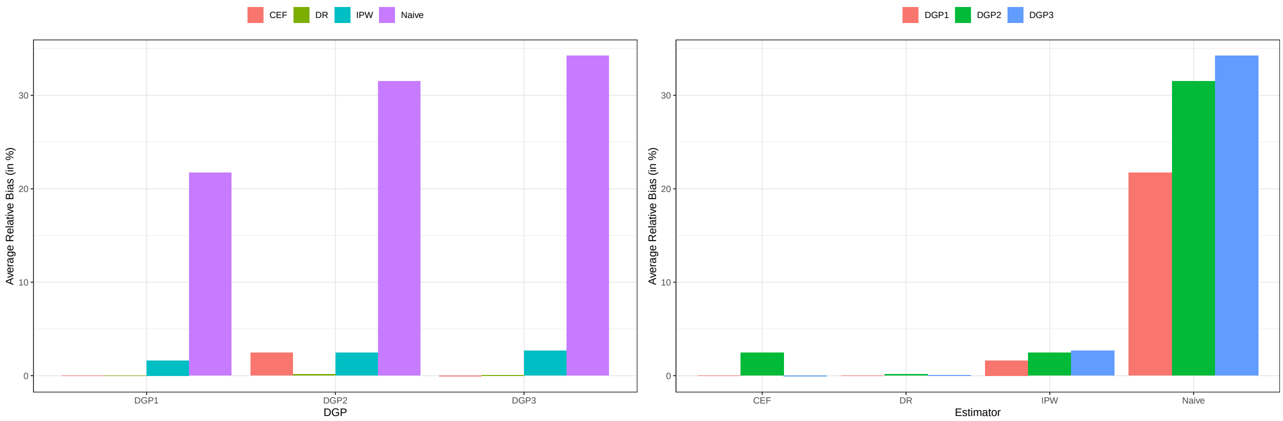


Figure 1: Monte Carlo Simulation Results

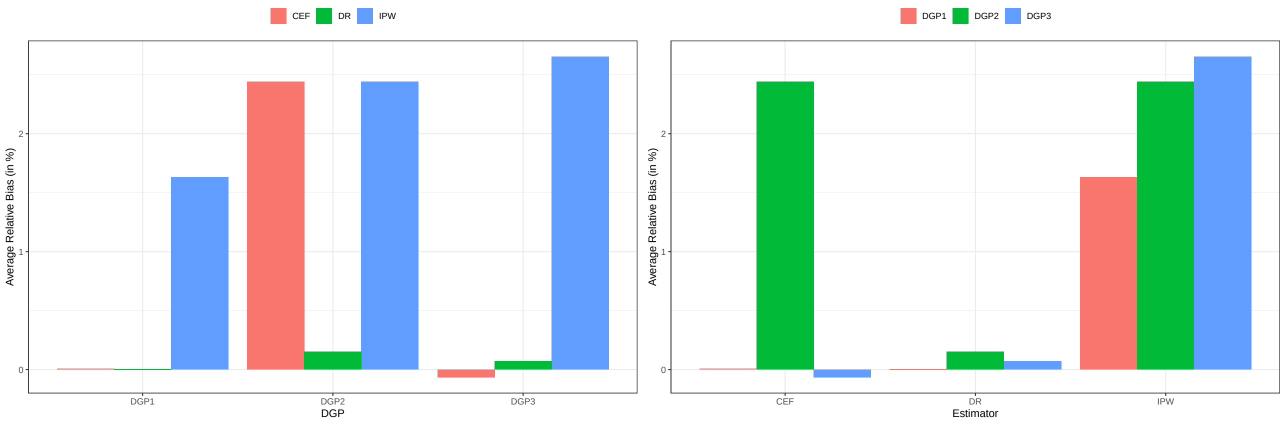


Figure 2: Monte Carlo Simulation Results – Without Naive

The results are as expected. The Naive estimator is always biased, and its bias is always larger than the ones in the other estimators (which partially or totally control for selection on X_1 and X_2).

The CEF estimator is inconsistent in the second DGP, as the omitted variable X_2 is related to both potential outcomes and treatment assignment. We note that this bias is relatively small (over 2%), but large when compared to the ones obtained in DGPs 1 and 3.

The IPW is also always consistent, but, due to its low efficiency, has average biases that are not negligible, ranging around 1.5 to 2.5%. Finally, the DR estimator is consistent in all specifications, as expected.

We also plot the variance of the estimators in Figure 3, where we see the higher variance of the IPW estimator relative to the regression-based estimators (CEF and DR) and even the Naive estimator in DGPs 1 and 3.

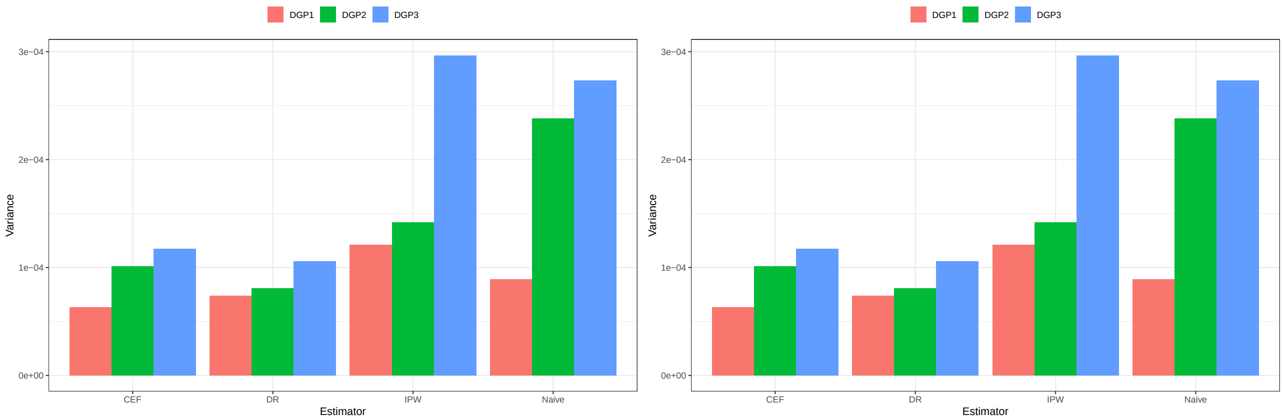


Figure 3: Monte Carlo Simulation Results – Estimator Variance

3 Quantile Regressions

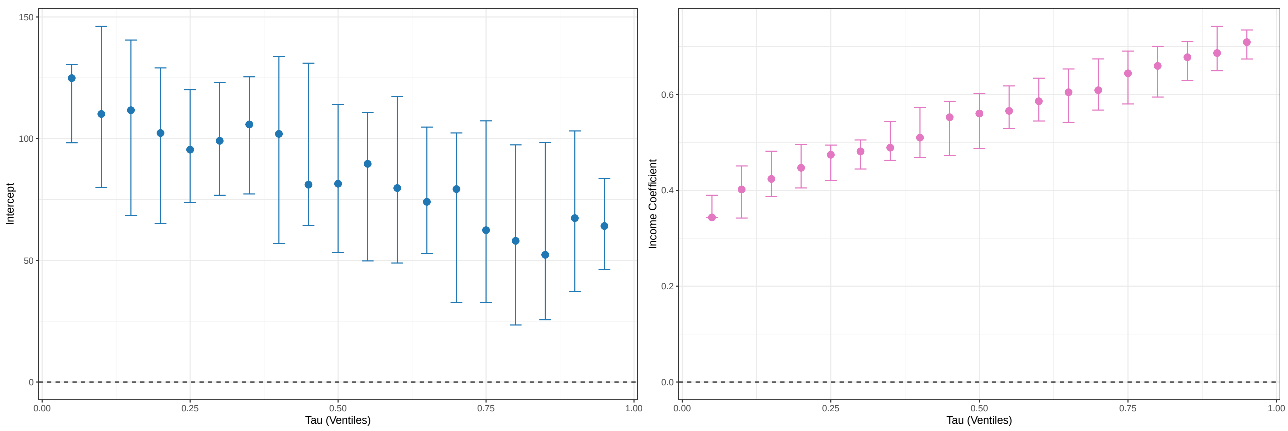
Using the `engel` dataset, we estimated a quantile regression of food expenditure on income for the median ($\tau = 0.5$) of the conditional distribution. Results are shown in Table 1.

Table 1: Quantile Regression – Food Expenditure on Income – Coefficients at the Median

	Coefficient	Lower Bound	Upper Bounnd
Intercept	81.48	53.26	114.01
Income	0.560	0.487	0.602

We note that we should not interpret these results as pertaining to the *unconditional* distribution of food expenditure, but rather to the *conditional* distribution on income. Thus, on the median of this distribution, a unit increase in income leads to a 0.56 increase on food expenditure. Furthermore, the intercept for households on the median of the conditional distribution indicates that the "subsistence" expenditure is about 81.5, as measured by the point estimate.

We plot the estimated coefficients for all ventiles ($\tau \in \{0.05, 0.10, 0.15, \cdots, 0.90, 0.95\}$) of the distribution of food expenditure conditional on income in Figure 4, along with the corresponding lower and upper bounds of the coefficients. We see that, in this simple analysis, the effect of income on food expenditure increases along the conditional distribution, while the intercept – which may be viewed as "basic" food expenditure – decreases..



(a) Intercept

(b) Income Coefficient

Figure 4: Quantile Regression Results for All Ventiles

Following Koenker and Hallock (2001), we also plot the conditional quantile curves for $\tau \in \{0.05, 0.10, 0.15, \cdots, 0.90, 0.95\}$. The median is the solid black curve, while the dashed black line is the conditional expectation function estimated using OLS to minimize the sum of the squared residuals.

We see that the dispersion of the Engel Curves increases with income, as measured by the increasing distance between the lines when we move along the x-axis. Additionally, the mean and median functions are quite different, which may be explained by the outliers seen in the graph, which affect the conditional expectation function, but not the median. As discussed in their paper, this causes the least squares fit to provide a rather poor approximation of the conditional mean of the poorest households in the sample.

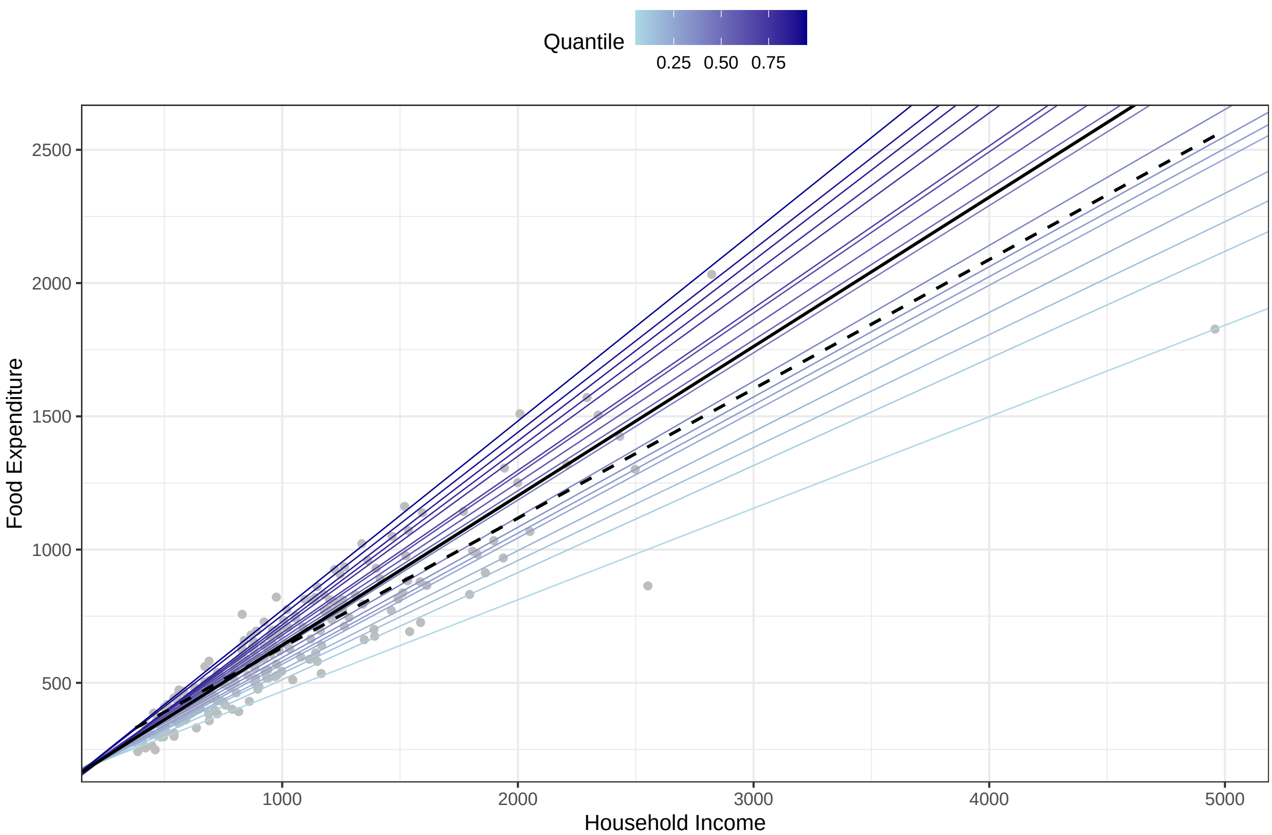


Figure 5: Conditional Quantile Functions