Expectile Treatment Effects: An efficient alternative to compute the distribution of treatment effects

Stephan Stahlschmidt*
Matthias Eckardt*
Wolfgang K. Härdle*

* Humboldt-Universität zu Berlin, Germany

BERLIN

ECONOMIC RISK

SFB 649

Expectile Treatment Effects: An efficient alternative to compute the distribution of treatment effects*

Stephan Stahlschmidt

School of Business and Economics

Humboldt–Universität zu Berlin

Germany

Matthias Eckardt

Department of Computer Science

Humboldt–Universität zu Berlin

Germany

Wolfgang K. Härdle

C.A.S.E Center for Applied Statistics and Economics

Humboldt–Universität zu Berlin

Germany

Working Paper Version: 20141014

Abstract

The distribution of treatment effects extends the prevailing focus on average treatment effects to the tails of the outcome variable and quantile treatment effects denote the predominant technique to compute those effects in the presence of a confounding mechanism.

The underlying quantile regression is based on a L_1–loss function and we propose the technique of expectile treatment effects, which relies on expectile regression with its L_2–loss function. It is shown, that apart from the extreme tail ends expectile treatment effects provide more efficient estimates and these theoretical results are broadened by a simulation and subsequent analysis of the classic LaLonde data.

Whereas quantile and expectile treatment effects perform comparably on extreme tail locations, the variance of the expectile variant amounts in our simulation on all other locations to less than 80% of its quantile equivalent and under favourable conditions to less than 2/3. In the LaLonde data expectile treatment effects reduce the variance by more than a quarter, while at the same time smoothing the treatment effects considerably.

Keywords: distributional treatment effects, efficiency, expectile treatment effects, LaLonde data, quantile treatment effects

JEL codes: C21, C31, C54, J64

1 Introduction

The analysis of a particular intervention and especially the effect it exhibits on a specific outcome variable has been a central issue in statistical data analysis and several classic techniques were invented to answer this question (Fisher, 1935). Whereas at first the design and analysis of experiments were of much interest, more recently the focus in social sciences has shifted to the evaluation of treatment effects from observational data, as in many circumstances the execution of experiments is infeasible or too costly. In this context treatment effects quantify causal outcomes arising from a clearly identifiable intervention and are for example applied to assess new policies or programmes (e.g. LaLonde, 1986; Dehejia and Wahba, 1999).

These applications most often concentrate on an average effect on the population or some subgroup thereof. However, the predominant Rubin Causal Model (Rubin, 1974) allows for heterogeneity in response to a treatment and the focus on average affects inhibits the disclosure of possibly deviant effects in the tails of the outcome distribution. Contrary quantile treatment effects (Doksum, 1974; Lehmann, 1974) allow to analyse the effect distribution and, given covariates, apply quantile regression techniques (Koenker and Bassett, 1978) to infer the causal effect. Quantile treatment effect have been applied to such diverse topics as teacher shortage in African countries (Bourdon et al., 2010), returns to education in China (Messinis, 2013) or export promotion programmes in developing countries (Volpe Martincus and Carballo, 2010). Recent methodology advances have addressed the usage of instrument variables (Abadie et al., 2002; Chernozhukov and Hansen, 2005) or regression discontinuity designs (Frandsen et al., 2012). Furthermore Machado and Mata (2005) and Firpo et al. (2009) extend quantile treatment effects to the unconditional distribution of the outcome variable.

The application of quantile regression to compute the distribution of treatment effects seems to denote an obvious choice, as its usage of the inverse c.d.f. allows for an intuitive understanding of the location parameter $\tau \in (0,1)$ on the distribution of the outcome variable. Accordingly any quantile q_τ may be denoted as $q_\tau = \mathrm{F}^{-1}(\tau)$. Furthermore quantiles may also be defined as a minimum contrast parameter estimator, as

$$q_\tau = \underset{q_\tau}{\arg\min} \left((1-\tau) \int_{-\infty}^{q_\tau} |v - q_\tau| \, \mathrm{f}(v) \mathrm{d}v + \tau \int_{q_\tau}^{\infty} |v - q_\tau| \, \mathrm{f}(v) \mathrm{d}v \right).$$

Quantile regression subtitutes the quantile q_τ for a regression model based on covariates.

However, given the somewhat arbitrary loss (contrast) function above it may be noted that quantile regression describes only one particular candidate to infer the (conditional) distribution of an outcome variable and expectiles, respectively expectile regression (Newey and Powell, 1987), denote a suitable alternative. Given some location parameter $\alpha \in (0,1)$, expectiles e_α are based on a slight variation of the minimization problem detailed above, as they solve the optimization

$$e_\alpha = \underset{e_\alpha}{\arg\min} \left((1-\alpha) \int_{-\infty}^{e_\alpha} (v - e_\alpha)^2 \, \mathrm{f}(v) \mathrm{d}v + \alpha \int_{e_\alpha}^{\infty} (v - e_\alpha)^2 \, \mathrm{f}(v) \mathrm{d}v \right)$$

and consequently exchange the L_1–loss function of quantiles for a L_2 variant. As before expectile regression generates a regression model for e_α given a set of covariates.

Although some authors have assigned expectiles to the "spittoon" (Koenker, 2013), they have recently gained considerably attention and a lively debate can be found in Kneib (2013a) and the accompanying discussion. Recent applications of expectiles were presented in financial statistics (Taylor, 2008), business statistics (Kokic et al., 2000) and demographics (Schnabel and Eilers, 2009b) and much methodology research on expectiles has concentrated on confidence measures (Duran et al., 2011; Guo and Härdle, 2012; Sobotka et al., 2013).

Whereas expectiles miss the intuitive indication of the location the parameter α defines, Newey and Powell (1987) note that expectiles are more efficient than quantiles. Hence, as both techniques present unbiased estimates, any comparison in terms of mean squared error is restricted to analysing the estimators' variances and in this respect expectiles predominately outperform quantiles. Furthermore the smaller variance in expectile regression affects the computation of treatment effects twice, as the potential outcome framework (Neyman, 1923; Rubin, 1974) demands a regression for the programme group separately from the regression in the control group.

Given this efficiency gain of expectiles, we adopt them to compute the distribution of treatment effects along an outcome variable and consequently introduce the technique of expectile treatment effects. In order to demonstrate the advantages of this novel approach we present a proof that apart from extreme tail ends expectile regression is more efficient than quantile regression, and this advantage of expectiles causes expectile treatment effects to be more efficient than quantile treatment effects.

In a subsequent simulation we drop some of the assumptions, which are necessary for the theoretical proof, and quantify the efficiency gain on broader terms. We find, that apart from extreme tail locations expectile treatment effects possess an at least 20% smaller variance and under favourable conditions the variance of the expectile procedure amounts to only 2/3 of the quantile variant. However, this advantage of expectile treatment effect is not observed at the more difficult tail ends, where both approaches perform comparably.

In an application of the novel expectile treatment effects procedure we analyse the classic LaLonde data (LaLonde, 1986) and contrast the new approach with the standard quantile treatment effect. In our implementation of a counterfactual decomposition approach (Machado and Mata, 2005) we observe that expectile treatment effects and quantile treatent effects present comparable medial results, but the variance of expectile variant amounts to less than 3/4 of the quantile equivalent.

In the next section we will detail the expectile treatment effects technique and present a theoretical comparison with quantile treatment effects in the end. Afterwards in section 3 the efficiency gain of expectile treatment effects is quantified in a simulation and section 4 describes the aforementioned application of both techniques to analyse the LaLonde data. Finally section 4 concludes.

2   Expectile Treatment Effects

The analysis of treatment effects strives to quantify causal consequences generated by a specific intervention, e.g. the implementation of a new programme, a change of policy or the execution of an experiment, and denote a popular tool in the evalua-

tion of medicine, public policies and other interventions. In these settings a portion of the statistical units $i \in \{1, \ldots, N\}$ is treated by a particular measure and the resulting effects, if any, are evaluated by some outcome variable Y_i. An overview on the applied econometric techniques and the accompanying literature is given by, among others, Imbens and Wooldridge (2009), while Pearl (2000) presents a statistical approach to intervention effects based on graphical models and structural equation models, which is, however, not of direct interest to our work.

The evaluation is typically performed by comparing statistical units affected by the measure, denoted by $W_i = 1$, with other units which are not exposed to the measure, that is $W_i = 0$. Obviously at any point in time the same unit can only be analysed in one of the two settings and a direct comparison at the individual level is therefore infeasible (Holland, 1986). The potential outcome framework (Neyman, 1923; Rubin, 1974) overcomes this issue by defining the realized outcome Y_i as a function of the two potential outcomes $Y_i(W_i)$ which arise if the observation is, respectively is not, treated:

$$Y_i = Y_i(W_i) = Y_i(0)(1 - W_i) + Y_i(1)W_i. \tag{1}$$

Apart from the effect of W_i the realized outcome Y_i is often also influenced by other factors. Experiments guarantee by their random assignment into treatment or control group, that any side effects are balanced between the two groups and consequently any observable difference might solely be caused by W. An application of this approach can for example be found in Heinrich et al. (2001) for behavioural economics and several other examples in development economics are cited in the review of Duflo et al. (2008). However, the use of experiments is often limited by moral constrains, e.g. the participation by a free agent can obviously not be forced, but is of voluntary character, or practical issues, e.g. additional costs for setting up an experiment.

On the other hand the quantification of treatment effects from observational data is often obscured by a selection mechanism, if either the statistical units or the corresponding administrator relates the potential participation in the intervention with the expected individual outcome. Such confounding mechanisms render a direct comparison between the treated and untreated statistical units unsuitable and Rosenbaum and Rubin (1983) propose to utilize the conditional independence between intervention participation and outcome given a vector of covariates $\mathbf{X}_i$

$$W_i \perp\!\!\!\perp Y_i(W_i) | \mathbf{X}_i$$

to account for the confounding effect of $\mathbf{X}_i$ and consequently control its influence on the treatment effect. Ideally this approach ensures a so–called unconfounded assignment, also known as selection on observables or exogeneity and depicts the workhorse of non-experimental treatment analysis. However, its properness in a certain research context can not be verified by a statistical test (Pearl, 2000).

Whereas the potential outcome framework allows for heterogeneity in response to the treatment due to its broad modelling approach, most studies are primarily concerned with medial treatment effects on the population, e.g. the so–called average treatment effect (ATE), or some specific subgroups thereof, e.g. the average treatment effect on the treated (ATOT). This limited analysis neglects any deviant treatment effects in the tails of the outcome's distribution and consequently might miss important characteristics of the intervention of interest.

Quantile treatment effects (QTE) describe not only an average effect in terms of the effect on the median, but offer also effects at any quantile of interest and therefore describe the treatment effect's distribution. Doksum (1974) and Lehmann (1974) define QTE as the horizontal difference between the quantiles of two potential outcome functions:

$$QTE_\tau = \mathrm{F}_{Y(1)}^{-1}(\tau) - \mathrm{F}_{Y(0)}^{-1}(\tau), \tag{2}$$

where $\tau \in (0,1)$ defines the quantile and F_Y^{-1} denotes the inverse c.d.f., also known as quantile function. In case of additional covariates, either included for precision or their confouding nature, the conditional quantile treatment effects can be denoted by

$$cQTE_\tau = \mathrm{F}_{Y(1)|\mathbf{X}}^{-1}(\tau) - \mathrm{F}_{Y(0)|\mathbf{X}}^{-1}(\tau). \tag{3}$$

Linear quantile regression (Koenker and Bassett, 1978) denotes one parametric approach to infer the conditional quantiles and facilitates the computation (3). In an i.i.d. setting the linear model

$$Y_i = \mathbf{X}_i^\top \beta_\tau + \varepsilon_i, \text{ with } \mathrm{F}_{\varepsilon_{i,\tau}}(0) = \tau$$

is solved for the estimate $\widehat{\beta}_\tau$ via the optimization constraint

$$\widehat{\beta}_\tau = \arg\min_{\beta_\tau} n^{-1} \sum_{i=1}^n \varrho_\tau \left(Y_i - \mathbf{X}_i^\top \beta_\tau \right), \tag{4}$$

where $\varrho(u) = u\left(\tau - \mathbf{I}\left(u < 0\right)\right)$ denotes the check function and $\mathbf{I}(\cdot)$ the indicator function. Consequently quantile regression is based on an asymmetrically weighted L_1–loss function.

Expectile regression (Newey and Powell, 1987) denotes an alternative approach to infer the conditional distribution $Y_i|\mathbf{X}_i$ at some location $\alpha \in (0,1)$, which also weights the error terms asymmetrically, but optimizes the L_2–loss function

$$\widehat{\beta}_\alpha = \arg\min_{\beta_\alpha} n^{-1} \sum_{i=1}^n \rho_\alpha \left(Y_i - \mathbf{X}_i^\top \beta_\alpha \right) \tag{5}$$

with the asymmetric least squares check function $\rho(u) = u^2 \left|\tau - \mathbf{I}\left(u < 0\right)\right|$. Expectile regression can be understood as a particular form of M–quantile estimation (Breckling and Chamber, 1988), but usually α differs from τ to describe the same location on the conditional distribution $Y_i|\mathbf{X}_i$. Indeed there exists a one–to–one relation between expectiles and quantiles (Jones, 1994; Yao and Tong, 1996) and Schulze Waltrup et al. (2013) exploit this relation to deduce quantiles from expectiles. This inference, although with less efficiency, has also been accomplished by Efron (1991), who obtains so–called percentiles from expectiles by counting the number of observations below the expectile of interest.

Any well defined relation between quantiles and expectiles increases the practicability of expectiles, as contrary to quantiles and due to the expectiles' dependence on the global properties of the respective distribution, expectiles miss an intuitive indication of the exact location any chosen α defines. A further adverse consequence of this dependence is found in the sensitivity of expectile regression to outliers, whereas quantile regression is well–known for its robustness (Koenker, 2013). On the other hand the estimation of expectiles relies on a least asymmetric weighted squares

algorithm (LAWS) and possess therefore a computational advantage if compared to the estimation of quantile regression (Kneib, 2013a). The quantile optimization constraint (4) is not continuously differentiable and instead linear programming techniques are applied. However, this difference affects the actual computation time less than the ease of implementing both approaches (Pertnoy and Koenker, 1997).

Concerning the application of both regression techniques to the analysis of treatment effects a crucial difference between quantile and expectile regression arises from the unequal variances of the respective estimators. Whereas both estimators are unbiased and asymptotically return the true coefficients β, the expectile regression estimator is predominately more efficient than the quantile regression one. This property was at first observed by Newey and Powell (1987) and can be explained by the implied smoothing of expectile regression (Kneib, 2013b) and the information available to each estimator. Whereas expectiles utilise the actual distance between data points and fitted values, quantiles rely solely on the binary information of the data points lying below or above the fitted values (Schnabel and Eilers, 2009a).

In detail, He and Shao (1996) present the asymptotic variance–covariance matrix for the estimate $\widehat{\beta}_\tau$ of quantile regression in the case of independently and possibly identically distributed observations by

$$\mathrm{ACov}(\widehat{\beta}_\tau) = \tau(1-\tau) \left(\sum_{i=1}^{n} \mathrm{f}_{\varepsilon,i}(0) \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top \left(\sum_{i=1}^{n} \mathrm{f}_{\varepsilon,i}(0) \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1}, \quad (6)$$

where $\mathrm{f}_{\varepsilon,i}$ describes the density of the error. The presence of this term complicates the actual computation of (6), as its estimation, and consequently the estimation of $\mathrm{ACov}(\widehat{\beta}_\tau)$, depends on the applied smoothing.

This issue does not translate to expectile regression, as Sobotka et al. (2013) report the equivalent for $\widehat{\beta}_\alpha$ as

$$\mathrm{ACov}(\widehat{\beta}_\alpha) = \left(\sum_{i=1}^{n} \overline{\omega}_{\alpha,i}^\circ \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \sum_{i=1}^{n} \mathrm{Var} \left(\overline{\omega}_{\alpha,i}^\circ \left(y_i - \mathbf{X}_i^\top \beta_\alpha \right) \right) \mathbf{X}_i \mathbf{X}_i^\top \left(\sum_{i=1}^{n} \overline{\omega}_{\alpha,i}^\circ \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1},$$

where $\overline{\omega}_{\alpha,i}^\circ = \mathrm{E}\left[\omega_{\alpha,i}^\circ \right]$ defines the expectation of the true individual weights of the asymmetric weighting schema (5).

An indication for the greater efficiency of expectile regression can be derived from the following lemma and its proof in the appendix.

Lemma 1. *For* $\mathbf{Z}_i = \{Y_i, \mathbf{X}_i\} \sim$ *identically and independently* $\forall$ i *with a distribution according to the c.d.f.* $\mathrm{F}(Y_i|\mathbf{X}_i)$ *and an error term following a normal distribution* $\varepsilon_i \sim \mathrm{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$, *which is independent of the explanatory variables* $\mathbf{X}_i$, *it can be shown that*

1. $\mathrm{ACov}(\widehat{\beta}_\tau) > \mathrm{ACov}(\widehat{\beta}_\alpha)$, *if the location* $\tau \in [\tau_l, \tau_u]$.

2. *The corresponding endpoints* $\tau_{l,u}$ *are defined by*

$$\tau_{l,u} = 0.5 \pm \sqrt{0.25 - \frac{\exp\left(-\frac{\mu_\varepsilon^2}{\sigma_\varepsilon^2}\right)}{2\pi}}.$$

Following the standard assumption of $\mu_\varepsilon = 0$ allows to infer the interval as $\tau \in [0.199, 0.801]$. Obviously $|\mu_\varepsilon| > 0$ increases the interval, whereas a large variance σ_ε^2 decreases it. Nevertheless, under the given assumptions expectile regression is more efficient on most parts of the distribution $Y_i|\mathbf{X}_i$ and is outperformed by quantile regression only on the extreme tail ends. In the subsequent simulation we will consider a reduced set of assumptions to broaden this theoretical result and quantify the actual efficiency gain.

Having noted the superior performance of expectile regression and mirroring QTE defined by (2) expectile treatment effects (ETE) may be introduced as

$$ETE_\alpha = \mathrm{G}_{Y(1)}(\alpha) - \mathrm{G}_{Y(0)}(\alpha),$$

where G_Y denotes the expectile function of Y_i corresponding to a particular location α. Conditional expectile treatment effects also make use of covariates to reduce variance or, in case of confouding variable, to identify the actual treatment effect:

$$cETE_\alpha = \mathrm{G}_{Y(1)|\mathbf{X}}(\alpha) - \mathrm{G}_{Y(0)|\mathbf{X}}(\alpha), \tag{7}$$

where $\mathrm{G}_{Y(W)|\mathbf{X}}$ can be obtained via expectile regression for any given vector $\mathbf{X}_i$.

The difference in variance between the expectile and quantile regression estimator detailed above affects naturally the efficiency of cETE and cQTE and the resulting corollary is stated in the following lemma.

Lemma 2. *Given the conditional distribution of $Y_i|\mathbf{X_i}$ denoted jointly by τ and α, a treatment effect in the form of a global location change and the efficiency gain of expectile regression observed before as $\mathrm{ACov}(\widehat{\beta}_\tau) > \mathrm{ACov}(\widehat{\beta}_\alpha)$, it can be shown that*

$$\mathrm{AVar}(\widehat{cETE}) < \mathrm{AVar}(\widehat{cQTE}).$$

The corresponding proof is stated in the appendix.

3 Simulation

Our simulation is split in two parts. At the start of each bootstrap run an artificial population of size $N = 200,000$ is generated to infer the quantile location parameter τ describing the same locations as its expectiles equivalent α on the distribution $Y_i|\mathbf{X}_i$. Afterwards a sample is drawn from this population and the cETE, respectively cQTE, are computed on this sample. Within a stochastic parametrization framework this two–step procedure is repeated $100,000$ times to observe the variance of both estimators.

The outcome Y_i of the $i \in \{1, \ldots, N\}$ observations is defined as the sum of the gaussian treatment variable $Z_i \sim \mathrm{N}(\mu_{Z,W}, 1)$, three standard normal confouding variables $X_i^{(1,2,3)} \sim \mathrm{N}(0,1)$ and a heteroscedastic error term ε_i to allow for varying treatment effects:

$$Y_i = Z_i + X_i^{(1)} + X_i^{(2)} + X_i^{(3)} + \varepsilon_i. \tag{8}$$

8

In order to provide robust simulation results independent of a specific parametrization, we vary certain parameters randomly along our simulation. Hence in the equation (8) above we allow the standard deviation $SD(\varepsilon_i)$ of the heteroscedastic error term $\varepsilon_i = \epsilon_i \frac{(Z_i+5)}{\kappa}$ to amount to a half or a quarter of the standard deviation of all explanatory variables $SD\left(Z_i + X_i^{(1)} + X_i^{(2)} + X_i^{(3)}\right)$. In detail, $\epsilon_i \sim N(0,1)$ describes a standard normal random variable and $\kappa \sim U(0.5, \{5, 10\})$ denotes a scaling parameter, which is uniformly distributed on the set $\{5, 10\}$ with probability 0.5 for each element. Furthermore in approximately half of the repetitions we randomly allow for a covariance of size $Cov\left(X_i^{(\cdot)}, X_i^{(\cdot)}\right) = 0.25$ among the three confounding variables and no covariance in the other half.

The treatment effect is modeled as a global location shift on the expectation $\mu_{Z,W}$ of the treatment variable Z_i. Whereas observations with a treatment status $W_i = 0$ are not affected by any treatment and consequently $\mu_{Z,0} = 0$, the expectation in the treatment group differs from zero and we randomly apply three different values by setting

$$\mu_{Z,1} \sim U\left(\frac{1}{3}, \{0.7, 1.4, 2.1\}\right)$$

to mirror the average effect each confounding variable $X_i^{(\cdot)}$ possess on the treatment effect in the outcome Y_i.

The assignment method to separate the treatment group from the control group is of stochastic nature to allow for an overlap between both groups and is based on the confounding variables. In detail, the individual participation W_i is determined by an application of the Bernoulli distribution

$$W_i \sim B(\pi_i),$$

where π_i denotes the individual propensity score which is computed as the logistic function

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}.$$

of the sum of the confounding variables $\eta_i = \sum_{p=1}^{3} X_i^{(p)}$.

Consequently due to their relatively high value in the confounding variables $X_i^{(\cdot)}$ approximately half of the population is assigned to the treatment group and consequently affected by the treatment. Their above average realization of $X_i^{(\cdot)}$ also affects the output variable Y_i via (8) and completes therefore the confounding mechanism which has to be controlled for to identify the treatment effect of Z_i. The other half of the population is not affected by the treatment and might serve in the control group.

Before we draw one sample from the artificial treatment population and another one from the artificial control population to compute the cETE and cQTE, we approximately determine τ corresponding to

$$\alpha = \{0.01, 0.1, 0.25, 0.75, 0.9, 0.99\}.$$

Hence we regress in an expectile regression framework the confouding variables $X_i^{(1,2,3)}$ on the outcome Y_i and observe how many observations possess smaller conditional outcome values $Y_i|X^{(1,2,3)}$ than the computed conditional expectiles $Y_\alpha|X^{(1,2,3)}$ (Efron, 1991):

$$\tau = N^{-1} \sum_{i=1}^{N} \mathbf{I}\left(Y_i < Y_\alpha \mid X^{(1,2,3)}\right). \tag{9}$$

To facilitate the computation (9) we condition on the small interval $[\overline{x} - 0.25\sigma_x, \overline{x} + 0.25\sigma_x]$, where $\overline{x}$ denotes the population wide multivariate mean of the confouding variables and σ_x describes the variable specific standard deviation. This definition and the size of the artificial population $N = 200,000$ ensure that in every repetition the observed empirical relation between τ and α is based on more than 1000 observations.

Having established the relation between quantiles and expectiles in our simulation, we continue with the second step of actually comparing cQTE and cETE. Therefore we draw a random sample of the treated population and the untreated population. This size of the treatment sample n_t, respectively control sample n_c, are determined stochastically according to

$$n_t \sim \mathrm{U}\left(\frac{1}{3}, \{100, 500, 1000\}\right)$$
$$n_c \sim \mathrm{U}\left(\frac{1}{3}, \{2n_t, 5n_t, 10n_t\}\right).$$

Hence the combined samples include between 300 and 11,000 observations and the control sample outnumbers the treatment sample to allow for the necessary overlap in the confouding variables between both samples.

Afterwards quantile regression, respectively expectile regression, are applied separately to both samples and the cQTE, respectively the cETE conditional on the population mean of the confouding variables $\overline{X}^{(1,2,3)}$ are computed according to (3) and (7). Figure 1 presents the distribution of both estimators, which arise due to the 100,000 simulation repetitions.

In our simulation both estimators report on average the same unbiased estimates, which however vary due to the introduced heteroscedasticity along the distribution of the conditional outcome variable $Y_i|\overline{X}_i$ with the location parameter α. For example, at $\alpha = 0.5$ both approaches generate unbiased estimates of the ATE $\mu_{Z,1} = 0.7$, as $\overline{cETE}_{0.5} = 0.700$ and $\overline{cQTE}_{0.5} = 0.704$, but in the lower tail at $\alpha = 0.1$ the mean cETE (cQTE) of 0.651 (0.652) is considerably lower than the reported 0.750 (0.757) at the upper tail location $\alpha = 0.9$.

However there is also some visible variation present in the estimates and for example at $\alpha = 0.1$ the variance of the expectile (quantile) treatment estimators amounts to 0.019 (0.027). As with the mean, the size of the variance varies with the location parameters α, respectively τ, as any treatment effects in the tails are more difficult to determine than the more centric ones. Apart
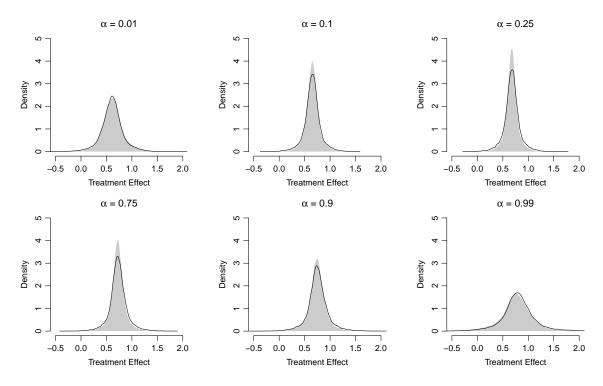
10

Figure 1: *Distribution of the cETE (grey area) and cQTE (solid black line) of a simulated treatment effect $\mu_{Z,1} = 0.7$ at various locations α.*

from this specific feature of the applied estimation procedures, the observed variation is also caused by the stochastic data generating process and the specific parametrization.

In detail, a treatment group size of $n_t = 100$ raises the variance sixfold ($\mu_{Z,1} = 2.1$) to tenfold ($\mu_{Z,1} = 0.7$), if compared to a treatment group of $n_t = 1000$ observations. The same reasoning holds for the size of the control group, as a being able to compare the treatment group to a control group of only double its size increases the variance of the treatment estimate by 20% ($\mu_{Z,1} = 2.1$) to 30% ($\mu_{Z,1} = 0.7$) if compared with a control group of ten times the size of the treatment group. Whereas the existence of covariance among the confounding variable increases the variance only slightly by around 7% for all values of $\mu_{Z,1}$, the incrases of the error's standard deviation from a quarter to a half of the standard deviation of all explanatory variables raises the variance sharply by 72% ($\mu_{Z,1} = 0.7$) to 103% ($\mu_{Z,1} = 2.1$). Finally our simulation indicates that the estimators' variances are positively correlated with the treatment size, as increasing the simulated treatment effect from $\mu_{Z,1} = 0.7$ to $\mu_{Z,1} = 2.1$ increases the variance by 36%. This observation might be explained by the resulting reduction in the essential overlap between both samples.

Apart from these general observations, which hold for both estimators, there persists, apart from extreme locations, a statistically significant difference between both estimators. As one may note in Figure 1, the density of the cETE estimator is visibly more concentrated on the mean than the density of the cQTE etimator, as the cETE estimates vary less than the cQTE estimates. However, this difference decreases with less centric locations and both estima-

11

α	$\mu_{Z,1}$		
	0.7	1.4	2.1
0.01	0.951	1.071	1.196
0.10	0.724	0.792	0.880
0.25	0.650	0.681	0.711
0.75	0.681	0.679	0.674
0.90	0.776	0.781	0.794
0.99	0.994	1.010	1.041

Table 1: *Ratio θ of* $\mathrm{Var}(cETE)$ *to* $\mathrm{Var}(cQTE)$ *for a simulated treatment effect* $\mu_{Z,1}$ *computed at several locations* α.

tors can not be distinguished visually at extreme locations. Still this empirical finding of a predominately smaller variance of the cETE estimator supports and, via the broad parametrization, extends the theoretical arguments detailed above.

In order to quantify this difference Table 1 compares both estimators and reports the ratio θ of the observed variance of the cETE estimates to the variance of the cQTE estimates

$$\theta = \frac{\mathrm{Var}(cETE)}{\mathrm{Var}(cQTE)}$$

for several locations α and various simulated treatment effect sizes $\mu_{Z,1}$.

Apart from extreme locations and one less pronounced occurrence at $\mu_{Z,1} = 2.1$ the cETE estimator possess in our simulation an at least 20% smaller variance than the cQTE estimator and under favourable conditions the variance of the cETE procedure amounts to only 2/3 of the cQTE equivalent. This substantial increase in efficiency is observable for all treatment values $\mu_{Z,1}$, but decreases slightly as the treatment effect size is raised. However, this gain is neutralised in the extreme tails, where both estimating procedures result in comparable variances. In summary, these findings fit well with the theoretical results above and show, that the potential gain of cETE is substantial, whereas the efficiency loss in the extreme tail ends seems to be negligible.

4 Application

Given their nature cETE and cQTE only report conditional results, which restricts their usefulness in many cases. Their generalization is limited by the observation that for example in the case of quantiles $\mathrm{F}^{-1}(\tau) \neq \mathrm{E}_X[\mathrm{F}^{-1}(\tau)|X]$ and therefore their direct application to compute the more interesting unconditional effect of the treatment on the output variable is inhibited. Indeed a conditional treatment effect at the upper tail of $Y_i|X_i$ may describe an effect at the lower tail for unconditional Y_i, if the particular $X_i = x_i$ values assign the observation to the lower tail of the unconditional distribution of Y_i.

One remedy to overcome these restrictions is denoted by the adoption of quantile regression to the counterfactual decomposition approach (Gosling et al., 2000; Machado and Mata, 2005), which has for example been implemented to study the gender wage gap in Sweden (Albrecht et al., 2003) and wage inequality in the U.S. (Melly, 2005). The counterfactual decomposition framework (Juhn et al., 1993) divides the total difference between the treatment and control group into a confounding and a treatment effect, that is

$$\mathrm{F}_{Y\langle 1|1\rangle} - \mathrm{F}_{Y\langle 0|0\rangle} = \left(\mathrm{F}_{Y\langle 1|1\rangle} - \mathrm{F}_{Y\langle 0|1\rangle}\right) + \left(\mathrm{F}_{Y\langle 1|1\rangle} - \mathrm{F}_{Y\langle 1|0\rangle}\right), \qquad (10)$$

where $\mathrm{F}_{Y\langle r|s\rangle}(Y) = \int \mathrm{F}_{Y_r}(Y|X)\,\mathrm{d}\,\mathrm{F}_{X_s}(X)$ denotes the distribution of the output variable Y_i arsing from integrating the characteristics X_i out of the conditional distribution $Y_i|X_i$ and $r, s \in \{0, 1\}$ indicate the control, respectively treatment group. Hence this approach identifies the treatment and confouding effect by allowing for a factual and counterfactual combination of characteristics X_i and valuation of these characteristics $Y_i|X_i$.

Accordingly the first subtraction on the RHS in equation (10) denotes the treatment effect arising from a difference in the valuation of equal characteristics among control and treatment group, whereas the second subtraction describes the confounding effect caused by a difference between control and treatment group in their particular characteristics. Consequently the counterfactual decomposition denotes a variant of the Blinder–Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) and we adapt expectile regression to this approach to study the classic LaLonde data set (LaLonde, 1986) on the effect of the National Supported Work (NSW) Demonstration programme.

The resulting unconditional quantile, respectively expectile treatment effect may be denoted as

$$uTE(l) = \phi(\mathrm{F}_{Y\langle 1|1\rangle}, l) - \phi(\mathrm{F}_{Y\langle 0|1\rangle}, l), \qquad (11)$$

where $\phi(\mathrm{F}_Y, l) \in \{\mathrm{F}_Y^{-1}(l), \mathrm{G}_Y(l)\}$ describes either a quantile or expectile at the location $l \in \{\alpha, \tau\}$ of the distribution of the outcome variable Y_i. In order to estimate the factual and counterfactual densities in equation (11) we follow the algorithm of Machado and Mata (2005) and note that instead of quantile regression also expectile regression may be deployed to estimate the conditional distribution $\mathrm{F}_{Y_r}(Y|X)$, as the estimated conditional expectile function also denotes a consistent estimator of the population expectile function and may describe the conditional distribution in its entirety (Newey and Powell, 1987).

Given the argumentation by Dehejia and Wahba (1999) we restrict our analysis to those male observations with information on their pre–treatment wage in 1975 and 1974 and consequently regress the 1978 wage linearly on age, education, an indicator for not possessing a degree, indicators for African Americans and Hispanic Americans, marital status and the aforementioned pre–treatment wages. We also define the four observations earning more than $30,000$USD as outliers and exclude them beforehand. Due to the restricted number of observations, 259 control and 182 treatment observations, we set the number
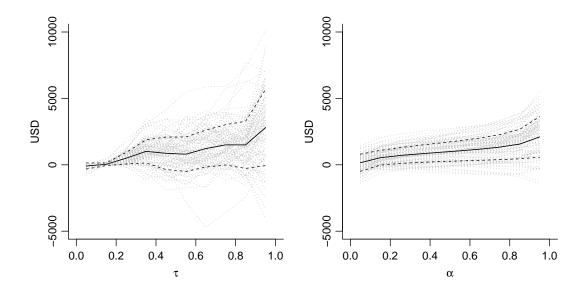
Figure 2: *Mean (black solid line) and standard deviation (black dashed line) from the distribution of the treatment effect from QTE (left) and ETE (right) accompanied by 100 random draws (grey dotted lines) from the 1000 bootstrap runs.*

of random draws in the algorithm to $m = 150$ and obtain a variation in the quantile and expectile estimation procedures via the same bootstrap samples arising from 1000 runs.

Due to the experimental character of the NSW data, we observe hardly any confounding effect at any location of the output variable. In addition the rather simplistic model and small set of confounding variables do not allow for the confouding effect and the displayed treatment effect to add up to the overall difference between both groups. A sizeable mean residual of up to 26% (ETE), respectively 28% (QTE) of the total difference can not be explained by the model and, together with the rather small number of observations, probably explain, why neither ETE nor QTE report average values close to the average treatment effect of 1314USD obtained by comparing the 1978 mean wage in the control and treatment group.

Figure 2 compares the estimates resulting from either applying quantile regression or expectile regression, where we compute quantiles, respectively expectiles on the locations $\alpha = \lambda = \{0.5, 0.15, \ldots, 0.85, 0.95\}$ for the factual and counterfactual margials. In detail, the left panel of Figure 2 is at first based on linear quantile regression to obtain the conditional distribution and secondly displays the difference in quantiles between the resulting factual and counterfactual marginals. Accordingly the right panel in Figure 2 is excluively based on linear expectile regression and expectiles of the resulting marginals.

Both procedures basically present the same results of an overall positive effect, which is rather small at the lower tail, but tripled at the upper tail. Consequently in absolute terms the NSW Demonstration programme benefited high

income earners stronger than low income earners. Comparing both procedures in Figure 2, it might be noted that due to its nature the ETE estimates present a much smoother picture. On the other hand QTE presents a broader picture as setting the same value for α and τ results in more extreme tail locations for QTE than for ETE. This inherited behaviour of QTE and ETE from its distinct regression techniques complicates any direct comparison of the respective variances, as our simulation has shown that the variance increases in the tails. Neglecting this nuisance ETE presents a substantially smaller variance than QTE, as the sum of the standard deviations at the aforementioned locations by applying ETE amounts to only 71% of its QTE equivalent, well in line with our simulation results. In addition some bootstrap samples causes QTE to report treatment effects relatively far from its overall mean, whereas the corresponding ETE estimates stick much closer to the ETE average treatment effects.

5   Conclusion

Observing the distribution of a treatment effect instead of solely relying on an average facilitates the detection of deviant treatment effects in the tails. Quantile treatment effects and expectile treatment effects have both been shown to generate unbiased estimates and whereas quantile treatment effects have been applied frequently, expectile treatment effects denote a related, but novel technique with substantial advantages in terms of mean squared error.

Its smaller variance, if compared to quantile treatment effects, has been proven for the main chunk of the distribution, whereas in the more difficult extreme tail ends expectile treatment effects has been observed to be as efficient as quantile treatment effects. This performance stems from the predominately more efficient expectile regression estimation, which due to the potential outcome framework in the Rubin Causal Model affects the estimation of treatment effects twice, once in the estimation for the programme group and once in the equivalent procedure for the control group.

Furthermore, Schulze Waltrup et al. (2013) indicate that also in an experimental setting, in which confounding variables are balanced between both groups and can consequently be omitted, the application of expectiles could be advantageous. And finally the limitation to linear models throughout our study has been taken for simplicity and computation convenience. The difference in the contrast function between the exepctile and quantile approach is unaffected by the form of the regression model applied to estimate q_τ, respectively e_α in the aforementioned optimization problems and consequently the observations made in the linear case should also carry over to more flexible non–linear regression models.

Still in certain settings efficiency in mean squared error will not denote the only criterion to choose between expectile treatment effects and its quantile cousin. Whereas Koenker (2013) points out that in theory quantiles can be computed as fast as expectiles and therefore the slower computation of quan-

tiles is only due to the ease of implementing both approaches, the superior robustness of quantiles could favour the application of QTE. Contrary to expectiles, quantiles are hardly affected by any outliers and consequently expectile treatment effects may result in biased estimates, if such outliers are not excluded beforehand or the data is not well behaved. These observations will disproportionately affect the extreme tail ends, at which expectile treatment effects performs only on a par with quantile treatment effects even on standard data.

Hence, if the extreme tail ends are of particular interest, quantile treatment effects may present more reliable results. If on the other hand the entire treatment effect distribution is of concern and general observations on deviant treatment effects in the tails are of interest, than the more efficient expectile treatment effects seem to denote a reasonable choice. In these cases the missing intuitive indication of the exact location any chosen α defines seems irrelevant, as not a specific α is of interest, but several values and in particular the resulting distribution of the treatment effect. In many applications the analysis of the entire treatment distribution seems more expedient and under these circumstances expectile treatment effects propose an efficient alternative to the established quantile treatment effects.

A   Proofs

A.1   Lemma 1

We assume $\mathbf{Z}_i = \{Y_i, \mathbf{X}_i\} \sim$ identically and independently $\forall\ i$ according to the c.d.f. $F(Y_i|\mathbf{X}_i)$ and define a gaussian error term $\varepsilon_i \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$, which is independent of the explanatory variables $\mathbf{X}_i$. Furthermore we allow τ and α to denote the same location and, if feasible, suppress the location subscript to simplify the notation.

Given the stated proposition

$$\mathrm{ACov}\left(\widehat{\beta}_\tau\right)\ >\ \mathrm{ACov}\left(\widehat{\beta}_\alpha\right),$$

we can combine the definitions of the respective variances to the inequality

$$\tau(1-\tau)\left(\sum_{i=1}^{n}\mathbf{x}_i\,\mathrm{f}_{i,\varepsilon}(0)\mathbf{x}_i^\top\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\right) > \left(\sum_{i=1}^{n}\overline{\omega}_i^\circ\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\,\mathrm{Var}\left(\overline{\omega}_i^\circ\varepsilon_i\right)\right)$$

$$\times\left(\sum_{i=1}^{n}\mathbf{x}_i\,\mathrm{f}_{i,\varepsilon}(0)\mathbf{x}_i^\top\right)^{-1} \qquad\qquad \times\left(\sum_{i=1}^{n}\overline{\omega}_i^\circ\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1},$$

where we separate the p.d.f. of the error $\mathrm{f}_{i,\varepsilon}(0)$, respectively the expectation of the true individual weights $\overline{\omega}_i^\circ$ from the inverse summation terms due to the i.i.d. assumption and subsequently reduce those summation terms:

$$\tau(1-\tau)\,\mathrm{f}_\varepsilon(0)^{-2}\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\right)\ >\ (\overline{\omega}_i^\circ)^{-2}\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\,\mathrm{Var}\left(\overline{\omega}_i^\circ\varepsilon_i\right)\right).$$

Excluding the parameter $\overline{\omega}_i^{\circ}$ from the variance terms allows to cancel them and the independence between the error variance $\text{Var}(\varepsilon_i) = \sigma_{\varepsilon_i}^2$ and $\mathbf{X}_i$ allows to cancel the remaining summation terms. Entering the assumed functional form of the error, the inequality becomes

$$\tau(1-\tau)\left(\frac{1}{\sqrt{2\pi}\sigma_{\varepsilon}}\exp\left(-\frac{(0-\mu_{\varepsilon})^2}{2\sigma_{\varepsilon}^2}\right)\right)^{-2} > \sigma_{\varepsilon}^2,$$

which we can be solved for τ to obtain the aforementioned solution for $\tau_{l,u}$ as

$$\tau_{l,u} = 0.5 \pm \sqrt{0.25 - \frac{\exp\left(-\frac{\mu_{\varepsilon}^2}{\sigma_{\varepsilon}^2}\right)}{2\pi}}.$$

A.2 Lemma 2

We assume a treatment effect in form of a global location change and allow as before τ and α to jointly denote the same location on $Y_i|\mathbf{X_i}$.

The stated proposition

$$\text{AVar}\left(\widehat{cETE}\right) < \text{AVar}\left(\widehat{cQTE}\right)$$

$$\iff \text{AVar}\left(\widehat{G_{Y(1)}}(\alpha) - \widehat{G_{Y(0)}}(\alpha)\right) < \text{AVar}\left(\widehat{F_{Y(1)}^{-1}}(\tau) - \widehat{F_{Y(0)}^{-1}}(\tau)\right)$$

can be rewritten as a sum of variances due to the independent samples:

$$\text{AVar}\left(\widehat{G_{Y(1)}}(\alpha)\right) + \text{AVar}\left(\widehat{G_{Y(0)}}(\alpha)\right) < \text{AVar}\left(\widehat{F_{Y(1)}^{-1}}(\tau)\right) + \text{AVar}\left(\widehat{F_{Y(0)}^{-1}}(\tau)\right).$$

Owning to the definition of the treatment effect, this inequality is fulfilled, if

$$\text{AVar}\left(\widehat{G_{Y(\cdot)}}(\alpha)\right) < \text{AVar}\left(\widehat{F_{Y(\cdot)}^{-1}}(\tau)\right)$$

$$\iff \mathbf{X}^{\top}\text{ACov}\left(\widehat{\beta}_{\alpha}\right)\mathbf{X} < \mathbf{X}^{\top}\text{ACov}\left(\widehat{\beta}_{\tau}\right)\mathbf{X}$$

$$\iff \text{ACov}\left(\widehat{\beta}_{\alpha}\right) < \text{ACov}\left(\widehat{\beta}_{\tau}\right),$$

which has been shown to hold given the conditions stated in lemma 1.

References

Abadie, A., Angrist, J. and Imbens, G. (2002). Instrumental variables estimation of qunatile treatment effects. *Econometrica*, **70**, 91–117.

Albrecht, J., Björklund, A. and Vroman, S. (2003). Is their a glass ceiling in Sweden?. *Journal of Labor Economics*, **21**, 145–177.

Blinder, A.S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, **8**, 436–455.

Bourdon, J., Fröhlich, M. and Michaelowa, K. (2010). Teacher shortage, teacher contracts and their effect on education in Africa. *Journal of the Royal Statistical Society: Series A*, **173**, 93–116.

Breckling, J. and Chambers, R. (1988). M–qunatiles. *Biometrika*, **75**, 761–771.

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73**, 245–261.

Dehejia, R.H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, **94**, 1053–1062.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two–sample case. *Annals of Statistics*, **2**, 221–233.

Duflo, E., Glennerster, R. and Kremer, M. (2008). Using randomization in development economics research: A toolkit. In *Handbook of Development Economics, Volume 4*, ed. Schultz, T.P. and Strauss, J., 3895–3962, Amsterdam: Elsevier.

Duran, E.A., Guo, M. and Härdle, W.K. (2011). A confidence corridor for expectile functions. *CRC 649 Discussion Paper*, no. 2011–004, Humboldt–Universität zu Berlin.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, **1**, 93–125.

Firpo, S., Fortin, N.M. and Lemieux, T. (2009). Unconditional qunatile regression. *Econometrica*, **77**, 953–973.

Fisher, R.A. (1935). *The Design of Experiments* Edinburgh: Oliver and Boyd.

Frandsen, B., Fröhlich, M. and Melly, B. (2012). Qunatile treatment effects in the regression discontinuity design. *Journal of Econometrics*, **168**, 382–395.

Gosling, A., Machin, S. and Meghir, C. (2000). The changing distribution of male wages in the U.K. *Review of Economic Studies*, **67**, 635–666.

Guo, M. and Härdle, W.K. (2012). Simultaneous confidence bands for expectile functions. *Advances in Statistical Analysis*, **96**, 517–541.

He, X. and Shao, Q.M. (1996). A general Bahadur representation of M–estimators and its application to linear regression with nonstochastic designs. *Annals of Statistics*, **24**, 2608–2630.

Heinrich, J., Boyd, R., Bowies, S., Camerer, C., Fehr, E., Gintis, H. and McElreath, R. (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *The American Economic Review*, **91**, 73–78.

Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 946–960.

Imbens, G.W. and Wooldridge, J.M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, **47**, 5–86.

Jones, M. (1994). Expectiles and M–quantiles are quantiles. *Statistics & Probability Letters*, **20**, 149–153.

Juhn, C., Murphy, K.M. and Pierce, B. (1993). Wage inequality and the rise in returns to skills. *The Journal of Political Economy*, **101**, 410–442.

Kneib, T. (2013a). Beyond mean regression. *Statistical Modelling*, **13**, 275–303.

Kneib, T. (2013b). Rejoinder. *Statistical Modelling*, **13**, 373–385.

Kokic, P., Cahmbers, R. and Beare, S. (2000). Microsimulation of business performance. *International Statistical Review*, **68**, 259–275.

Koenker, R. (2013). Discussion: Living beyond our means. *Statistical Modelling*, **13**, 323–333.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, **46**, 33–50.

LaLonde, R. (1986). Evaluating the econometric evaluation of training programs. *American Economic Review*, **76**, 604–620.

Laplace, P.S. (1820). Théorie analytique des probabilités. Paris: Courcier.

Lehmann, E.L. (1974). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Machado, J. A.F. and Mata, J. (2005). Counterfactual decomposition of changes in wage distribution using quantile regression. *Journal of Applied Econometrics*, **20**, 445–465

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

Messinis, G. (2013). Returns to education and urban–migrant wage differentials in China: IV quantile treatment effects. *China Economic Review*, **26**, 39–55.

Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, **10**, 1–51.

Oaxaca, R.L. (1973). Male–Female Wage Differentials in Urban Labor Markets. *International Economic Review*, **14**, 693–709.

Pearl, J. (2000). *Causality: models, reasoning and inference.* Cambridge: MIT Press.

Portnoy, S. and Koenker, R. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of squared-error vs. absolute-error estimators. *Statistical Science*, **12**, 279–300.

Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects *Biometrika*, **70**, 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology, **66**, 688–701.

Schnabel, S.K. and Eilers, P.H.C. (2009a). Optimal expectile smoothing. *Computational Statistics and Data Analysis*, **53**, 4168–4177.

Schnabel, S.K. and Eilers, P.H.C. (2009b). An analysis of life expectancy and economic production using expectile frontier zones. *Demographic Research*, **21**, 109–134.

Schulze Waltrup, L., Sobotka, F., Kneib, T. and Kauermann, G. (2013). Expectile and quantile regression – David and Goliath? *Working paper, Ludwig–Maximilians–Universität Munich.*

Sobotka, F., Kauermann, G., Schulze Waltrup, L. and Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, **23**, 135–148.

Taylor, J.W. (2008). Estimating value at risk and expected shortfall using expectiles *Journal of Financial Econometrics*, **6**, 231–252.

Volpe Martincus, C. and Carballo, J. (2010). Beyond the average effect: The distributional impacts of export promotion programs in developing countries. *Journal of Development Economics*, **92**, 201–214.

Yao, Q. and H. Tong (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, **6**, 273–292.

SFB 649 Discussion Paper Series 2014

001 "Principal Component Analysis in an Asymmetric Norm" by Ngoc Mai Tran, Maria Osipenko and Wolfgang Karl Härdle, January 2014.

002 "A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data" by Lijie Gu, Li Wang, Wolfgang Karl Härdle and Lijian Yang, January 2014.

003 "An Extended Single Index Model with Missing Response at Random" by Qihua Wang, Tao Zhang, Wolfgang Karl Härdle, January 2014.

004 "Structural Vector Autoregressive Analysis in a Data Rich Environment: A Survey" by Helmut Lütkepohl, January 2014.

005 "Functional stable limit theorems for efficient spectral covolatility estimators" by Randolf Altmeyer and Markus Bibinger, January 2014.

006 "A consistent two-factor model for pricing temperature derivatives" by Andreas Groll, Brenda López-Cabrera and Thilo Meyer-Brandis, January 2014.

007 "Confidence Bands for Impulse Responses: Bonferroni versus Wald" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, January 2014.

008 "Simultaneous Confidence Corridors and Variable Selection for Generalized Additive Models" by Shuzhuan Zheng, Rong Liu, Lijian Yang and Wolfgang Karl Härdle, January 2014.

009 "Structural Vector Autoregressions: Checking Identifying Long-run Restrictions via Heteroskedasticity" by Helmut Lütkepohl and Anton Velinov, January 2014.

010 "Efficient Iterative Maximum Likelihood Estimation of High-Parameterized Time Series Models" by Nikolaus Hautsch, Ostap Okhrin and Alexander Ristig, January 2014.

011 "Fiscal Devaluation in a Monetary Union" by Philipp Engler, Giovanni Ganelli, Juha Tervala and Simon Voigts, January 2014.

012 "Nonparametric Estimates for Conditional Quantiles of Time Series" by Jürgen Franke, Peter Mwita and Weining Wang, January 2014.

013 "Product Market Deregulation and Employment Outcomes: Evidence from the German Retail Sector" by Charlotte Senftleben-König, January 2014.

014 "Estimation procedures for exchangeable Marshall copulas with hydrological application" by Fabrizio Durante and Ostap Okhrin, January 2014.

015 "Ladislaus von Bortkiewicz - statistician, economist, and a European intellectual" by Wolfgang Karl Härdle and Annette B. Vogt, February 2014.

016 "An Application of Principal Component Analysis on Multivariate Time-Stationary Spatio-Temporal Data" by Stephan Stahlschmidt, Wolfgang Karl Härdle and Helmut Thome, February 2014.

017 "The composition of government spending and the multiplier at the Zero Lower Bound" by Julien Albertini, Arthur Poirier and Jordan Roulleau-Pasdeloup, February 2014.

018 "Interacting Product and Labor Market Regulation and the Impact of Immigration on Native Wages" by Susanne Prantl and Alexandra Spitz-Oener, February 2014.

SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

SFB 649 Discussion Paper Series 2014

040 "Localising Forward Intensities for Multiperiod Corporate Default" by Dedy Dwi Prastyo and Wolfgang Karl Härdle, August 2014.

041 "Certification and Market Transparency" by Konrad Stahl and Roland Strausz, September 2014.

042 "Beyond dimension two: A test for higher-order tail risk" by Carsten Bormann, Melanie Schienle and Julia Schaumburg, September 2014.

043 "Semiparametric Estimation with Generated Covariates" by Enno Mammen, Christoph Rothe and Melanie Schienle, September 2014.

044 "On the Timing of Climate Agreements" by Robert C. Schmidt and Roland Strausz, September 2014.

045 "Optimal Sales Contracts with Withdrawal Rights" by Daniel Krähmer and Roland Strausz, September 2014.

046 "Ex post information rents in sequential screening" by Daniel Krähmer and Roland Strausz, September 2014.

047 "Similarities and Differences between U.S. and German Regulation of the Use of Derivatives and Leverage by Mutual Funds – What Can Regulators Learn from Each Other?" by Dominika Paula Gałkiewicz, September 2014.

048 "That's how we roll: an experiment on rollover risk" by Ciril Bosch-Rosa, September 2014.

049 "Comparing Solution Methods for DSGE Models with Labor Market Search" by Hong Lan, September 2014.

050 "Volatility Modelling of CO_2 Emission Allowance Spot Prices with Regime-Switching GARCH Models" by Thijs Benschop, Brenda López Cabrera, September 2014.

051 "Corporate Cash Hoarding in a Model with Liquidity Constraints" by Falk Mazelis, September 2014.

052 "Designing an Index for Assessing Wind Energy Potential" by Matthias Ritter, Zhiwei Shen, Brenda López Cabrera, Martin Odening, Lars Deckert, September 2014.

053 "Improved Volatility Estimation Based On Limit Order Books" by Markus Bibinger, Moritz Jirak, Markus Reiss, September 2014.

054 " Strategic Complementarities and Nominal Rigidities" by Philipp König, Alexander Meyer-Gohde, October 2014.

055 "Estimating the Spot Covariation of Asset Prices – Statistical Theory and Empirical Evidence" by Markus Bibinger, Markus Reiss, Nikolaus Hautsch, Peter Malec, October 2014.

056 "Monetary Policy Effects on Financial Intermediation via the Regulated and the Shadow Banking Systems" by Falk Mazelis, October 2014.

057 "A Tale of Two Tails: Preferences of neutral third-parties in three-player ultimatum games" by Ciril Bosch-Rosa, October 2014.

058 "Boiling the frog optimally: an experiment on survivor curve shapes and internet revenue" by Christina Aperjis, Ciril Bosch-Rosa, Daniel Friedman, Bernardo A. Huberman, October 2014.

058 "Expectile Treatment Effects: An efficient alternative to compute the distribution of treatment effects" by Stephan Stahlschmidt, Matthias Eckardt, Wolfgang K. Härdle, October 2014.