Swiss Institute for International Economics
and Applied Economic Research

University of St.Gallen

Estimation of counterfactual distributions using quantile regression

3rd Conference on Policy Evaluation
ZEW, Mannheim, 27.10.2006

Blaise Melly
www.siaw.unisg.ch/lechner/melly

Motivation

- Most of the parametric and nonparametric literature assuming exogeneity has focused on the estimation of average treatment effects
- Parametric setting: Oaxaca (1973) / Blinder (1973) decomposition
- Nonparametric setting: matching literature surveyed by Imbens (2004)

Motivation

- However, there is an interest in knowing the distributional effects of minimum wages, training programs, education, gender,…
- The increase in wage inequality has recently motivated new estimators. Machado and Mata (2005) and Gosling, Machin and Meghir (2000) have proposed estimators based on parametric quantile regression but they do not give asymptotic justification and analytical inference procedures

Contributions

- In this paper we propose and derive the asymptotic distribution of:
  - a quantile equivalent of the Oaxaca (1973) / Blinder (1973) decomposition
  - a quantile equivalent of the Heckman, Ichimura and Todd (1998) matching estimator
- Estimators of the variances, Monte Carlo simulations and application to the black-white wage gap
- Illustration for this talk: effects of training programs in West Germany

Structure of the talk

1. **Identification of quantile treatment effects**
2. Estimation of quantile treatment effects
   - Parametric estimator
   - Nonparametric estimator
3. Public sector sponsored training in West Germany
4. Conclusions

Notation

- Notation

  Binary treatment: T

  Potential outcomes: $Y(0), Y(1)$

  Covariates: X

- Distribution and quantile functions of Y

  $F_{Y(t)}(q) = \theta \Leftrightarrow F_{Y(t)}^{-1}(\theta) = q$

- Propensity score

  $\Pr(T = 1 | X) = p(X)$

- Observed: $T, X,\ Y = TY(0) + (1 - T)Y(1)$

Identification of *ATE*

- In this paper, we assume selection on observables:
  $Y(0), Y(1) \amalg T | X$
  and common support $0 < p(X) < 1$.
- These assumptions identify *ATE*s. For instance:

$$E\left[Y(0)\right] = E\left[E\left[Y(0)|X\right]\right]$$

$$= E\left[E\left[Y(0)|X, T=0\right]\right]$$

$$= E\left[E\left[Y|X, T=0\right]\right]$$

- Alternatives: instrumental variables, DiD, bounds

Estimands: Quantile Treatment Effects

Quantile treatment effect: $F_{Y(1)}^{-1}(\theta) - F_{Y(0)}^{-1}(\theta)$

Quantile treatment effect on the treated:

$$F_{Y(1)}^{-1}(\theta | T = 1) - F_{Y(0)}^{-1}(\theta | T = 1)$$

- In order to identify *QTE*, we cannot use the strategy used for *ATE* because the unconditional quantiles cannot be represented as linear functions of the conditional quantiles: $F_{Y(0)}^{-1}(\theta) \neq E\left[F_{Y(0)}^{-1}(\theta | X)\right]$

Identification of *QTE*

- However, this property holds for the distribution function. The exogeneity assumption implies

$$F_{Y(0)}(y) = E\left[1\left(Y(0) \le y\right)\right] = E\left[E\left[1\left(Y(0) \le y\right) \middle| X\right]\right]$$

$$= E\left[E\left[1\left(Y(0) \le y\right) \middle| X, T = 0\right]\right] = E\left[E\left[1\left(Y \le y\right) \middle| X, T = 0\right]\right]$$

- Similar situation to *ATE*:

- exogeneity and a parametric assumption for the conditional distribution function identify the *QTE*s
- exogeneity and the common support assumption identify the *QTE*s without parametric restrictions

Propensity score

- A well-known result of Rosenbaum and Rubin (1983): $Y(0),Y(1) \amalg T \mid X \Rightarrow Y(0),Y(1) \amalg T \mid p(X)$
  Direct extension to the estimation of QTE
- Similarly: direct extension for multiple treatments
- All strategies used to estimate *ATE* can be used to estimate *QTE*:
  - Regression estimators
  - Matching estimators
  - Propensity score estimators (weighting, regression, blocking,…)

Structure of the talk

1. Identification of quantile treatment effects
2. **Estimation of quantile treatment effects**
   - **Parametric estimator**
   - **Nonparametric estimator**
3. Public sector sponsored training in West Germany
4. Conclusions

Parametric estimation procedure

- We assume linearity of all conditional quantiles (similar results for other parametric assumptions)

  Procedure:

1. We estimate the conditional quantiles using quantile regression (Koenker and Bassett 1978) for all $\tau \in [0,1]$

   $$\hat{\beta}_t(\tau) = \arg\min_{b \in \mathbb{R}^K} n_0^{-1} \sum_{i:T_i=t} \rho_\tau(Y_i - X_i b), \; t = 0,1$$

   and get J quantile regression coefficient vectors

   $$\hat{\beta}_t = \left(\hat{\beta}_t(\tau_1), \ldots, \hat{\beta}_t(\tau_J)\right)$$

Parametric estimation procedure

2. Problem: the estimated conditional quantiles can cross: $\tau \leq \tilde{\tau} \not\Rightarrow X_i \hat{\beta}(\tau) \leq X_i \hat{\beta}(\tilde{\tau})$. It is therefore not possible to invert the estimated conditional quantile function. But, by definition,

$$F_{Y|X}\left(q|X_i\right) = \int_0^1 1\left(F_{Y|X}^{-1}\left(\tau|X_i\right) \leq q\right) d\tau = \int_0^1 1\left(X_i'\beta\left(\tau\right) \leq q\right) d\tau$$

The sample equivalent is given by

$$\hat{F}_{Y(t)}\left(q|X_i\right) = \int_0^1 1\left(X_i \hat{\beta}_t\left(\tau\right) \leq q\right) d\tau$$

$$= \sum_{j=1}^{J} \left(\tau_j - \tau_{j-1}\right) 1\left(X_i' \hat{\beta}_t\left(\tau_j\right) \leq q\right)$$

Parametric estimation procedure

3. Integration of the estimated conditional distribution function over the range of X

$$\hat{F}_{Y(t)}\left(q \middle| T = t'\right) = \int \hat{F}_{Y(t)}\left(q \middle| x\right) dF_X\left(x \middle| T = t'\right) = n_{t'}^{-1} \sum_{i:T_i = t'} \hat{F}_{Y(t)}\left(q \middle| X_i\right)$$

4. Inversion of the estimated unconditional distribution function

$$\hat{q}_c(\theta) = \inf\left\{q : \hat{F}_{Y(0)}\left(q \middle| T = 1\right) \geq \theta\right\}$$

$$\hat{q}_0(\theta) = \inf\left\{q : \hat{F}_{Y(0)}\left(q \middle| T = 0\right) \geq \theta\right\}$$

$$\hat{q}_1(\theta) = \inf\left\{q : \hat{F}_{Y(1)}\left(q \middle| T = 1\right) \geq \theta\right\}$$

Asymptotic distribution

- The estimators of *QTE* and *QTET* are root *n* consistent and asymptotically normally distributed
- In the paper: asymptotic variance, analytic estimator of the asymptotic variance, Monte Carlo simulations
- Remarks:
  - If we estimate observed quantiles, the proposed estimators are more precise than the sample quantiles
  - This suggest an "Hausman" test in order to test the validity of the parametric model

Nonparametric estimator: procedure

- No parametric assumption, smoothness of the conditional distribution, common support restriction, undersmoothing, higher-order kernel
- The procedure is basically the same but the conditional quantile functions are estimated nonparametrically (Chaudhuri 1991). Thus, instead of estimating global parameters we estimate different parameters for each combination of *X:*

$$\hat{\beta}_t(\tau, x) = \arg\min_b n_t^{-1} \sum_{i:T_i=t} K\left(\frac{X_i - x}{h_n}\right) \rho_\tau(Y_i - X_i b)$$

Nonparametric estimator: Asymptotic distribution

- The estimators of *QTE* and *QTET* are root *n* consistent and asymptotically normally distributed. In the paper: asymptotic variance, analytic estimator of the asymptotic variance, Monte Carlo simulations
- Corollary: the estimators of *QTE* and *QTET* achieve the semiparametric efficiency bounds derived by Firpo (forthcoming)
- Corollary: asymptotic equivalence of his propensity score weighting estimator and the local linear based matching estimator

Structure of the talk

1. Identification of quantile treatment effects
2. Estimation of quantile treatment effects
   - Parametric estimator
   - Nonparametric estimator
3. **Public sector sponsored training in West Germany**
4. Conclusions

Public sponsored training in West Germany

- Active labor market policies (ALMP) are considered as important tools to reduce Europe's notoriously high levels of unemployment. Germany, for instance, spends about 1% of GDP for ALMP

- We consider 4 groups of programmes

Program	Description
Practice firm	Further training that simulates a job in a specific field of profession
Short training	Further training to obtain additional qualification in the profession held; planned duration $\leq$ 6 months.
Long training	Same types as short training with a planned duration > 6 months.
Retraining	Training to obtain a new professional degree in a field other than the profession currently held.

Data

Same samples as in Lechner, Miquel and Wunsch (2005). Same data sets as in Fitzenberger and Speckesser (2005) and Fitzenberger, Osikominu and Völter (2006)

	Employment subsample	**Benefit payment register**	**Training participant data**
Source	Employer supplied mandatory social insurance entries.	Benefit payment register of the FEA.	Questionnaires filled in by caseworker for statistical purposes (*ST35*).
Population	1% random sample of persons covered by social insurance for at least one day 1975-1997. Self-employed, civil servants, students are not included. Data 1975-2002	Recipients of UA, UB, or MA 1975-2002.	Participants in further training, retraining, short programmes (§41a EPA), German language courses and temporary wage subsidies 1975-1997.
Available information	Personal characteristics and history of employment.	Information about the receipt of benefits, mainly *UB, UA, MA*.	Personal characteristics of participants and information about training programmes.
Important variables	Gender, age, nationality, education, profession, employment status, industrial sector, firm size, earnings, regional information	Type and amount of benefits received.	Type, duration and result of the programme, type of income support paid during participation.
Structure	Spells based on daily information.	Spells based on daily info.	Spells based on monthly information.

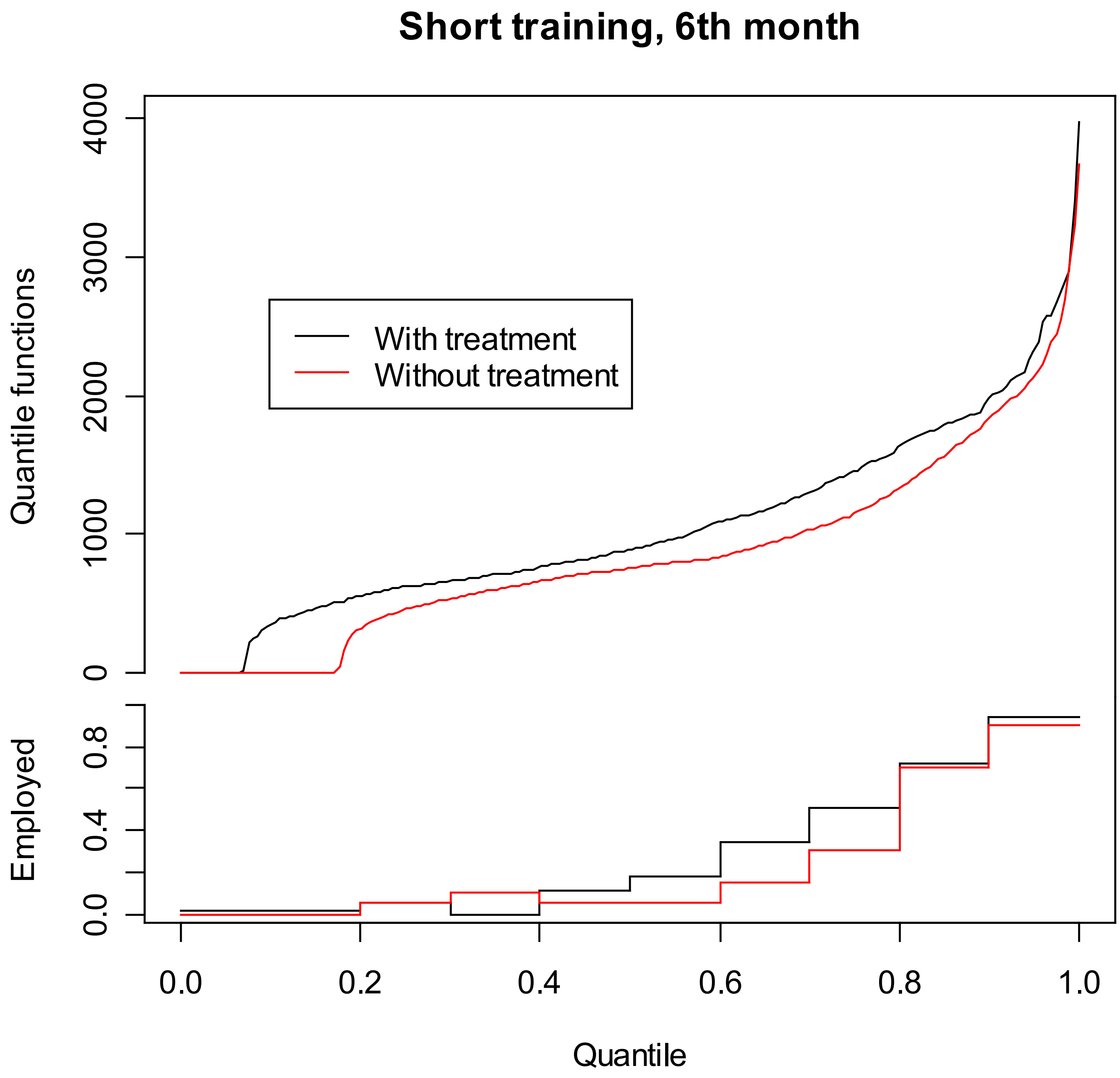
Outcome variables

- Employment status: *S*, Wage: *Y,* UE benefits: *B*
  Outcome: (consumptive) income = $SY+(1-S)B$
- We assume selection on observables for *S, B* and *Y*. Selection on observables is plausible given the rich datasets: employment history, detailed personal, regional and employer information, remaining UE benefits claim. Missing: health, motivation
- Alternative outcome: productive income = SY
- Ideally: effect on the potential wage *Y* but it is not identified without further (unreasonable) assumptions

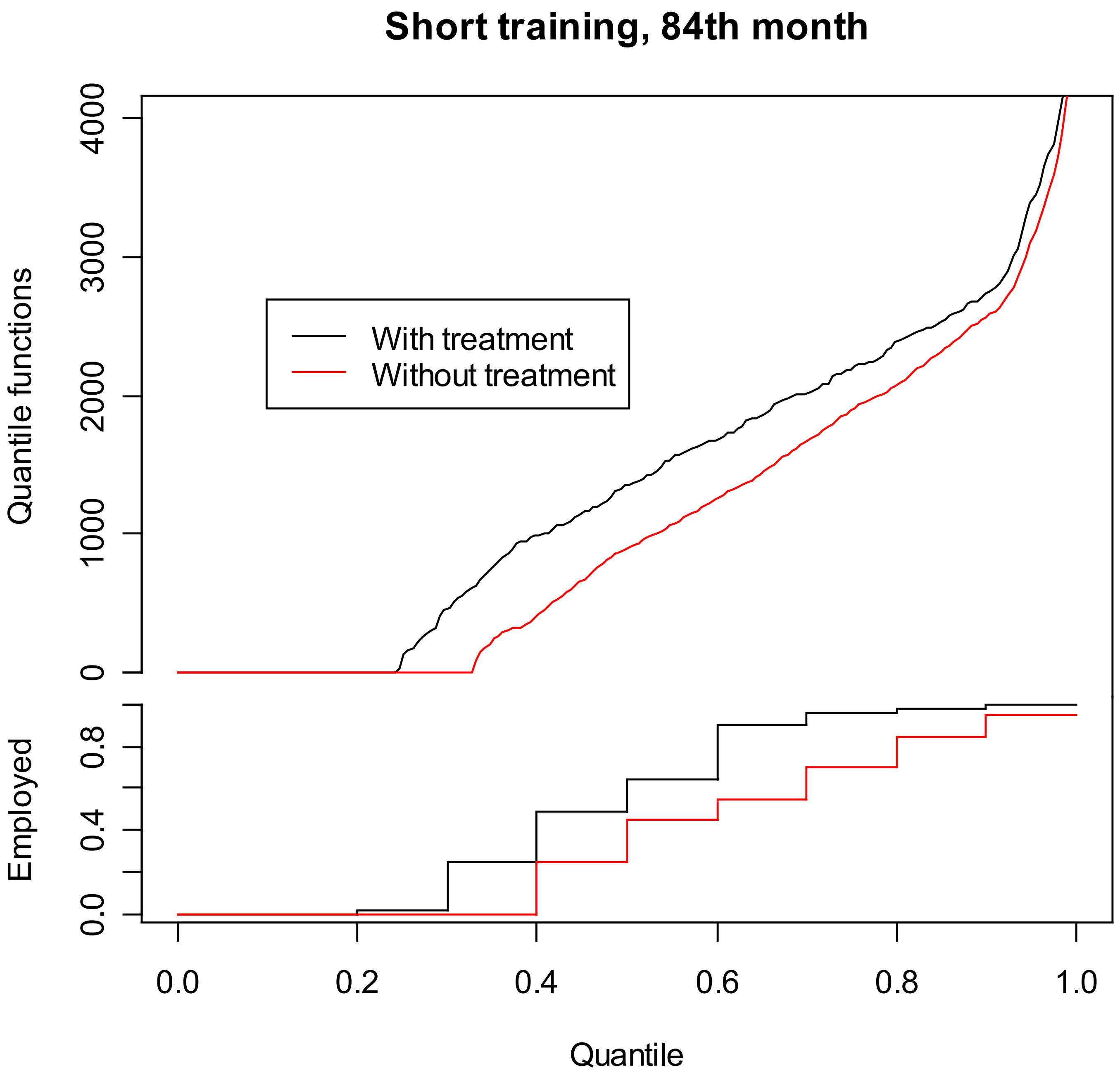
Estimation procedures

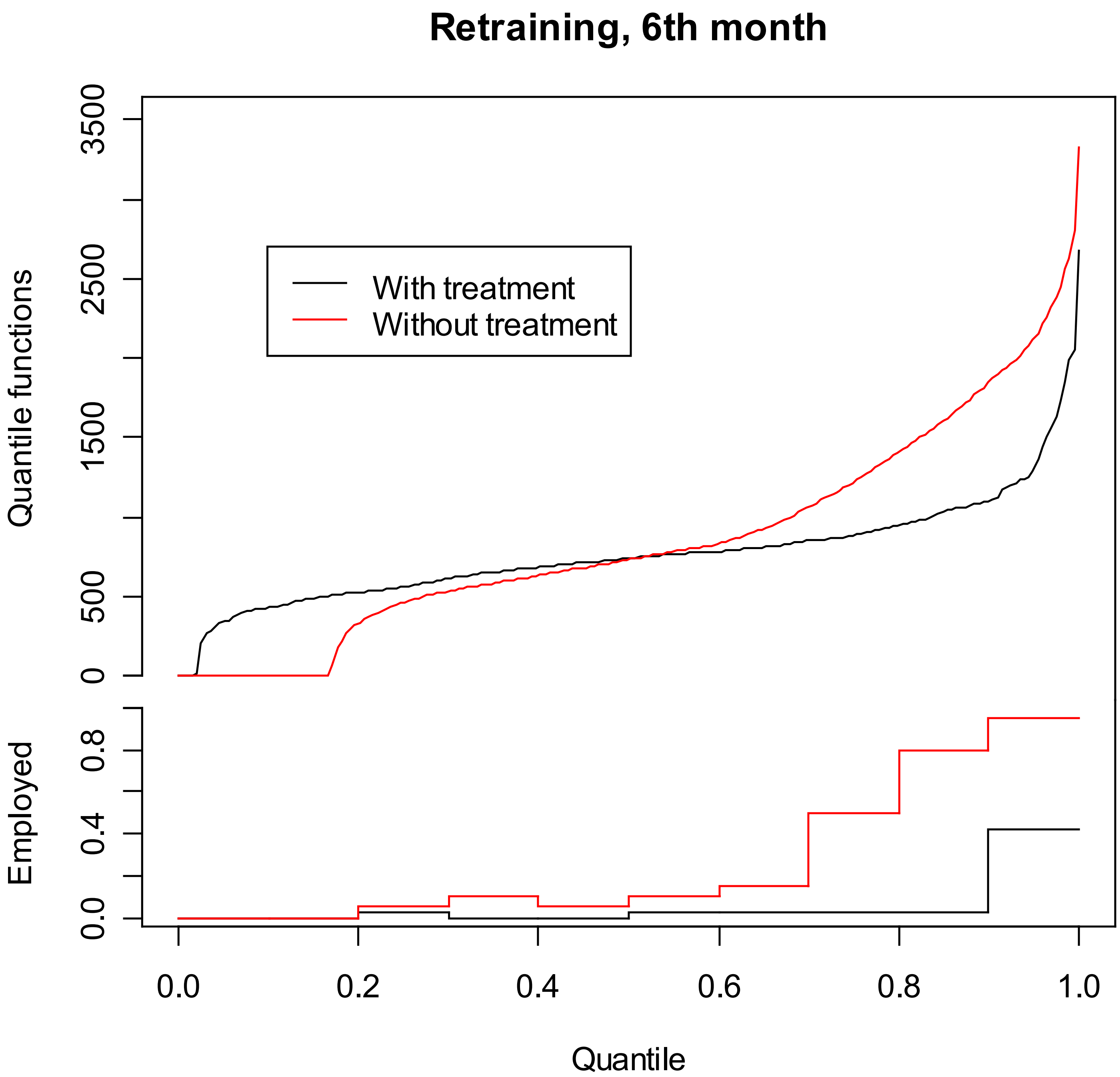
- Estimation of *QTET*; comparison only between each program and nonparticipation
- Estimator: nonparametric regression on the propensity score with exact match on gender. The propensity score is estimated by a parametric probit.
- The conditional distribution function at 0 is estimated by local logit. The rest of the distribution is estimated by local linear quantile regression. We use an Epanechnikov kernel with bandwidth $= sd\left(\hat{p}\left(X\right)\right)n^{-1/4}$
- We plot also the estimated probability of working for each decile of both distributions

$$\Pr\left(D\left(t\right)=1\middle|\underline{q}\leq Y\left(t\right)\leq\overline{q},T=1\right)$$

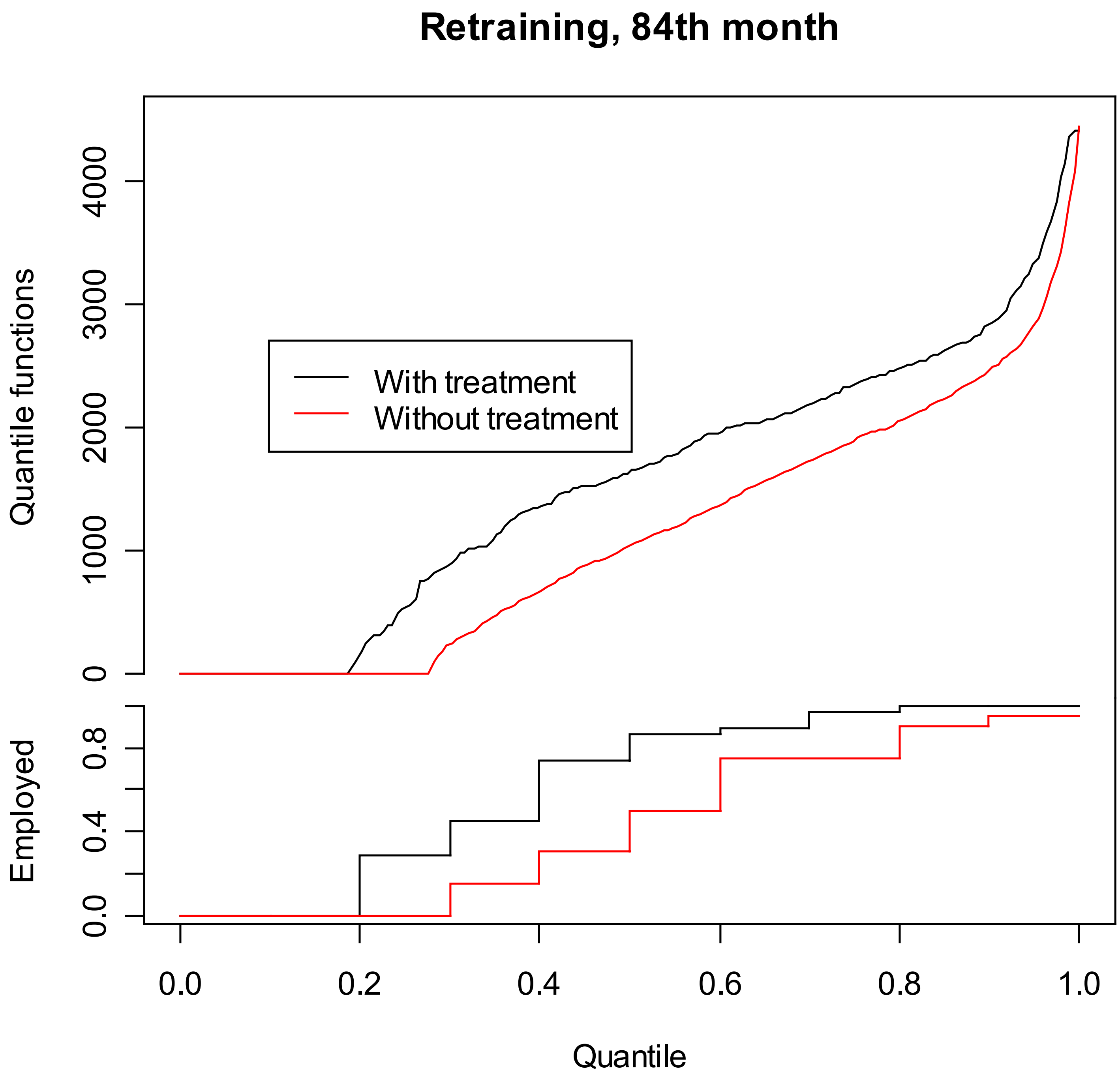


Short training, 6th month
Quantile functions
With treatment
Without treatment
0
1000
2000
3000
4000
Employed
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Quantile

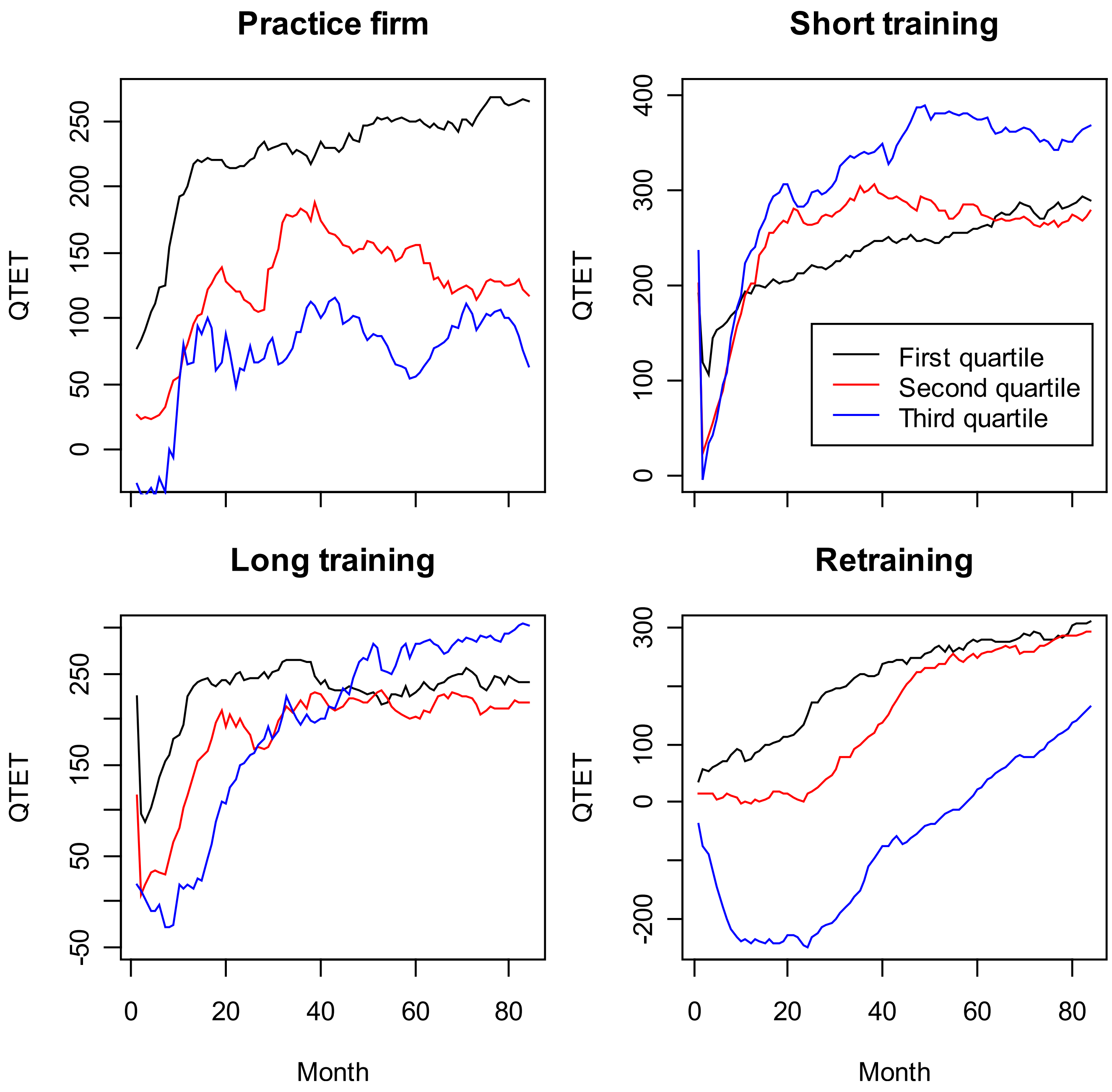

Short training, 84th month
With treatment
Without treatment
Quantile functions
0
1000
2000
3000
4000
Employed
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Quantile



Retraining, 6th month
With treatment
Without treatment
Quantile functions
Employed
Quantile
0
500
1500
2500
3500
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0


Retraining, 84th month
With treatment
Without treatment
Quantile functions
0
1000
2000
3000
4000
Employed
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Quantile

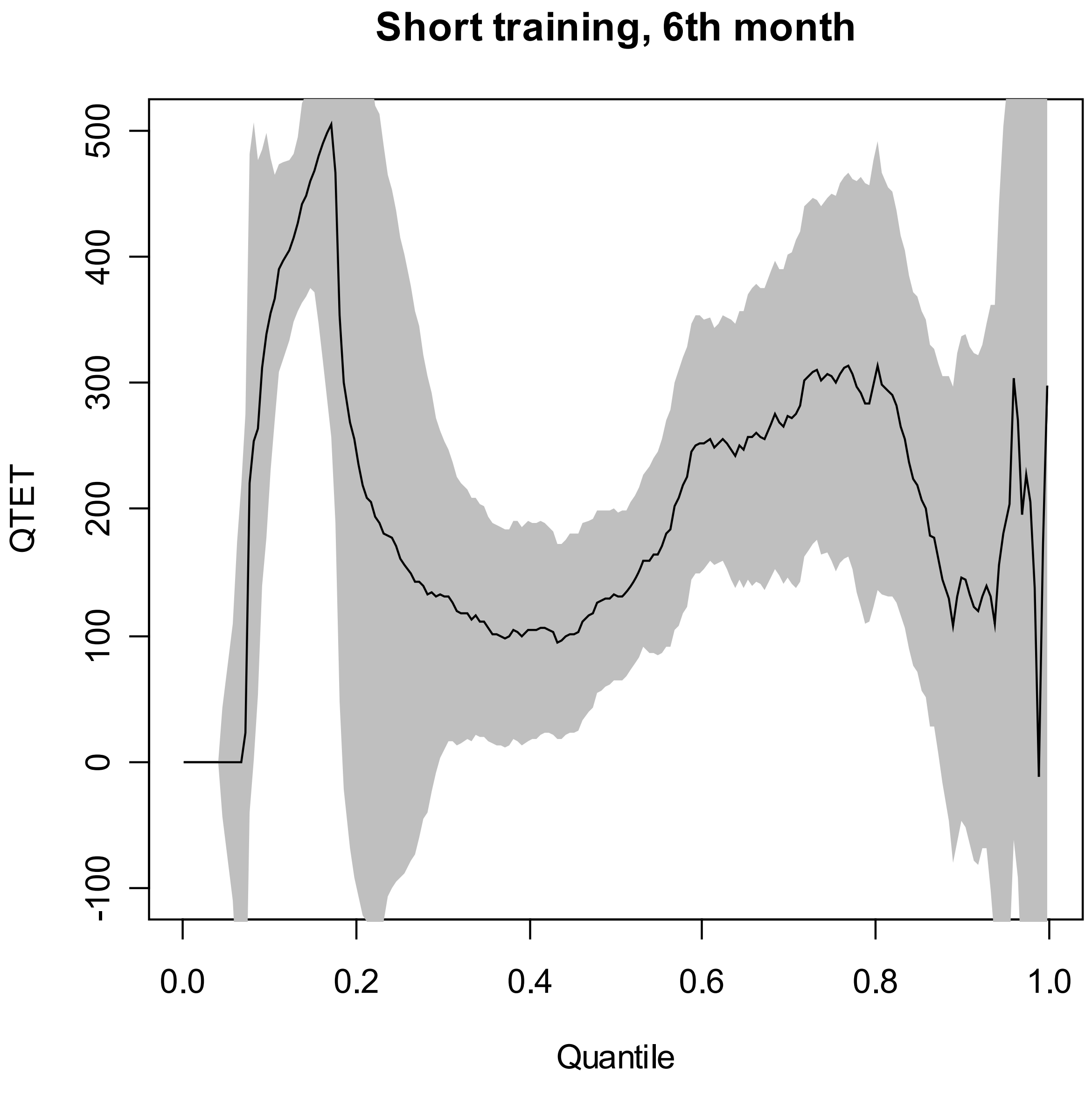
QTET, accumulated incomes

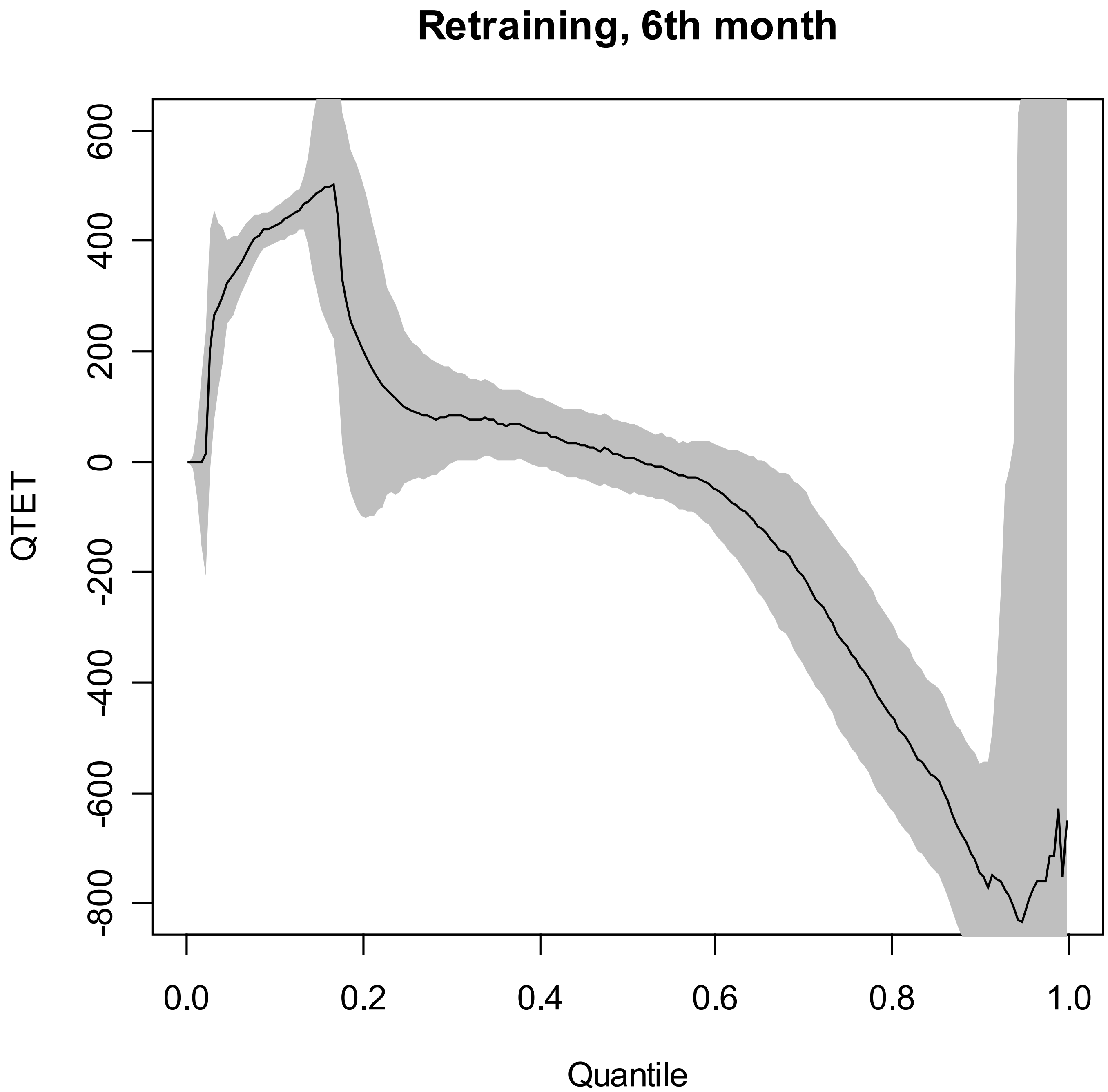


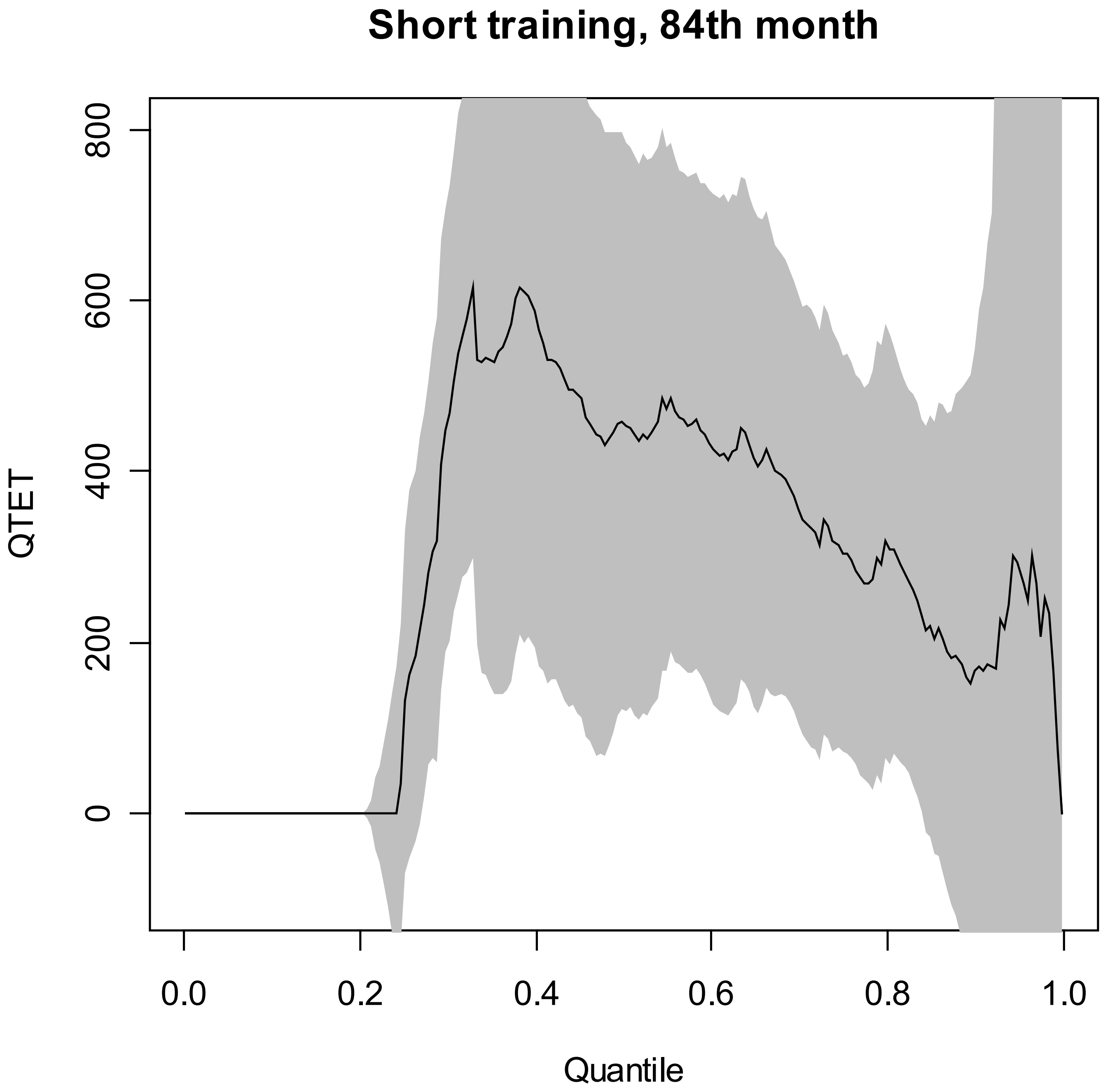
Training programs effects

- For an unemployed, the training programs have 3 effects: on UE benefits claim, employment probability and potential wage. The first affects the lower part of the distribution while the second tends to affect the middle part and the third the upper part.
- Different patterns emerge for the programs, depending on the length of the programs, the severity of the lock-in effect and the amount of human capital added by the participation

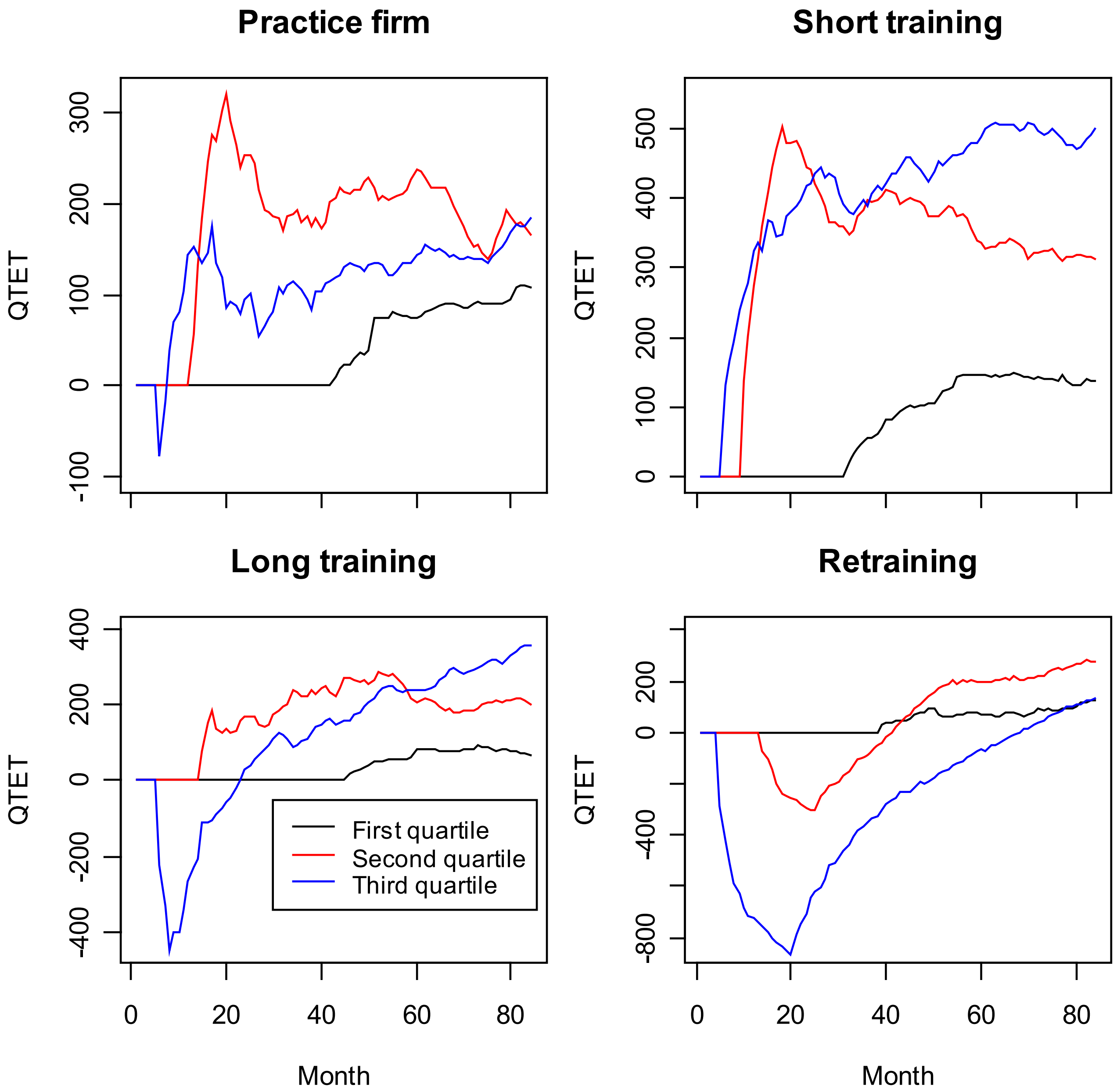
Bootstrap confidence intervals




Retraining, 6th month
QTET
Quantile
600
400
200
0
-200
-400
-600
-800
0.0
0.2
0.4
0.6
0.8
1.0


Short training, 84th month
QTET
Quantile
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0

Accumulated productive incomes (*SY*)



Conclusion: comparison of approaches

- The estimation of *QTE*s under exogeneity can be performed using the same strategies that were proposed for the estimation of mean effects:
  - Matching: no asymptotic results for *QTE*, not efficient for *ATE* (probably also for *QTE*)
  - Propensity score based methods: Firpo (forthcoming), asymptotic efficient, smoothnes of the treatment probability, Monte Carlo simulation for *ATE*s: high variance of the weighting estimator
  - Regression estimators: asymptotic efficient, smoothness of the distribution function

Conclusion: Public sponsored training

- In an application to the evaluation of German ALMP, we have estimated the *QTE*s of training programs for unemployed on their consumptive income. This outcome is interesting: it represents their effective earnings, it would be the basis for deciding to participate or not if this choice is given to the unemployed
- Of course, we would like to know the effects on the potential wage. However, it is not reasonable to assume selection on observables conditionally on working. In the absence of a continuous instrument, only a bounding strategy seems reasonable. We will explore this approach in future works