

# Causal Analysis at Extreme Quantiles with Application to London Traffic Flow Data

Prajamitra Bhuyan<sup>1§</sup>, Kaushik Jana<sup>1§</sup>, Emma J. McCoy<sup>\*,†</sup>

<sup>§</sup>Indian Institute of Management, Calcutta, India

<sup>§</sup> Ahmedabad University, India

<sup>\*</sup>London School of Economics and Political Science, United Kingdom

<sup>†</sup>The Alan Turing Institute, United Kingdom

## Abstract

Transport engineers employ various interventions to enhance traffic-network performance. Quantifying the impacts of Cycle Superhighways is complicated due to the non-random assignment of such an intervention over the transport network. Treatment effects on asymmetric and heavy-tailed distributions are better reflected at extreme tails rather than at the median. We propose a novel method to estimate the treatment effect at extreme tails incorporating heavy-tailed features in the outcome distribution. The analysis of London transport data using the proposed method indicates that the extreme traffic flow increased substantially after Cycle Superhighways came into operation.

*Key words: Causality; Extreme value analysis; Heavy-tailed distribution; Potential outcome; Quantile regression; Transport engineering*

## 1 Introduction

In the last couple of decades, metropolitan areas in both the developed and the developing countries have been affected by increasing traffic congestion mainly due to population explosion. Urban

---

<sup>1</sup>The authors contribute equally to this paper.

population in developing nations is projected to grow, adding 2.5 billion people by 2050 ([UN Environment Programme, 2019](#)). In addition to the negative impacts on mobility and air quality, previous studies indicate that severe congestion has a negative impact on GDP, the city’s economic competitiveness ([Slawson, 2017](#); [Jin and Rafferty, 2017](#)), and road safety, thus imparting socio-economic distress ([Retallack and Ostendorf, 2019](#)). Intelligent transportation systems can revolutionize traffic-mobility management and offer an integrated approach to infrastructure development ([Xian et al., 2021](#)). Transport engineers employ network interventions (i.e. treatment) to control high-consequence traffic events. Often such interventions fail to achieve the intended objective, and as a consequence transport networks perform poorly concerning traffic flow, capacity utilisation, safety, and economic and environmental impacts. The underlying reason for such failure could be that the interventions are often made without statistical evidence to guarantee the intended outcomes. In general, the interventions are typically targeted to address specific network problems, and the challenge is that their effect is ‘confounded’ if the treated and control units differ systematically with respect to several characteristics which may also affect the outcome of interest.

In the last decade, cycling has been promoted as an affordable, clean and environment friendly sustainable means of transportation around the world. The bicycle contributes to cleaner air and less congestion that promotes economic growth and reduces inequalities while bolstering the fight against climate change to achieve sustainable development goals ([United Nations, 2021](#)). A recent study on Traffic Scorecard released in 2021 ranks London as the world’s most congested city out of more than 1,000 global cities ([Reid, 2021](#)). The Mayor of London aims for cycling journeys in London to increase from 2% of all journeys in 2001 to 5% by 2026. To promote cycling activity, several policy decisions including Cycle Superhighways (CS), Santander Cycles and Biking Boroughs have already been implemented ([Transport for London, 2010](#)). The CS are 1.5-meter wide barrier-free cycling paths connecting outer London to central London to provide adequate spatial capacity for existing cyclists and potential future commuters who adopt cycling as a mode of transport (see [Figure 1](#)). Twelve Cycle Superhighways were announced in 2008 with the aim of enabling faster and safer cycle journeys. As displayed in [Figure 2](#), these routes were designed in a clock faced layout to radiate from the city center towards greater London. Currently, only four routes are in operation, namely CS2 (Stratford to Aldgate); CS3 (Barking to Tower Gateway); CS7 (Merton to the City); and CS8 (Wandsworth to Westminster). In the very first year, cycling has increased by 83% along CS3 and 46% along CS7 after their inauguration in 2010. The effects of CS on heavy congestion are not evaluated in the report by [Transport for London \(2011\)](#) due to the lack of adequate data

and methodological framework for analysis. However, some contradictory reports about the effects of CS on road traffic congestion have been published in print and electronic media ([Norman, 2017](#); [Blunden, 2016](#)). Due to the intricate nature of the transport network, the quantification of the effects of CS on high consequence events like extreme traffic congestion is a complex problem. Moreover, various traffic and socioeconomic characteristics may act as confounders, affecting both the traffic flow and intervention simultaneously, therefore it is important to study the association of those factors with the interventions and outcome of interest.



Figure 1: Examples of London Cycle Superhighways

The need to develop statistical methods to study causal quantities is important due to their wide application and impact in various fields of science, e.g., medical science ([Imbens and Rubin, 2015](#)), transport engineering ([Li et al., 2017](#); [Zhang et al., 2021](#)), climate science ([Hannart and Naveau, 2018](#); [Naveau et al., 2020](#)) and public policy ([Gangl, 2010](#); [Freedman, 2010](#)). Most commonly, the literature on causal inference is focused on population means of potential outcomes. In many real scenarios, asymmetric and heavy-tailed distributions are frequently encountered and treatment effects are often better summarized in tail quantiles rather than in averages. There is a growing literature that focuses directly on the estimation of casual effects on quantiles. [Firpo \(2007\)](#) proposed estimating the quantiles by minimizing an inverse probability weighted check loss function, which achieves non-parametric consistency by means of a propensity score estimated as a logistic

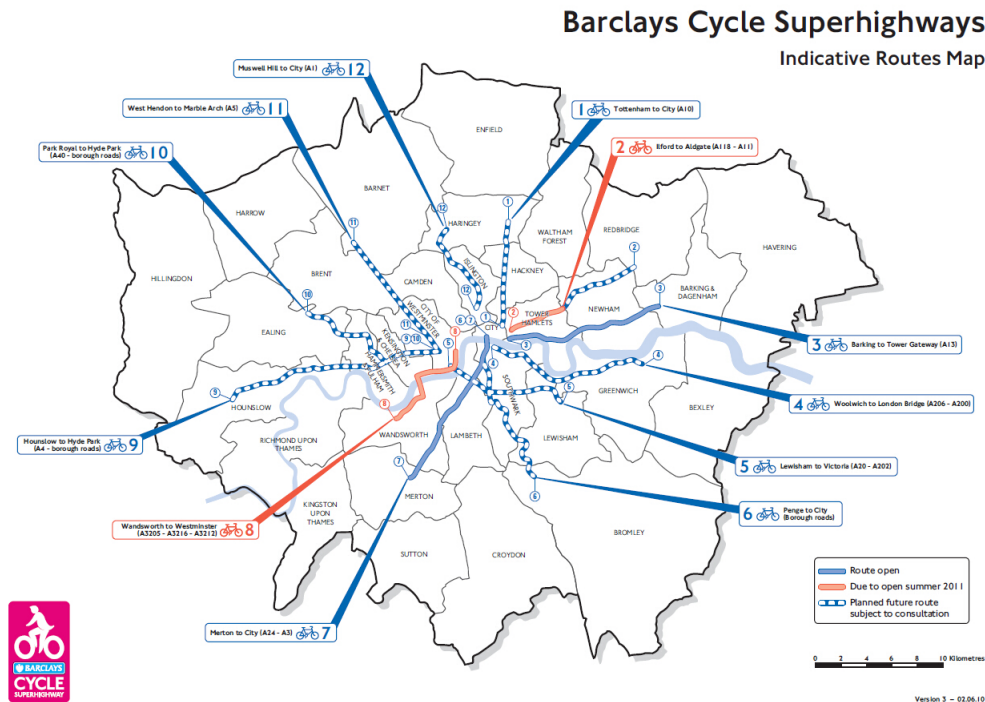


Figure 2: Route map of London Cycle Superhighways

power series with degree increases asymptotically. [Melly \(2006\)](#), and [Frölich and Melly \(2013\)](#); [Chernozhukov et al. \(2013\)](#) considered estimation of the quantiles under a linear parametric model for the distribution function, and quantile functions, respectively. [Zhang et al. \(2012\)](#) developed several methods for estimation of causal effects on quantiles which are analogous to the methods employed for the average causal effect. [Díaz \(2017\)](#) devised a doubly robust estimator based on a semiparametric approach using a targeted maximum likelihood estimator for the quantiles. In this context, [Xu et al. \(2018\)](#) introduced a Bayesian nonparametric approach through Bayesian additive regression trees to model the propensity score and a Dirichlet process mixture of normal models to construct the distribution of potential outcomes.

Estimation of extreme quantiles (or equivalently, so-called return levels) of univariate and conditional distributions is an important problem in various application areas, including meteorology, hydrology, climatology and finance ([Embrechts et al., 1997](#)). Extreme value theory provides tools to estimate the probability of extreme events based on the assumption that the underlying distribution of the normalized random variable resides in the domain of attraction of extreme value distribution ([Fisher and Tippett, 1928](#)). Thus extreme quantiles can be predicted by estimating

the parameters of the corresponding extreme value distribution and the normalizing constants. An alternative approach is to characterise the distribution with its tail index and utilise the same for estimating an extreme quantile. For a detailed review on different models and methods, the readers are referred to the classical books on extreme value inference by [Embrechts et al. \(1997\)](#) and [Coles \(2001\)](#), and the review paper by [Gomes and Guillou \(2015\)](#). In the context of causal inference, [Zhang \(2018\)](#) considered the estimator proposed by [Firpo \(2007\)](#) and developed asymptotic theory for the causal estimator for intermediate and extreme quantiles. Very little work exists in the literature relating causality and the occurrence of extreme events. Recently [Gnecco et al. \(2019\)](#) and [Engelke \(2020\)](#) combined causal inference with extreme value theory to characterise causal tail dependence of two random variables in the context of a directed acyclic graphical model. [Gissibl and Klüppelberg \(2018\)](#); [Gissibl et al. \(2017\)](#) studied the causal dependence structure through max-linear models on acyclic graphs. For a hydrological application, [Mhalla et al. \(2019\)](#) studied the causal relationship between two extreme random variables for modelling daily measurements of river discharges corresponding to riverflow-connected stations. This modelling approach uses a copula structure to estimate the dependence between two random variables with the restriction that there are no confounders (common set of covariates affecting treatment and response simultaneously) in the model. [Deuber et al. \(2021\)](#) developed a propensity score based estimator of extremal quantile treatment effect. This method relies on asymptotic tail approximations through the Hill estimator for the extreme value indices of potential outcome distributions. Recently, there has been a focus on studying the causal links between the climate system and climate change (attribution) and this causal link has been used to design event attribution; see [Hannart and Naveau \(2018\)](#) and [Ribes et al. \(2020\)](#) for more details.

## 1.1 London transport-network data

The current study is motivated by the problem of estimation of the causal effect of London Cycle Superhighways on extreme traffic congestion based on a data set over the period 2007-2014. The data consists of 75 treated zones and 375 control zones which were selected through stratified random sampling along the 40 km long main corridors radiating from central London to outer London. In observational data, the effect of CS is confounded due to various factors related to traffic dynamics, road characteristics, and socio-demographic conditions. Traffic data on both the major and minor road network are routinely collected by the Department for Transport, Government of the UK. Additional data on traffic flow and speed are collected from the London Atmospheric Emissions

Inventory. It has been observed that traffic congestion is associated with bus-stop density and road network density (George, 1970). An association between traffic congestion and socio-demographic characteristics, such as employment and land-use patterns, has also been indicated in previous studies (Badoe and Miller, 2000; Chen et al., 2001; Zhang et al., 2017; Bhuyan et al., 2021). A detailed description of the above data sets can be found in Li et al. (2017). To incorporate these socio-demographic effects, we obtained relevant data on population and employment density, as well as the information of land-use patterns from the Office for National Statistics. The data that are available from the aforementioned sources and the rationale to construct the response and covariates are described in Section 5.

Recent literature on transport engineering emphasized the importance of developing inferential tools to predict extreme consequences of interventions or shocks to the system, enabling one to identify the most effective available interventions to mitigate the consequences of such events (Zheng and Sayed, 2019; Xu and Nusholtz, 2017; FarahaCarlos and Azevedo, 2017). As mentioned above, the motivating data on traffic system arises from a study of various interventions targeted at traffic conditions in the city of London. A vital task for the policy makers is to quantify the causal effects of these interventions on the extreme traffic flow and investigate its associated consequences. The outcome variable (traffic flow) under consideration exhibits heavy tails as there are small but non-trivial number of locations with very high traffic volume. Heavy-tailed distributions are often characterized by very large variance. For such cases, standard error of the existing estimators of the causal quantities are also very high and subsequently precludes statistical significance at most plausible sample sizes (Díaz, 2017). In this work, we address this issue and propose a method of estimation of the causal effect at the extreme tails by combining tools of causal inference and extreme value theory. The method is applied to analyse the motivating data from the transport engineering application outlined above.

The outline of the article is as follows, in Section 2 we describe the existing models and methods used in the traditional potential outcome framework. In Section 3, we propose a new method to estimate the causal effects at extreme quantiles. Simulation studies are performed to assess the effectiveness of the proposed methods and the results are summarized in Section 4. In Section 5, we discuss the results obtained by analysing the London Cycle Superhighways data. We summarise the key findings and conclude with a discussion of future research in Section 6.



## 2 Potential outcome framework

In usual causal inference problems, the data available for estimation of the causal effects are realisations of a random vector,  $Z_i = (Y_i, D_i, \mathbf{X}_i)$ ,  $i = 1, \dots, n$ , where for the  $i$ th unit of observation,  $Y_i$  denotes the response,  $D_i$  the treatment (intervention or exposure) received, and  $\mathbf{X}_i$  a vector of confounders or covariates. Let the triplet  $(Y, D, \mathbf{X})$  represents a generic form of  $(Y_i, D_i, \mathbf{X}_i)$ ,  $i = 1, \dots, n$ . The treatment can be binary, multi-valued or continuous but essentially it is not assigned randomly. Therefore, the simple comparisons of average responses across different treatment groups will not in general reveal a ‘causal’ effect (causation) due to potential confounding. Confounding can be addressed if the vector of covariates  $\mathbf{X}$  is sufficient to ensure unconfoundedness, or conditional independence of potential outcomes and the treatment. In the context of binary treatment, the conditional independence assumption requires that  $(Y(0), Y(1)) \perp\!\!\!\perp D | \mathbf{X}$ , where  $D$  is the indicator function for receiving the treatment, and  $Y(1)$  and  $Y(0)$  indicate potential outcomes under treated or control status, respectively. That is, given the information contained in the covariate vector  $\mathbf{X}$ , the assigned treatment carries no additional information on the individual’s potential outcomes. This assumption is also referred to as conditional exchangeability or strong ignorability in the literature ([Rosenbaum and Rubin, 1983](#)). An additional requirement for valid causal inference is positivity or overlap assumption: the conditional probability of treatment assignment is strictly positive for every combination of values of the covariates, i.e.  $0 < \mathbb{P}[D = 1 | \mathbf{X} = \mathbf{x}] < 1$  for all  $\mathbf{x}$ . This assumption implies that there exists no strata of  $\mathbf{X}$  such that treatment assignment is uniquely determined. To illustrate this point, suppose that for a given value of  $\mathbf{X}$ , say  $\mathbf{x}^*$ ,  $\mathbb{P}[D = 0 | \mathbf{X} = \mathbf{x}^*] = 1$ . Then there will be no triplets of observed data  $(Y, D, \mathbf{X})$  such that  $D = 1$  and  $\mathbf{X} = \mathbf{x}^*$ ; thus, any estimates of treatment effects, say a conditional average treatment effect such as  $\mathbb{E}[Y | D = 1, \mathbf{X} = \mathbf{x}^*] - \mathbb{E}[Y | D = 0, \mathbf{X} = \mathbf{x}^*]$  would require extrapolation. The readers are referred to [Moodie and Stephens \(2022\)](#) for more details. Conventionally, the main interest is in estimating the average treatment effect  $\mu = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$ , which measures the difference in average outcomes under treatment and control status.

In practice, distributional features are often summarized using quantiles when asymmetric distribution for the response variable is encountered. In case of a skewed distribution, the median is a more appropriate measure of location than the mean, and statistics based on quantiles could be more meaningful measures of the spread of the distribution than the standard deviation. In many applications, practitioners are interested in the tails of the distribution (say higher than 95th per-

centiles) of the potential outcomes (Zhang et al., 2012; Díaz, 2017). In the presence of confounding, a simple comparison of the treated and untreated individuals in terms of empirical quantiles would not have a causal interpretation. In such situation, the causal effects of treatment on the target  $\tau_0$ -th ( $0 < \tau_0 < 1$ ) quantile is defined as

$$\eta_{\tau_0} = q_{Y(1)}(\tau_0) - q_{Y(0)}(\tau_0), \quad (1)$$

where  $q_{Y(t)}$  denotes the quantile associated with the distribution of potential outcome  $Y(t)$  for  $t = 0, 1$  (Zhang et al., 2012). In many contexts, it is also of interest to understand the quantile effect of applying versus withholding the intervention only for the treated units (rather than for all the units in the population). This causal quantity is called the quantile treatment effect on the treated and is defined as

$$\zeta_{\tau_0} = q_{Y(1)|D=1}(\tau_0) - q_{Y(0)|D=1}(\tau_0), \quad (2)$$

where  $q_{Y(t)|D=1}$  denotes the quantile associated with the distribution of the potential outcome  $Y(t)$  given  $D = 1$  for  $t = 0, 1$  (Zhang et al., 2012). When there exists no heterogeneity of the treatment effect induced by the covariates (i.e., no interactions effects of the treatment and the confounders), the quantile treatment effect on the population and the quantile treatment effect on the treated are identical provided the effects are additive. However, when treatment heterogeneity exists, these two treatment effects differ (Moodie et al., 2018).

Causal effects on quantiles can be estimated using quantile regression with appropriate adjustment for confounders. However, it should be noted that the quantile estimates from a quantile regression model are conditional on all the covariates in the model. When one is interested in estimating the marginal quantile at a given probability level (without conditioning on covariates), it has to be determined by all conditional quantiles. Zhang et al. (2012) studied the problem of estimating the causal quantities, defined in (1) and (2), under parametric assumption on the distribution of the outcome variable. In light of the conditional exchangeability assumption, the distribution of the potential outcome  $Y(t)$ , for  $t = 0, 1$ , (i.e. counterfactual distribution) is given by

$$\begin{aligned} F_t(y) &= \mathbb{E}[P(Y(t) \leq y | \mathbf{X})] \\ &= \mathbb{E}[P(Y(t) \leq y | D = t, \mathbf{X})] \\ &= \mathbb{E}[P(Y \leq y | D = t, \mathbf{X})] \\ &= \int G_{t,x}(y) dH(\mathbf{x}), \end{aligned} \quad (3)$$



where  $G_{t,x}(y) = P(Y \leq y | D = t, \mathbf{X} = \mathbf{x})$  for  $y \in \mathbb{R}$ , and  $H$  denotes the marginal distribution of  $\mathbf{X}$  for  $t = 0, 1$ . Similarly, the distribution of the potential outcome  $Y(t)$  conditional on  $D = 1$ , is given by

$$F_{t|1}(y) = E\{P(Y(t) \leq y | \mathbf{X}) | D = 1\} = \int G_{t,x}(y) dH_1(\mathbf{x}), \quad (4)$$

where  $H_1$  denotes the marginal distribution of  $\mathbf{X}$  given  $D = 1$  for  $t = 0, 1$ .

Zhang et al. (2012) estimated the counterfactual distribution, given in (3), by

$$\hat{F}_t(y) = n^{-1} \sum_{i=1}^n \hat{G}_{t,\mathbf{x}_i}(y), \quad (5)$$

where  $\hat{G}$  is an estimate of  $G$  under the assumption that a normal linear model holds after a Box-Cox transformation. Finally the target  $\tau_0$ -th quantile  $q_{Y(t)}(\tau_0)$  is estimated by inverting the estimated unconditional distribution function  $\hat{F}_t(y)$  and substituting it into the expression given in (1) to obtain the estimate of the quantile treatment effect for the population. This is known as outcome regression (OR) based estimate. Similarly, one can estimate  $F_{t|1}(\cdot)$ , given in (4), by

$$\hat{F}_{t|1}(y) = \frac{1}{\sum_{i=1}^n D_i} \sum_{i=1}^n \hat{G}_{t,\mathbf{x}_i}(y) D_i. \quad (6)$$

One can then invert  $\hat{F}_{t|1}(y)$  to obtain estimates of  $q_{Y(t)|D=1}(\tau_0)$  and the quantile treatment effect for the treated  $\zeta_{\tau_0}$  defined in (2). An alternative set of estimators of the counterfactual distributions proposed by Zhang et al. (2012) are based on inverse propensity weighting (IPW) and matching techniques. Firpo (2007) proposed quantile regression approach for estimation of counterfactual quantiles using inverse propensity score as weights.

### 3 Proposed methodology

Note that the conditional distribution of  $Y(t)$  given  $\mathbf{X} = \mathbf{x}$  can be obtained from the conditional quantile function  $Q_{Y|t,\mathbf{x}}(\tau) = G_{t,\mathbf{x}}^{-1}(\tau)$  of the outcome  $Y$  for all  $\tau \in (0, 1)$  (Chernozhukov et al., 2013). Let  $U \sim \text{Unif}(0, 1)$ . Then we can write

$$\begin{aligned} G_{t,\mathbf{x}}(y) &= P[U \leq G_{t,\mathbf{x}}(y)] \\ &= P[Q_{Y|t,\mathbf{x}}(U) \leq y] \\ &= E[\mathbb{I}[Q_{Y|t,\mathbf{x}}(U) \leq y]] \\ &= \int_0^1 \mathbb{I}[Q_{Y|t,\mathbf{x}}(\tau) \leq y] d\tau, \end{aligned} \quad (7)$$

where  $\mathbb{I}[\cdot]$  is an indicator function. When the context suggests the presence of a heavy-tailed nature in the data generating process and one is interested in the extreme tails of the distribution  $F_t$  (i.e. probability levels are close to 0 or 1), it is more appropriate to model  $G_{t,\mathbf{x}}$  with some heavy-tailed components. To model the bulk and the extreme component of  $G_{t,\mathbf{x}}(y)$  efficiently, we choose a suitable transition probability level  $\tau_u \in (0, 1)$  and decompose (7) as:

$$G_{t,\mathbf{x}}(y) = \int_0^{\tau_u} \mathbb{I}[Q_{Y|t,\mathbf{x}}(\tau) \leq y] d\tau + \int_{\tau_u}^1 \mathbb{I}[Q_{Y|t,\mathbf{x}}(\tau) \leq y] d\tau. \quad (8)$$

The transition point  $\tau_u$  is assumed to represent the probability level of quantile above which the distribution (second part of (8)) exhibits heavy-tailed nature. At the same time a very high probability level as a choice of  $\tau_u$ , may prevent enough number of observations being generated from the extreme part of the distribution. In practice,  $\tau_u$  needs to be estimated from the data by balancing the above two issues. The next section provides a detailed outline of a data driven and efficient error control approach to determine the probability level  $\tau_u$ . Note that, any estimation of the unknown population quantile process involves the evaluation of the quantile process at finite grid points. Thus the integral in (8) needs to be evaluated at discrete points. A common requirement to reliably estimate the quantile process is that the mesh width of grids should converge to 0 with a rate  $1/\sqrt{n}$  or faster (Chernozhukov et al., 2020).

We propose an estimate of  $G_{t,\mathbf{x}}(\cdot)$  as

$$\hat{G}_{t,\mathbf{x}}(y) = \sum_{j=1}^u (\tau_j - \tau_{(j-1)}) \mathbb{I}[\hat{Q}_{Y|t,\mathbf{x}}(\tau_j) \leq y] + \sum_{j=u+1}^J (\tau_j - \tau_{(j-1)}) \mathbb{I}[\tilde{Q}_{Y|t,\mathbf{x}}(\tau_j) \leq y], \quad (9)$$

where the integer  $u$  is determined from  $\tau_u$ , and  $\hat{Q}_{Y|t,\mathbf{x}}(\tau_j)$  and  $\tilde{Q}_{Y|t,\mathbf{x}}(\tau_j)$  are estimates of  $Q_{Y|t,\mathbf{x}}(\tau_j)$  based on simultaneous quantile regression and parametric extreme value modelling, respectively, such that the issue of quantile crossing (within each of the summands) does not arise. The details of the above estimation methods are discussed in Subsections 3.1-3.2. The estimated conditional distribution function is then plugged into (5) to obtain the estimate of counterfactual distributions  $\hat{F}_t(y)$  for  $t = 0, 1$ . Then  $\hat{F}_t(y)$  is inverted to find the quantile  $\hat{q}_{Y(t)}(\tau_0) = \inf\{z : \hat{F}_t(z) \geq \tau_0\}$  for  $t = 0, 1$ , and finally the estimate of quantile causal effect  $\eta_{\tau_0}$  for the population is given by

$$\hat{\eta}_{\tau_0} = \hat{q}_{Y(1)}(\tau_0) - \hat{q}_{Y(0)}(\tau_0).$$

Similarly, we obtain  $\hat{q}_{Y(t)|D=1}(\tau_0) = \inf\{z : \hat{F}_{t|1}(z|D=1) \geq \tau_0\}$  from (6), and the estimate of  $\zeta_{\tau_0}$  is given by

$$\hat{\zeta}_{\tau_0} = \hat{q}_{Y(1)|D=1}(\tau_0) - \hat{q}_{Y(0)|D=1}(\tau_0).$$

### 3.1 Selection of transition point

The value of  $\tau_u$ th quantile in (8) determines the transition point of the outcome distribution, below which the quantile regression model is fitted to the bulk part of the distribution, and above which a heavy-tailed distribution is fitted. Let  $Y$  be a random variable with distribution function  $\Psi$  with right end point  $y_\Psi$ . Then by the Pickands-Balkema-de Haan Theorem (Balkema and de Haan, 1974; Pickands, 1975), the conditional distribution of  $Y - U$  above an excess  $U < y_\Psi$  is well approximated by the generalized Pareto distribution (GPD) for large  $U$  under some general regularity conditions. That is,

$$\lim_{y \uparrow y_\Psi} P[Y - U \leq y | Y > U] \rightarrow H(y),$$

and  $H$  is the distribution function of GPD, given as:

$$H(y) = \begin{cases} 1 - [1 + \xi (\frac{y}{\sigma})]^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp[-\frac{y}{\sigma}], & \text{if } \xi = 0, \end{cases}$$

where  $y \geq 0$  if  $\xi \geq 0$ , and  $0 \leq y \leq -\sigma/\xi$  if  $\xi < 0$ . Here  $\sigma$  and  $\xi$  are the scale and shape parameters, respectively. The value of the shape parameter  $\xi = 0$ , interpreted as  $\xi \rightarrow 0$ , leads  $H$  to be the exponential distribution with mean  $\sigma$ . The negative values of  $\xi$  imply the right tail of  $H$  is short and light, and it has a finite right end point. In particular, for  $\xi = -1$ ,  $H$  is a uniform distribution over  $[0, \sigma]$ . When  $\xi > 0$ , the higher the value of  $\xi$  the heavier the tail of the distributions. Choice of the threshold and fitting of a GPD to the exceedences of the threshold is a widely studied problem in the extreme value analysis literature. Too low and too high value of the threshold should generally be avoided (Coles, 2001). A very high threshold leaves fewer exceedences to fit the GPD model and would yield estimates with high variance, and a too low threshold leads to an inadequate fit to the GPD model and may produce biased estimates. In the light of the decomposition of the estimate  $\hat{G}_{t,x}$ , provided in (9), the threshold requires to be a suitable quantile of the conditional distribution of the outcome. Many threshold selection procedures are available in the literature; see Scarrott and MacDonald (2012) for a review.

There are broadly four categories of approaches for threshold selection. The first, and most common approach is based on a graphical diagnosis method (Davison and Smith, 1990; Drees et al., 2000; Coles, 2001). The second approach involves methods that minimize the asymptotic mean-squared error of the estimators of the GPD parameters or of the extreme quantiles, under particular assumptions on the upper tail of the GPD (Hall, 1990; Beirlant et al., 2004; Langousis et al., 2016).

The third category of methods are based on the goodness-of-fit (gof) test of the GPD (Davison and Smith, 1990; Bader et al., 2018), where the threshold is selected at a level to ensure that the GPD provides an adequate fit to the exceedances. The fourth method is automatic threshold selection method of Bader et al. (2018) for choosing a suitable threshold. Most of the above methods can incorporate additional covariate information in different levels of the modelling. See Davison and Smith (1990) and Beirlant et al. (2005) for an extensive review of covariate dependent GPD modelling.

### 3.1.1 Two-step approach for selection of $\tau_u$

To select the transition point  $\tau_u$  in (7), we develop a two-step approach. In the first step we use covariates to generate multiple competing and ordered thresholds through constraint quantile regression (discussed in the following Section 3.2). And in the second step, we adopt the automatic threshold selection method of Bader et al. (2018) controlling false discovery rate associated with multiple goodness-of-fit tests. Specifically, in the first step, for a given number  $l$ , we consider  $l$  competing thresholds as conditional quantiles obtained by quantile regression (Koenker, 2005) and denote the fitted quantiles as  $U(\tau|\mathbf{w}) = \mathbf{w}^T \hat{\boldsymbol{\beta}}(\tau)$ , where  $\mathbf{v} = (t, \mathbf{x})$  and  $\mathbf{w} = (1, \mathbf{v})^T$  is a covariate vector of dimension  $(m+1)$  including 1, and  $\hat{\boldsymbol{\beta}}(\tau) = (\hat{\beta}_0(\tau), \dots, \hat{\beta}_m(\tau))^T$  are the estimates of the regression parameter vector at  $\tau = \tau_1, \dots, \tau_l$ . Therefore, for a given set of  $l$  probability levels, the method produces thresholds  $U(\tau_1|\mathbf{w}) \leq \dots \leq U(\tau_l|\mathbf{w})$ . We then use each of the thresholds, and fit a GPD to the exceedances above the threshold. A different GPD is obtained for each of the covariate adjusted threshold. We use maximum likelihood method to estimate the unknown GPD parameters. Let us assume that there are  $n_k$  exceedances corresponding to the  $k$ th threshold, for  $k = 1, \dots, l$ . Then the sequence of null hypotheses to be tested are  $H_0^{(k)}$ : the distribution of  $n_k$  observations above  $U(\tau_k|\mathbf{w})$  follows the GPD.

Bader et al. (2018) adopted and improved the ForwardStop rule of G'Sell et al. (2015) for the sequential testing (ordered) of null hypotheses  $H_0^{(1)}, \dots, H_0^{(l)}$  when the tests are dependent. The method uses the  $p$ -values corresponding to the tests of  $l$  hypotheses as  $p_1, \dots, p_l \in [0, 1]$ . For testing multiple hypothesis with an ordered nature, G'Sell et al. (2015) proposed transforming the sequence of  $p$ -values to a monotone sequence and then applying the original method of Benjamini and Hochberg (1995) on the transformed sequence. The rejection rule is constructed by returning a cutoff  $\hat{k}$  such that  $H_0^{(1)}, \dots, H_0^{(\hat{k})}$  are rejected. They prescribe that if no  $\hat{k} \in 1, \dots, l$  exists, then no

rejection is made. ForwardStop is given by

$$\hat{k} = \max \left\{ k \in \{1, \dots, l\} : -\frac{1}{k} \sum_{i=1}^k \log(1 - p_i) \leq \lambda \right\},$$

where  $\lambda$  is a pre-specified level. In this sequential testing, stopping at  $k$  implies that the goodness-of-fit test of the GPD to the exceedances at the first  $k$  thresholds  $U(\tau_l|\mathbf{w}), \dots, U(\tau_k|\mathbf{w})$  are rejected. In other words, the set of first  $k$  null hypotheses  $H_0^{(1)}, \dots, H_0^{(k)}$  are rejected. Let us denote  $U(\tau_u|\mathbf{w})$  as the GPD threshold corresponding to the selected probability level  $\tau_u = \tau_{\hat{k}}$ .

### 3.2 Estimation of quantiles in the bulk part

After the selection of the transition point between the bulk and the extreme part of the distribution, we need to estimate the conditional quantiles involved in both the summands of (8). The first part requires estimation of multiple conditional quantiles at different probability levels. One can estimate these quantiles separately, but this approach has two major drawbacks. Firstly, it could be possible that multiple estimated quantiles cross each other even if their population versions never cross. For example, it may be possible that the estimated 90th percentile has a larger value than the estimated 95th percentile. This results in an invalid estimate of the corresponding distribution function. Secondly, it fails to incorporate common characteristic among the consecutive quantiles (Zou and Yuan, 2008). Simultaneous estimation borrows information from all the neighbouring quantiles that may improve the efficiency of individual quantile estimates (Liu and Wu., 2011). In their work, Schnabel and Eilers (2013) used splines to model the quantile regression functions simultaneously but, the drawback with the method is that there is no clear theoretical justification which guarantees that there will be no crossings.

Alternatively, the method of Bondell et al. (2010) ensures non-crossing for simultaneous quantile estimation (first part of (9)). Let  $\mathbf{C} \subset \mathbb{R}^m$ , be a closed convex polytope, represented as the convex hull of  $n$  points in  $m$ -dimensions. Recall that  $\mathbf{v} = (D, \mathbf{x})$  and  $\mathbf{w} = (1, \mathbf{v})^T$ , and we are interested in ensuring that the quantile curves do not cross for all values of the covariate  $\mathbf{v} \in \mathbf{C}$ . For estimating multiple quantiles at probability levels  $\tau_1 < \dots < \tau_u$ , Bondell et al. (2010) devised a simple and computationally efficient procedure using a version of constrained optimization that guarantees non-crossing of the estimated quantiles for all values of the covariate  $\mathbf{v} \in \mathbf{C}$ . This goal is achieved by imposing appropriate dominance in every consecutive pair of quantile regression planes. Under this setup, the parameters associated with the quantile function are estimated by solving the following

optimization problem:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \min_{\boldsymbol{\beta}} \sum_{j=1}^u \sum_{i=1}^n \rho_{\tau_j} (Y_i - \mathbf{w}_i^T \boldsymbol{\beta}(\tau_j)), \\ \text{subject to } &\mathbf{w}^T \boldsymbol{\beta}(\tau_{j+1}) \geq \mathbf{w}^T \boldsymbol{\beta}(\tau_j) \\ \text{for } &j = 1, \dots, u-1, \quad \forall \quad \mathbf{v} \in \mathbf{C}, \end{aligned} \quad (10)$$

where  $\rho_{\tau_j}(w) = \tau_j(1 - \mathbb{I}(w < 0))$  is the check function. Note that, the estimated conditional quantile of  $Y$  given  $D = t$  and  $\mathbf{X} = \mathbf{x}$  in the first summand of equation (9) is given by  $\hat{Q}_{Y|t,\mathbf{x}}(\tau) = \hat{\beta}_0(\tau) + t\hat{\beta}_1(\tau) + \mathbf{x}^T \hat{\boldsymbol{\beta}}^*(\tau)$ , where  $\hat{\boldsymbol{\beta}}^*(\tau)$  is the  $(m-1)$  dimensional vector of estimated coefficients corresponding to the covariate vector  $\mathbf{x}$ , for  $t = 0, 1$ ,  $\tau = \tau_1, \dots, \tau_u$ .

The proposed method involves estimation of the quantile regression at several probability levels, and it also involves a fixed number of covariates. For quantile regression with constraints and with the assumption that the covariate space is bounded by the convex polytope, the computation becomes a linear programming problem with  $u-1$  constraints (Bondell et al., 2010). Karmakar (1984) presented a polynomial-time algorithm for linear programming that requires  $O(m^{3.5}L)$  arithmetic operations on  $O(L)$  bit numbers, where  $L$  is the number of bits in the input expressed as a function of  $u$  and number of covariates  $m$ .

### 3.3 Estimation of quantiles in the extreme part

Note that the choice of transition point or threshold  $U(\tau_u|\mathbf{w})$  provided in Subsection 3.1 guarantees the suitability of GPD to model the exceedences  $Y - U(\tau_u|\mathbf{w})$ , where  $U(\tau_u|\mathbf{w}) = \hat{\beta}_0(\tau_u) + t\hat{\beta}_1(\tau_u) + \mathbf{w}^T \hat{\boldsymbol{\beta}}^*(\tau_u)$  for  $t = 0, 1$ , and  $\hat{\boldsymbol{\beta}}^*(\tau_u)$  is the estimated regression coefficient obtained from Section 3.2. Then the estimated conditional quantile of  $Y$  given  $D = t$ , and  $\mathbf{X} = \mathbf{x}$  at probability level  $\tau = \tau_{u+1}, \dots, \tau_J$ , in the second summand of (9) is given by the approximation (see Coles (2001), Section 4.3.3) :

$$\tilde{Q}_{Y|t,\mathbf{w}}(\tau) \approx \begin{cases} U(\tau_u|\mathbf{w}) + \frac{\hat{\sigma}}{\hat{\xi}} \left[ \left( \frac{\hat{\zeta}_U}{1-\tau} \right)^{\hat{\xi}} - 1 \right], & \text{if } \hat{\xi} \neq 0, \\ U(\tau_u|\mathbf{w}) + \hat{\sigma} \log \left( \frac{\hat{\zeta}_U}{1-\tau} \right), & \text{if } \hat{\xi} = 0, \end{cases} \quad (11)$$

where  $\hat{\zeta}_U = \frac{n_U}{n}$ ,  $n_U$  is the number of exceedences over the threshold  $U(\tau_u|\mathbf{w})$  for the given sample of size  $n$ , and  $\hat{\sigma}$  and  $\hat{\xi}$  are the maximum likelihood estimates of  $\sigma$  and  $\xi$ , respectively. Here, the treatment and other covariates influence the aforementioned estimate of the quantiles only through the threshold  $U(\tau_u|\mathbf{w})$ .



Note that, under the assumption of a heavy-tailed response distribution (with shape parameter  $\xi > 0$ ), the above estimated quantile  $\tilde{Q}_{Y|t,\mathbf{w}}(\tau)$  in (11) is an increasing function of  $\tau$ . This fact ensures no crossings for the fitted quantiles at extreme tails. Also, each of the quantile estimate  $\tilde{Q}_{Y|t,\mathbf{w}}(\tau)$  is larger than the threshold  $U(\tau_u|\mathbf{w})$  which is the highest quantile of the bulk part. This fact guarantees that there is no crossing at the transition point between the extreme and the bulk part.

### 3.4 Bootstrap algorithm

Zhang et al. (2012) used the standard full sample bootstrap method to compute the standard error and confidence interval of the proposed estimators (OR, IPW, Matching, etc.) for ease of implementation, avoiding the difficulty of deriving asymptotic variance. However, the validity of the bootstrap method has not been studied. Zhang et al. (2021) studied the estimator proposed by Firpo (2007), and showed that the estimator is asymptotically normal and suggested a valid full-sample bootstrap confidence interval to quantify the uncertainty. For the moderately extreme case, the paper showed that the limiting distribution of the estimator is no longer Gaussian, and proposed a  $b$ -out-of- $n$  bootstrap for valid inference. Note that, our proposed method is different from the existing methods, and it involves estimation of the conditional distributions (see Section 3). Recently, Litvinova and Silvapulle (2020) established the asymptotic validity of the full-sample bootstrap for constructing confidence intervals for high-quantiles, tail probabilities, and other tail parameters of a univariate distribution. We employ the full sample bootstrap considering the following steps to compute the standard error and confidence interval of the proposed estimator.

- Step 1. Sample  $(Y_i^*, D_i^*, \mathbf{X}_i^*)$  from the observed data  $(Y_i, D_i, \mathbf{X}_i)$ ,  $i = 1, \dots, n$ .
- Step 2. Based on the data  $(Y_i^*, D_i^*, \mathbf{X}_i^*)$ ,  $i = 1, \dots, n$ , obtain the estimates  $\hat{\eta}_{\tau_0}^*$  and  $\hat{\zeta}_{\tau_0}^*$ .
- Step 3. Replicate Step 1 to Step 2 for  $N_B$  number of times, where  $N_B$  is a large number, say 1000.
- Step 4. Estimate the bias, standard error and confidence interval of  $\hat{\eta}_{\tau_0}$  and  $\hat{\zeta}_{\tau_0}$  from  $N_B$  bootstrap estimates of  $\hat{\eta}_{\tau_0}^*$  and  $\hat{\zeta}_{\tau_0}^*$ , respectively.

## 4 Simulation study

In this section, we generate synthetic data resembling the main features of the London traffic flow data set to compare different competing methods. It has been observed that the performance of the IPW-based method are very poor when the estimated propensity score are very small (Díaz, 2017; Kang and Schafer, 2007). We seek to avoid such situations for a valid comparison. For this purpose, the parameters are chosen such that at least 25% of the observations belong to the treated group. We consider both symmetric and skewed distributions to generate observations for the covariates. One of the two confounding variables, denoted by  $X_1$  follows the normal distribution with mean 15 and standard deviation 6, and the other confounder  $X_2$  follows the exponential distribution with mean 2. We then include a non-confounding variable  $X_3$  generated from the normal distribution with both mean and standard deviation 1, and specify the following relationships between the response variable  $Y$  and a binary treatment  $D$ :

$$\text{logit}[\mathbb{E}(D|X_1, X_2, X_3)] = -3 + 0.1X_1 + 0.1X_2 + 0.2X_3$$

$$Y = 10 + 15D + X_1 + 3X_2 + 2X_1D + (1 + 4X_2 + 3D)\epsilon,$$

where error  $\epsilon$  is generated from two different symmetric distributions with mean 0. In the first case, we consider  $\epsilon$  following the normal distribution with mean 0 and standard deviation 10 and in the second case  $\epsilon$  follows the  $t$ -distribution with degrees of freedom 1. These two choices of error distribution allows us to compare the performance of the causal estimators for both light-tailed and heavy-tailed response variable. We generate each of the above data sets for two sample sizes  $n = 500$  and  $n = 1000$ .

The transition point representing the shift from the bulk to the extreme part ( $\tau_u$  in (8)) is selected by the automated threshold selection method described in Section 3.1. As a set of competing thresholds, we consider 10 quantiles with equally spaced probability levels ranging from 0.75 to 0.99. We then use quantile regression at each probability level  $\tau$  with covariates as treatment  $D$ , and three more predictors,  $X_1$ ,  $X_2$  and  $X_3$ . The fitted quantiles are then used as covariate dependent thresholds for the GPD. The sequential multiple hypothesis approach based on Anderson-Darling goodness-of-fit test of Bader et al. (2018) is then used to select the best threshold with given level  $\lambda = 0.05$ . The selected threshold is used as the transition point and the corresponding probability level is taken as the transition point  $\tau_u$ . For evaluating the quantiles involved in the bulk part, as discussed in Section 3.2, we chose probability levels with an interval of 0.01 (mesh width). This

choice of mesh width guarantees the required convergence rate to 0. We choose a total of 100 grid points in the computation of (9) with 75 and 25 equidistant points for computing the first (bulk part) and the second component (extreme part), respectively. For computing the bulk part of (9), we adopt the simultaneous quantile regression approach of Bondell et al. (2010) and use the associated R package. In the extreme part of (9), the GPD threshold is modelled with all the aforementioned covariates and the shape and scale parameters are estimated using the maximum likelihood method.

We compare the performance of the proposed estimator with the existing estimators for both population and treated subpopulation. As a set of competing estimators, we consider the targeted maximum likelihood estimate (TMLE) of Díaz (2017) and the estimator of Firpo (2007) along with the OR and IPW estimates by Zhang et al. (2012). All these methods are based on correctly specified OR and/or PS models. The competing estimators are then compared based on 1000 iterations with respect to the following statistics: the absolute relative bias (ARB), the relative variance (RV) and relative mean squared error (RMSE). More precisely, we compute the following statistics with  $L$  iterations:

$$ARB = \left| \frac{1}{L} \sum_{i=1}^L \frac{\hat{\eta}_{\tau_0}^{(i)} - \eta_{\tau_0}}{\eta_{\tau_0}} \right|, \quad RV = \frac{\sum_{i=1}^L \left[ \hat{\eta}_{\tau_0}^{(i)} - \frac{1}{L} \sum_{i=1}^L \hat{\eta}_{\tau_0}^{(i)} \right]^2}{\sum_{i=1}^L \left[ \hat{\eta}_{\tau_0}^{(i)} - \frac{1}{L} \sum_{i=1}^L \hat{\eta}_{\tau_0}^{(i)} \right]^2}, \quad RMSE = \frac{\sum_{i=1}^L \left[ \hat{\eta}_{\tau_0}^{(i)} - \eta_{\tau_0} \right]^2}{\sum_{i=1}^L \left[ \hat{\eta}_{\tau_0}^{(i)} - \eta_{\tau_0} \right]^2},$$

where  $\hat{\eta}_{\tau_0}^{(i)}$  and  $\hat{\eta}_{\tau_0}^{(i)}$  are the estimates of the population quantile effect at  $\tau_0$  based on the proposed method and a competitor in the  $i$ th iteration, respectively, with  $L = 1000$ . Similarly, we compute these metrics for the treated subpopulation (i.e., for  $\zeta_{\tau_0}$ ). The results for the quantile effects on population are summarized graphically in Figures 3-4 based on the aforementioned statistics for all the estimators under consideration. The values for all the statistics are presented on a natural log-scale. Similar comparison are performed for the quantile treatment effects on the treated subpopulation and the findings are presented in the Figures 5-6.

First, we discuss the results corresponding to quantile causal effects for the population (Figures 3-4). The ARB associated with IPW, TMLE and Firpo estimates are small for all the probability levels except at  $\tau_0 = 0.995$ . The ARB increases significantly for all the estimates at probability level  $\tau_0 = 0.995$ . In particular, the ARB of IPW and Firpo estimates are turns out to be very large compared to the other methods at probability level  $\tau_0 = 0.995$  when the data generation mechanism involves a heavy-tailed error distribution (see Figure 4). For all the probability levels, the RV and RMSE of the proposed method are smaller compared to all other methods when the data generation

mechanism involves a normal error distribution (see Figure 3). Similar patterns are also observed at  $\tau_0 = 0.95$  when the data generation mechanism involves a heavy-tailed error distribution (see Figure 4). In this setup, the RV and RMSE of all the competing estimates are exorbitantly high compared to the proposed method at  $\tau_0 = 0.995$ . Next we consider the results corresponding to quantile causal effects for the treated (see Figures 5-6). Here, the findings are very similar to the case of causal effects on the population except for a few notable differences. The performance of all estimates except the OR method at  $\tau_0 = 0.995$  are comparable with respect to the ARB when the data generation mechanism involves a normal error distribution (see Figure 5). In contrast, when the data generation mechanism involves a heavy-tailed error distribution, the ARB of the proposed method is considerably smaller compared to all other competitive methods for  $n = 500$ . The RV and RMSE of all the competitive estimates are considerably higher compared to the proposed method at probability level  $\tau_0 = 0.95$  and  $\tau_0 = 0.995$ . This pattern is more dominant when the data generation mechanism involves a heavy-tailed error distribution (see Figure 6). These findings for the above simulation study demonstrate the effectiveness of the proposed method over the existing methods for estimating the causal effect at extreme quantiles.

## 5 Analysis of London cycle superhighways effects

The primary focus of this study is to evaluate the effect of London cycle superhighways (intervention) to contain extreme congestion. The data that are available and the rationale to construct the response (Annual Average Daily Traffic) and covariates are described below.

- (a) Annual Average Daily Traffic (AADT): the total volume of vehicle traffic of a highway or a road in a year divided by 365 days. To measure AADT on individual road segments, traffic data is collected by an automated traffic counter, hiring an observer to record traffic or licensing estimated counts from GPS data providers. AADT is a simple but useful measurement to indicate busyness of a road.
- (b) Total Cycle Collisions (TCS): total number of cycle collisions causing injury based on police records from the STATS19 accident reporting form and collected by the UK Department for Transport. The location of an accident is recorded using coordinates which are in accordance with the British National Grid coordinate system. The CS routes were intended to reduce the risk of accidents for cyclists and the route allocation is possibly influenced by TCS. It is

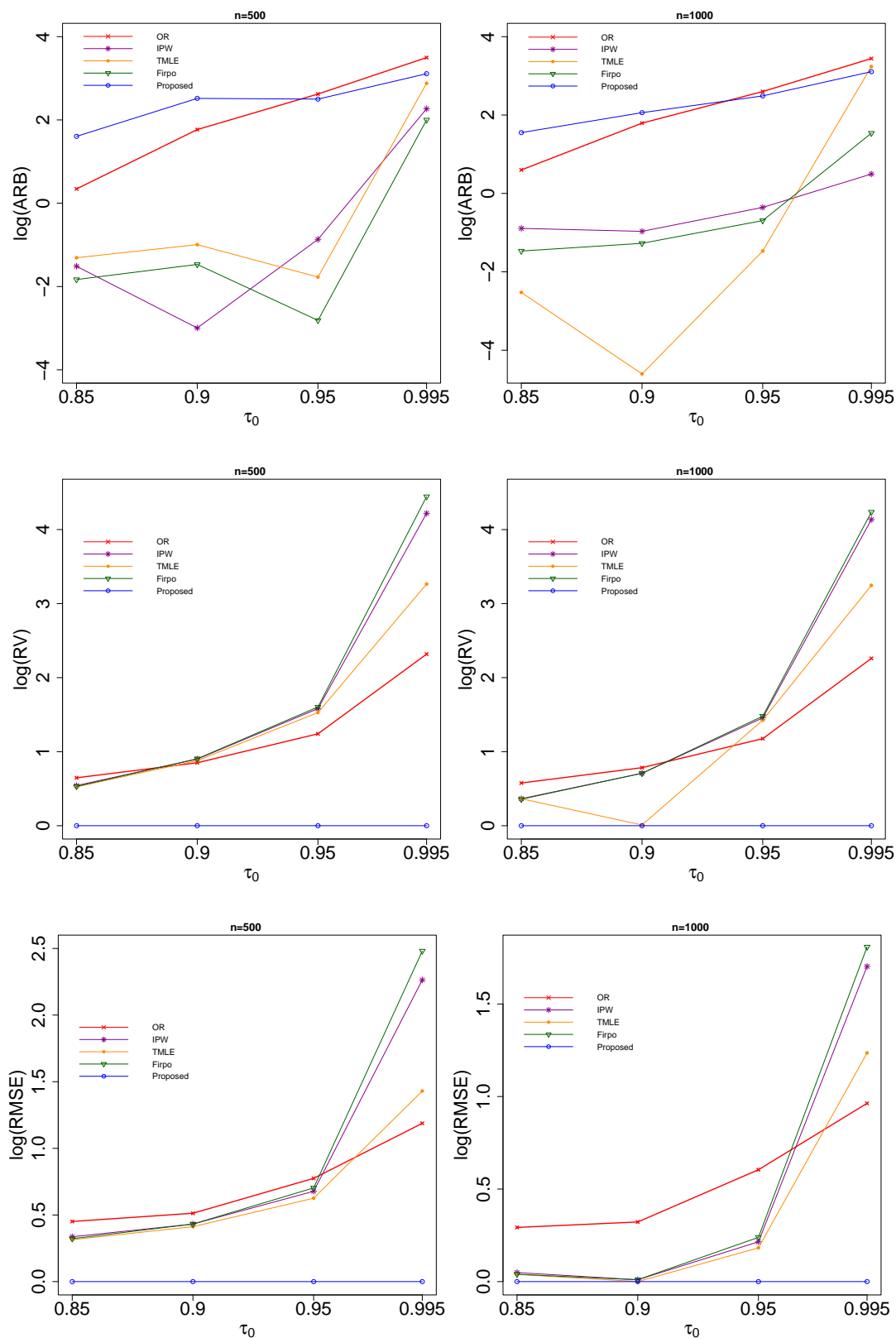


Figure 3: Comparison of proposed estimator, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the population at different probability levels based on the data generating process with Gaussian error. The top, middle and bottom rows show the comparison through three statistics: ARB, RV and RMSE, respectively, for two different sample sizes 500 and 1000. The numbers on the horizontal axis of each plot indicate the quantile levels, and five different symbols on the plot corresponds to five different estimators.

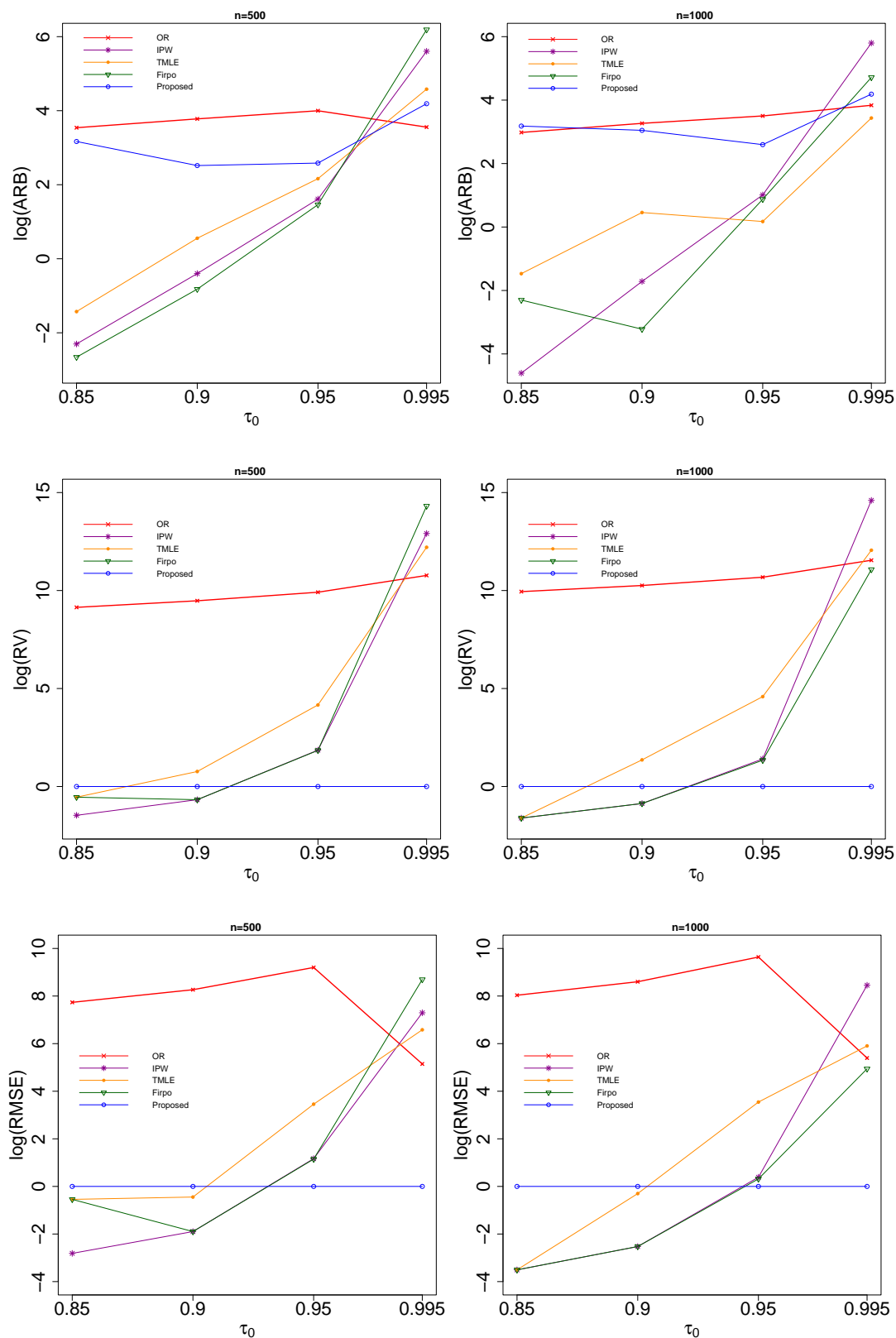


Figure 4: Comparison of proposed estimator, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the population at different probability levels based on the data generating process with heavy-tailed error from t-distribution. The top, middle and bottom rows show the comparison through three statistics: ARB, RV and RMSE, respectively, for two different sample sizes 500 and 1000. The numbers on the horizontal axis of each plot indicate the quantile levels, and five different symbols on the plot corresponds to five different estimators.



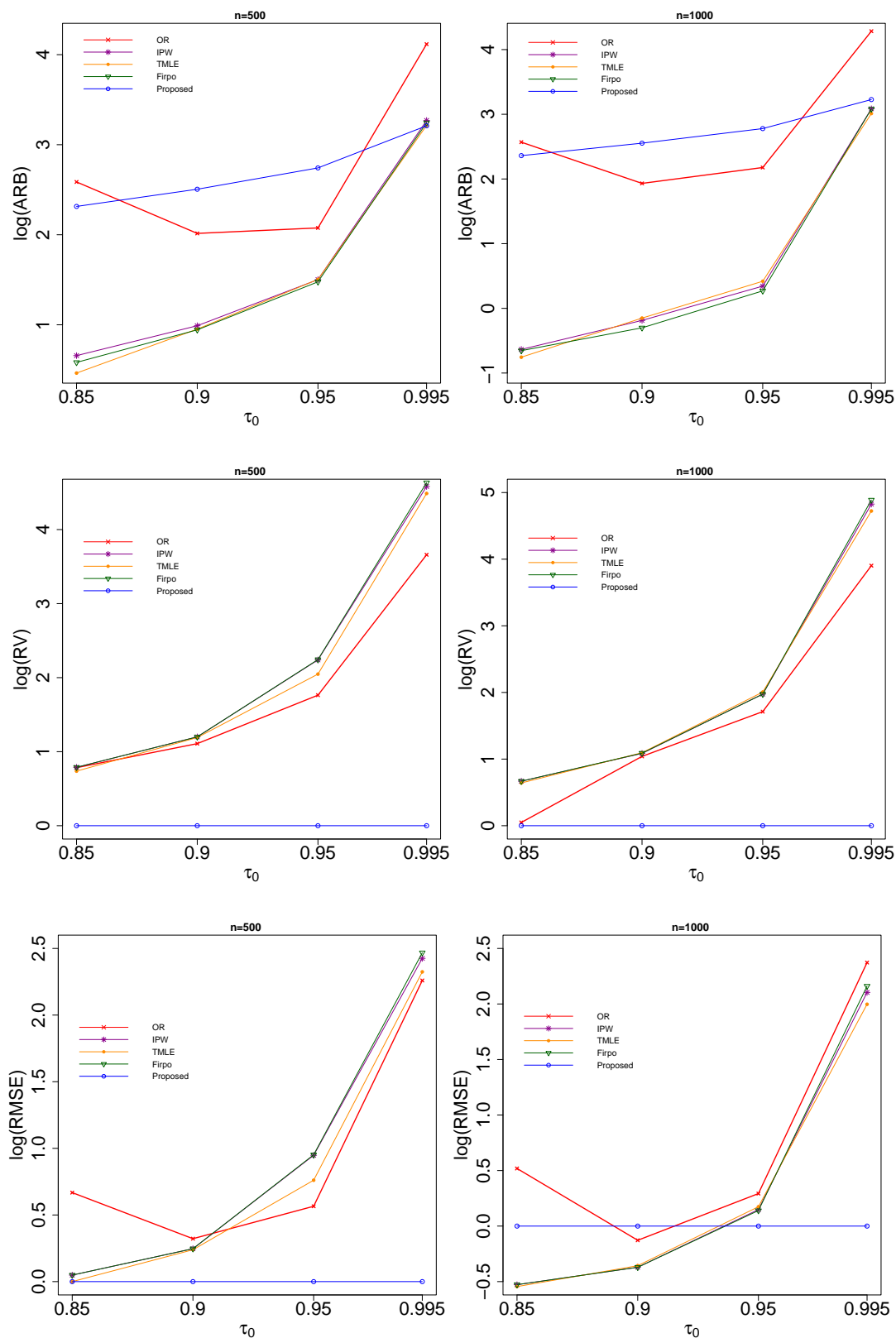


Figure 5: Comparison of proposed semiparametric, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the treated at different probability levels based on the data generating process with Gaussian error. The top, middle and bottom rows show the comparison through three statistics: ARB, RV and RMSE, respectively, for two different sample sizes 500 and 1000. The numbers on the horizontal axis of each plot indicate the quantile levels, and five different symbols on the plot corresponds to five different estimators.

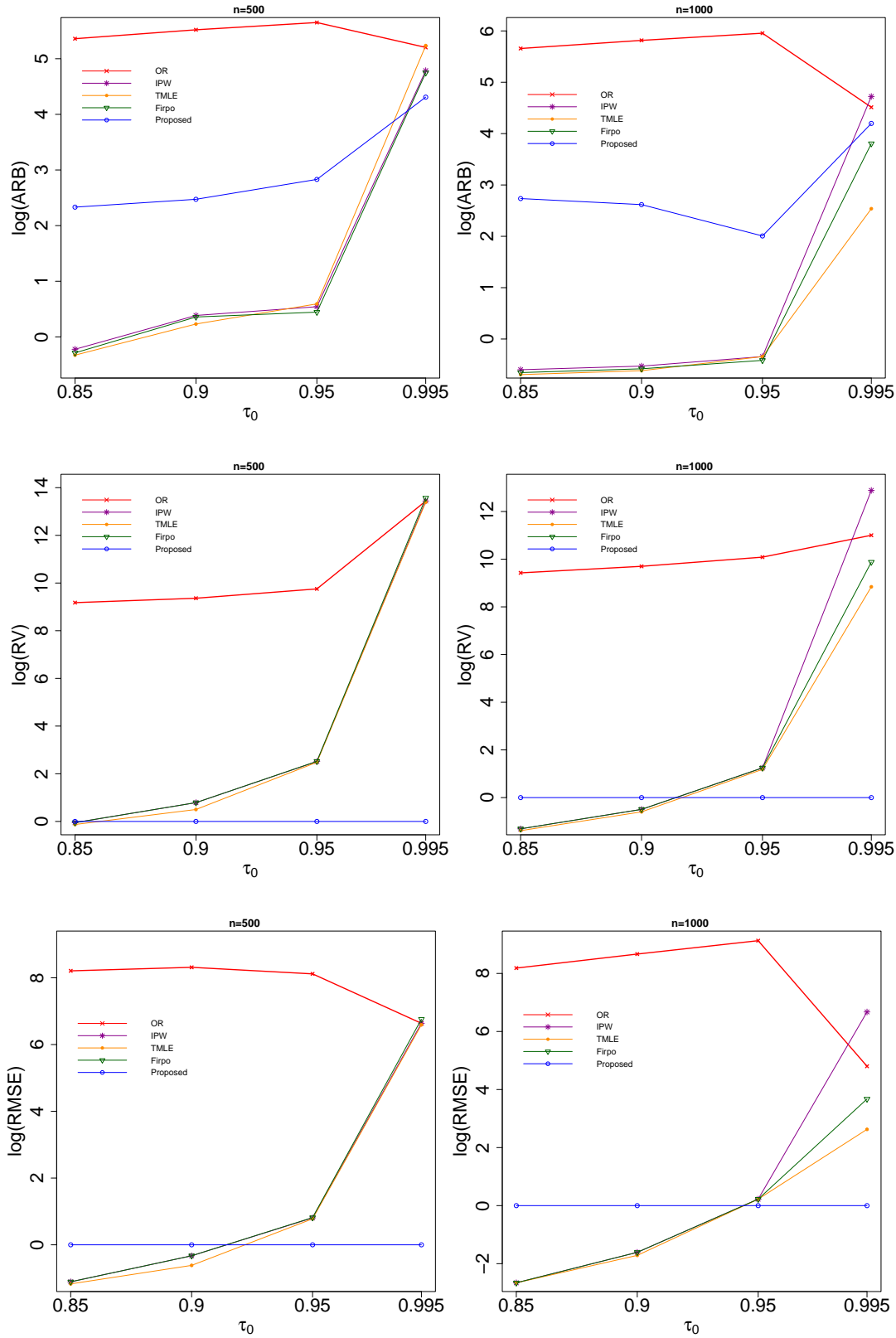


Figure 6: Comparison of proposed semiparametric, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the treated at different probability levels based on the data generating process with heavy-tailed error from t-distribution. The top, middle and bottom rows show the comparison through three statistics: ARB, RV and RMSE, respectively, for two different sample sizes 500 and 1000. The numbers on the horizontal axis of each plot indicate the quantile levels, and five different symbols on the plot corresponds to five different estimators.

expected that the accident rates will affect the traffic characteristics.

- (c) Bus-stop density: the ratio of the number of bus-stops to the road length. The presence of bus-stops is expected to affect the traffic flow and speed due to frequent bus-stops and pedestrian activities. The allocation of CS routes were designed to avoid areas with high bus-stop density for safety of the cyclists.
- (d) Road network density: with the available geographical information system we could also represent the road network density in each zone by using a measure of the number of network nodes per unit of area. A network node is defined as the meeting point of two or more links. To safeguard from conflicting turning movements the CS paths are routed through the areas with high road network density.
- (e) Road length: high capacity networks tend to depress land values which in turn will influence the socio-economic profile of the people who live close together. Data for road length for each zone was generated using geographical information system software.
- (f) Road type: a binary variable where '1' represents dual-carriageway and '0' represents single-carriage. This is an important feature since we might expect traffic congestion in single-carriage roads.
- (g) Density of domestic buildings: this is a potentially useful feature since we might expect congestion to be associated with the nature of land use and the degree of urbanization. Also, the allocation of the CS paths are possibly influenced by land use characteristics.
- (h) Density of non-domestic buildings: rising housing costs in business and office districts force people to live further away, lengthening commutes, and affecting traffic flow and speed. As mentioned before, this feature may influence allocation of the CS paths.
- (i) Road area density: the ratio of the area of the zone's total road network to the land area of the zone. The road network includes all roads in the including motorways, highways, main or national roads, secondary or regional roads. It is expected that the traffic flow is associated with road density.
- (j) Employment density: traffic generation potential depends on economic activity and we proxy this by employment density. High employment density tends to influence pedestrian activity

which in turn affects traffic speed. The CS paths are designed to provide coverage in the areas with high employment density and encourage commuters to use cycling as a regular mode of transport.

Figure 7 shows density plot of AADT, exhibiting heavy-tailed nature of the response variable. This feature of the response variable and total sample size of 450 observations is adequate to apply extreme theory and related inference ([Hosking and Wallis, 1987](#)). Using the proposed method (with

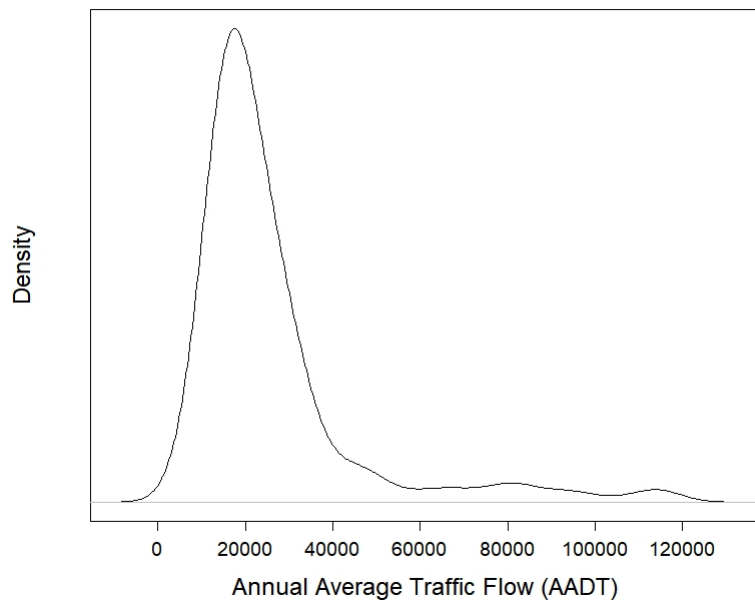


Figure 7: Density plot of Annual Average Traffic Flow during 2011-2014 in London metropolitan area.

same set of grid points and level of test as used in the preceding simulation study) we estimate the quantile treatment effects for the population and for the treated at different intermediate and extreme quantiles corresponding to probability levels  $\tau_0=0.85, 0.90, 0.95$  and  $0.995$ . We consider the OR, IPW, TMLE and Firpo methods along with the proposed method for the analysis. Propensity score is estimated using a logit model with the following factors: total cycle collision, bus-stop density, road network density, road length, density of domestic buildings, density of non-domestic buildings, road area density, and employment density. The OR method by [Zhang et al. \(2012\)](#) and the proposed method use density of domestic buildings, density of non-domestic buildings,

road area density, road network density, road type, bus-stop density as covariates. The estimated quantile treatment effects of CS on annual average daily traffic (AADT) relative to the respective quantile of AADT in the pre-intervention period are presented in Tables 1-2. The pre-intervention (2007 to 2010) quantiles of AADT are computed based on the observations from the same locations as considered for the post-intervention (2011-2014) period. The standard errors (SEs) and the corresponding 95% confidence intervals (CI) for all the estimates, with the exception of Firpo estimate at probability level  $\tau_0 = 0.995$ , are obtained from 1000 full bootstrap samples. To find the SE and CI of the causal effects based on Firpo method at probability level  $\tau_0 = 0.995$ , we employ *b*-out-of-*n* bootstrap proposed by Zhang et al. (2021). See Section 3.4 for details.

The result based on the proposed estimate indicate roughly 29% increase in traffic flow compared to the pre-intervention period at probability levels  $\tau_0 = 0.995$  both for the entire city and the treated locations. The 95% bootstrap CI for the proposed estimate exclude 0, suggesting a significant effect of CS on AADT. However, as observed in the simulation study, the SE of the estimated causal effects of CS at probability level  $\tau_0 = 0.995$  based on OR, IPW, and Firpo methods are very high and the most of the results are not significant except the estimate based on TMLE. The SE of the proposed method is significantly lower than those of the OR, IPW, TMLE and Firpo methods which is also reflected in the width of the confidence intervals at all probability levels for treated. A similar pattern is also observed for the population estimates at probability levels  $\tau_0 = 0.95$  and  $\tau_0 = 0.995$ .

Overall, our analysis indicates that the introduction of CS can trigger extreme traffic flows and increase traffic mobility. The protected cycleways in London including CS is just over 1% of the capital's roads and CS is only a quarter of the protected cycleways. In 2020, the five busiest roads in the U.K. were in London, however, none of these five roads is within proximity of cycleways. The U.K. has 38.3 million registered motor vehicles, up from 27 million in 2007. With cities such as London not appreciably adding road capacity with a massive expansion of motor vehicles will lead to traffic congestion (Reid, 2021). It is not unexpected that several unobserved factors associated with the transport network and other interventions play a crucial role in mitigating potential traffic problems anticipated by the introduction of cycle lanes. The results from our analysis provide more insights, but for a comprehensive understanding to formulate effective transport strategies, the effect of the cycleways on traffic speed requires further investigation.

Table 1: The quantile causal effect of CS on AADT relative to the respective quantile of AADT in the pre-intervention period for population at different probability levels  $\tau_0$ .

$\tau_0$	Method	Estimate (%)	SE	95% CI
0.85	OR	43.34	11.50	(25.36, 70.85)
	IPW	51.41	73.06	(-3.27, 219.09)
	TMLE	38.08	26.85	(4.47, 135.68)
	Firpo	108.17	22.37	(41.85, 135.01)
	Proposed	98.52	19.76	(65.19, 135.09)
0.90	OR	41.33	11.18	(22.88, 65.65)
	IPW	128.95	51.70	(0.35, 101.92)
	TMLE	100.47	38.50	(3.24, 98.34)
	Firpo	121.26	15.39	(30.40, 97.85)
	Proposed	71.60	14.40	(47.07, 97.63)
0.95	OR	37.60	10.43	(19.07, 59.49)
	IPW	61.50	19.66	(0.31, 83.81)
	TMLE	55.03	12.15	(2.03, 61.76)
	Firpo	59.69	9.71	(19.21, 61.45)
	Proposed	44.98	8.91	(28.57, 60.86)
0.995	OR	36.71	20.81	(-9.16, 69.66)
	IPW	2.08	9.32	(-4.78, 22.92)
	TMLE	15.37	6.26	(1.32, 39.97)
	Firpo	-0.95	15.88	(-19.21, 40.90)
	Proposed	29.11	5.86	(19.31, 39.77)



Table 2: The quantile causal effect of CS on AADT relative to the respective quantile of AADT in the pre-intervention period for treated at different probability levels  $\tau_0$ .

$\tau_0$	Method	Estimate (%)	SE	95% CI
0.85	OR	41.27	15.90	(14.56, 71.44)
	IPW	49.69	17.10	(21.55, 81.00)
	TMLE	49.22	20.75	(2.56, 77.02)
	Firpo	50.80	23.11	(26.97, 99.01)
	Proposed	44.65	8.76	(30.56, 62.12)
0.90	OR	50.27	10.09	(21.61, 62.42)
	IPW	60.94	13.51	(19.69, 71.28)
	TMLE	55.43	18.22	(-2.32, 66.11)
	Firpo	67.85	19.45	(18.78, 82.99)
	Proposed	37.71	7.27	(25.95, 52.19)
0.95	OR	45.04	8.65	(26.46, 59.49)
	IPW	38.19	13.46	(14.21, 69.27)
	TMLE	31.07	14.33	(-3.63, 56.57)
	Firpo	71.63	17.82	(20.72, 76.37)
	Proposed	34.31	7.04	(23.06, 48.46)
0.995	OR	-8.51	25.50	(-80.79, 21.02)
	IPW	15.53	8.79	(-2.37, 26.51)
	TMLE	-3.60	12.32	(-22.56, 24.75)
	Firpo	17.48	15.62	(3.08, 63.71)
	Proposed	28.80	5.90	(19.27, 40.13)

## 6 Discussions

In this article we have introduced a methodology that can be used to draw inference for causal effects at extreme quantiles. We have modelled the relationships between the covariates and the response and use those relationships to predict the outcome for both the treatment statuses and to estimate the treatment effect at certain extreme quantiles. The key methodological contribution is the improvisation of the conventional outcome regression (OR) model to combine a semi-parametric quantile regression framework with a heavy-tailed parametric component for the extreme tail of the response distribution. This approach also addresses issues with model misspecification due to parametric assumptions and accounts for the high variability at the extreme tails of the distribution. The inverse propensity weighted (IPW) estimate of [Zhang et al. \(2012\)](#) and the estimator proposed by [Firpo \(2007\)](#) are the usual alternatives for estimating causal effects at intermediate quantiles of a heavy-tailed distribution that avoid some of the parametric assumptions associated with the regression function and/or link function embedded in the OR approach. But these estimators exhibit high volatility at the extreme tails similar to that of the OR method, whereas the proposed method performs substantially better than all the competitors under consideration. Recently, [Deuber et al. \(2021\)](#) considered the estimator proposed by [Firpo \(2007\)](#) and adjusted it with the Hill estimator for the extreme value indices to estimate the extremal quantile treatment effect. The Hill estimator depends on the choice of selection of the number of upper order statistics  $\kappa$  such that  $\kappa \rightarrow \infty$  and  $\kappa/n \rightarrow 0$ . The choice of  $k$  is a crucial issue and several practical choices are prescribed in the literature. Our proposed method is different from this method and it doesn't require for selection the upper order statistics. Also, the performance of the Hill estimator based method is very similar to the method proposed by [Firpo \(2007\)](#) with respect to mean squared error.

In observational studies, one can never be sure that a model for the treatment assignment mechanism or an outcome regression model are correct. An alternative approach is to develop a doubly-robust (DR) estimator. Several DR estimation methodologies are proposed in the literature ([Kang and Schafer, 2007](#)). The approach suggested by [Zhang et al. \(2012\)](#) for doubly robust estimation for quantile causal effects is not readily extendable to our case. However, one can use augmented regression methods considering a suitable function of the propensity score  $\phi(\pi(D_i|x_i, \hat{\gamma}))$  as an additional covariate in [\(10\)](#) and [\(11\)](#). Although empirical validation of the DR property seems possible, the study of theoretical properties is a challenging problem under the current set-up. One can further develop a new doubly robust estimation methodology based on the targeted maximum

likelihood approach considered by [Díaz \(2017\)](#).

We have used the proposed method to analyse the effect of London Cycle Superhighways. Our results suggest that the introduction of Cycle Superhighways can increase extreme traffic flow, encouraging further research and cautious approaches for any possible extension of Cycle Superhighways in metropolitan cities like London, which are heavily affected by congestion. In this context, it would be an interesting problem to develop a statistical framework that can be used to derive inference for quantile causal effects incorporating both pre-intervention and post-intervention data as discussed in [Bhuyan et al. \(2021\)](#). Also, it is of great interest for general public and epidemiologists, particularly in light of the increase in cycling and the recent proliferation of emergency cycle lanes to support safe community during the pandemic. Despite these interventions, we are also seeing more commuting by private vehicles as people are trying to avoid public transport ([Reid, 2021](#)), hence an understanding of their impact on cycle lanes on congestion is important.

It is also worth noting that we are obliged to assume that the outcomes of one unit are not affected by the treatment assignment of any other units. While not always plausible, we attempt to reduce the “spillover” effects by reducing interactions between the treated and control units. One possible direction of further studies could be using an improved design and defining several types of treatment effects following [Hudgens and Halloran \(2008\)](#) and [Barkley et al. \(2020\)](#), and develop associated estimation methodologies for the setting where there may be clustered interference. Assessment of competing estimators of quantile effect at extreme tails which lies beyond the range of sample data could be another direction of future research. The scoring method of [Gandy et al. \(2021\)](#) for unconditional extreme quantile estimates can be extended to assess estimates of causal effects at extreme tails.

## Acknowledgement

The authors would like to acknowledge the Lloyd’s Register Foundation for funding this research through the programme on Data-Centric Engineering at the Alan Turing Institute. The authors are grateful to Dr. Haojie Li for providing the London Traffic data set, and Dr. Almut Veraart and Dr. Joydeep Chowdhury for helpful comments and suggestions. The authors also like to express their sincere thanks to the two anonymous reviewers and the Associate Editor for many helpful comments which have improved the presentation of this paper.

## References

- Bader, B., Yan, J., and Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Ann. Appl. Stat.*, 12(1):310–329. [12](#), [16](#)
- Badoe, D. A. and Miller, E. J. (2000). Transportation-land-use interaction: empirical findings in north america, and their implications for modeling. *Transportation Research Part D: Transport and Environment*, 5(4):235–263. [6](#)
- Balkema, A. A. and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792 – 804. [11](#)
- Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., and Emch, E. E. (2020). Causal inference from observational studies with clustered interference, with application to a cholera vaccine study. *Annals of Applied Statistics*, 14(3):1432–1448. [29](#)
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons. [11](#)
- Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2005). *Statistics of Extremes*. Wiley-Blackwell, 1 edition. [12](#)
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300. [12](#)
- Bhuyan, P., McCoy, E. J., Li, H., and Graham, D. J. (2021). Analysing the causal effect of london cycle superhighways on traffic congestion. *The Annals of Applied Statistics*, 15(4):1999–2022. [6](#), [29](#)
- Blunden, M. (2016). Cycle superhighways make traffic worse in the city, report reveals. *Evening-Standard*, Oct 5. [3](#)
- Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838. [13](#), [14](#), [17](#)

- Chen, C., Jia, Z., and Varaiya, P. (2001). Causes and cures of highway congestion. *IEEE Control Systems Magazine*, 21(6):26–32. 6
- Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268. 4, 9
- Chernozhukov, V., Fernández-Val, I., and Melly, B. (2020). Fast algorithms for the quantile regression process. *Empirical Economics*, 62:7–33. 10
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. 5, 11, 14
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442. 11, 12
- Deuber, D., Li, J., Engelke, S., and Maathuis, M. H. (2021). Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions. *arXiv:2110.06627*. 5, 28
- Drees, H., de Haan, L., and Resnick, S. (2000). How to make a hill plot. *Ann. Statist.*, 28(1):254–274. 11
- Díaz, I. (2017). Efficient estimation of quantiles in missing data models. *Journal of Statistical Planning and Inference*, 190:39–51. 4, 6, 8, 16, 17, 29
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Berlin: Springer. 4, 5
- Engelke, S. (2020). Graphical models and causality for extreme events. 5
- FarahaCarlos, H. and Azevedo, L. (2017). Safety analysis of passing maneuvers using extreme value theory. *IATSS Research*, 41:12–21. 6
- Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75:259–276. 3, 5, 9, 15, 17, 28
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190. 4

- Freedman, D. A. (2010). *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge University Press. 3
- Frölich, M. and Melly, B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, 31:346–357. 4
- Gandy, A., Jana, K., and Veraart, A. E. D. (2021). Scoring predictions at extreme quantiles. *Advances in Statistical Analysis*. 29
- Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology*, 36:21–47. 3
- George, K. A. (1970). Transportation compatible land uses and bus-stop location. *Transactions on The Built Environment*, 44. 6
- Gissibl, N. and Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720. 5
- Gissibl, N., Klüppelberg, C., and Otto, M. (2017). Tail dependence of recursive max-linear models with regularly varying noise variables. *arXiv: 1701.07351*. 5
- Gnecco, N., Meinshausen, N., Peters, J., and Engelke, S. (2019). Causal discovery in heavy-tailed models. *arXiv: 1908.05097*. 5
- Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, 83(2):263–292. 5
- G’Sell, M. G., Wager, S., Chouldechova, A., and Tibshirani, R. (2015). Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2):423–444. 12
- Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32(2):177 – 203. 11
- Hannart, A. and Naveau, P. (2018). Probabilities of causation of climate changes. *Journal of Climate*, 31(14):5507–5524. 3, 5



- Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349. 24
- Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842. 29
- Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press. 3
- Jin, J. and Rafferty, P. (2017). Does congestion negatively affect income growth and employment growth? empirical evidence from us metropolitan regions. *Transport Policy*, 55:1–8. 2
- Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical*, 22(4):523–539. 16, 28
- Karmakar, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395. 14
- Koenker, R. (2005). *Quantile regression*. Cambridge University press. 12
- Langousis, A., Mamalakis, A., Puliga, M., and Deidda, R. (2016). Threshold detection for the generalized pareto distribution: Review of representative methods and application to the noaa ncdc daily rainfall database. *Water Resources Research*, 52(4):2659–2681. 11
- Li, H., J., G. D., and Liu, P. (2017). Safety effects of the london cycle superhighways on cycle collisions. *Accident Analysis and Prevention*, 99:90–101. 3, 6
- Litvinova, S. and Silvapulle, M. J. (2020). Consistency of full-sample bootstrap for estimating high-quantile, tail probability, and tail index. *arXiv:2004.12639*. 15
- Liu, Y. and Wu., Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of nonparametric statistics*, 23(2):415–437. 13
- Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. *Rev. Labor Econ.*, 68:543–572. 4

- Mhalla, L., Chavez-Demoulin, V., and Dupuis, D. J. (2019). Causal mechanism of extreme river discharges in the upper danube basin network. [5](#)
- Moodie, E. E. M., Saarela, O., and Stephens, D. A. (2018). A doubly robust weighting estimator of the average treatment effect on the treated. *Stat*, 7(1). [8](#)
- Moodie, E. E. M. and Stephens, D. A. (2022). Causal inference: critical developments, past and future. *arXiv:2204.02231*. [7](#)
- Naveau, P., Hannart, A., and Ribes, A. (2020). Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and Its Application*, 7(1):null. [3](#)
- Norman, W. (2017). Bike lanes don’t clog up our roads, they keep london moving. *The Gaurdian*, Dec 1. [3](#)
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, 3(1):119–131. [11](#)
- Reid, C. (2021). None of top five congested roads in u.k. feature adjacent cycleways. *Forbes*, Dec 7. [2](#), [25](#), [29](#)
- Retallack, A. and Ostendorf, B. (2019). Current understanding of the effects of congestion on traffic accidents. *Int J Environ Res Public Health*, 13(16(18)). [2](#)
- Ribes, A., Thao, S., and Cattiaux, J. (2020). Describing the relationship between a weather event and climate change: a new statistical approach. *Journal of Climate*, 0(0):null. [5](#)
- Rosenbaum, P. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effec. *Biometrika*, 40:41–55. [7](#)
- Scarrott, C. and MacDonald, A. (2012). Review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, 10(1). [11](#)
- Schnabel, S. K. and Eilers, P. H. C. (2013). Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, 97:77–87. [13](#)
- Slawson, N. (2017). Traffic jams on major uk roads cost economy around £9bn. *The Gaurdian*, Oct 18. [2](#)

- Transport for London (2010). Cycling revolution london. [2](#)
- Transport for London (2011). Barclays cycle superhighways evaluation of pilot routes 3 and 7. [2](#)
- UN Environment Programme (2019). Cycling, the better mode of transport. <https://www.unep.org/news-and-stories/story/cycling-better-mode-transport>, June 11. [2](#)
- United Nations (2021). World bicycle day. <https://www.un.org/en/observances/bicycle-day>, June 03. [2](#)
- Xian, X., Ye, H., Wang, X., and Lie, K. (2021). Spatiotemporal modeling and real-time prediction of origin-destination traffic demand. *63*, 1(77-89). [2](#)
- Xu, D., Daniels, M. J., and Winterstein, A. G. (2018). A bayesian nonparametric approach to causal inference on quantiles. *Biometrics*, 74:259–276. [4](#)
- Xu, L. and Nusholtz, G. (2017). Application of extreme value theory to crash data analysis. *The Stapp Car Crash Journal*, 61:287–298. [6](#)
- Zhang, K., Sun, D. J., Shen, S., and Zhu, Y. (2017). Analyzing spatiotemporal congestion pattern on urban roads based on taxi gps data. *Journal of Transport and Land Use*, 10(1):675–694. [6](#)
- Zhang, N., J., G. D., Hörcher, D., and Bansal, P. (2021). A causal inference approach to measure the vulnerability of urban metro systems. *Transportation*. [3](#), [15](#), [25](#)
- Zhang, Y. (2018). Extremal quantile treatment effects. *The Annals of Statistics*, 46(6B):3707 – 3740. [5](#)
- Zhang, Z., Chen, Z., Troendle, J. F., and Zhang, J. (2012). Causal inference on quantiles with an obstetric application. *Biometrics*, 68(3):697–706. [4](#), [8](#), [9](#), [15](#), [17](#), [24](#), [28](#)
- Zheng, L. and Sayed, T. (2019). Application of extreme value theory for before-after road safety analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2673. [6](#)
- Zou, H. and Yuan, M. (2008). Regularized simultaneous model selection in multiple quantiles regression. *Computational Statistics & Data Analysis*, 52(12):5296 – 5304. [13](#)