Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 1452–1474
https://doi.org/10.1093/jrsssc/qlad080
Advance access publication 7 September 2023
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Causal analysis at extreme quantiles with application to London traffic flow data

Prajamitra Bhuyan[1,†], Kaushik Jana[2,†] and Emma J. McCoy[3]

[1]Operations Management Group, Indian Institute of Management, Calcutta, India
[2]Mathematical and Physical Sciences Division, Ahmedabad University, Ahmedabad, India
[3]Department of Statistics, London School of Economics and Political Science, London, UK

Address for correspondence: Prajamitra Bhuyan, Operations Management Group, Indian Institute of Management, Calcutta, Diamond Harbour Road, Joka, Kolkata 700104, West Bengal, India. Email: prajamitra.bhuyan@gmail.com

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.

Keywords: 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 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 (Jin & Rafferty, 2017; Slawson, 2017), and road safety, thus imparting socio-economic distress (Retallack & Ostendorf, 2019). Intelligent transportation systems can revolutionise 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 arond 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

† P.B. and K.J. contributed equally to this paper.

Received: February 4, 2022. **Revised:** April 26, 2023. **Accepted:** August 7, 2023
© The Royal Statistical Society 2023. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

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-m 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 CS 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 centre 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 (Blunden, 2016; Norman, 2017). 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 socio-economic 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.

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 & Rubin, 2015), transport engineering (Li et al., 2017; Zhang & Bansal, 2021), climate science (Hannart & Naveau, 2018; Naveau et al., 2020), and public policy (Freedman, 2010; Gangl, 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 summarised 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 minimising an inverse probability weighted check loss function, which achieves non-parametric consistency by means of a propensity score estimated as a logistic power series with degree increases asymptotically. Frölich and Melly (2013), and 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 semi-parametric approach using a targeted maximum likelihood estimator (TMLE) for the quantiles. In this context, Xu et al. (2018) introduced a Bayesian non-parametric 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 normalised random variable resides in the domain of attraction of extreme value distribution (Fisher & Tippett, 1928). Thus, extreme quantiles can be predicted by estimating the parameters of the corresponding extreme value distribution and the normalising 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. (2021) and Engelke and Hitz (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) and Gissibl et al. (2017) studied the

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025



Figure 1. Examples of London Cycle Superhighways.

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 CS 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 & Miller, 2000; Bhuyan et al., 2021; Chen et al., 2001; Zhang et al., 2017). 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

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

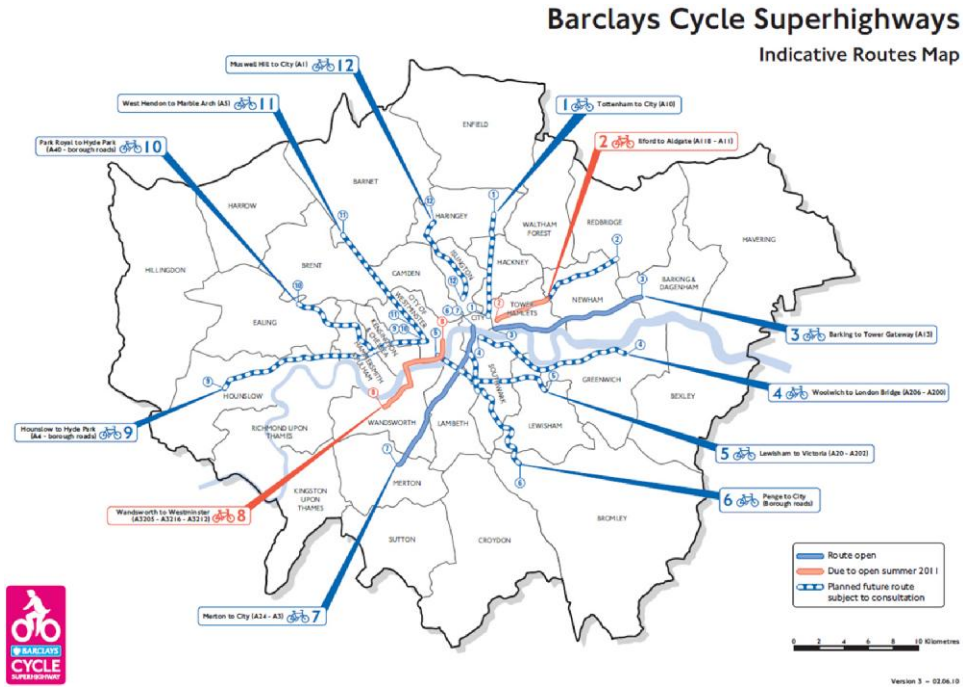


Figure 2. Route map of London Cycle Superhighways.

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 emphasised 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 (FarahaCarlos & Azevedo, 2017; Xu & Nusholtz, 2017; Zheng & Sayed, 2019). 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 consequnces. 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 characterised 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 summarised in Section 4. In Section 5, we discuss the results obtained by analysing the London CS 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, $Z_i = (Y_i, D_i, \boldsymbol{X_i})$, $i = 1, \ldots, n$, where for the ith unit of observation, Y_i denotes the response, D_i the treatment (intervention or exposure) received, and $\boldsymbol{X_i}$ a vector of

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

confounders or covariates. Let the triplet $(Y, D, \boldsymbol{X})$ represents a generic form of $(Y_i, D_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$. 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 $\boldsymbol{X}$ 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 $(Y(0), Y(1)) \perp\!\!\!\perp D \mid \boldsymbol{X}$, 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 $\boldsymbol{X}$, 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 & 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. $0 < \mathbb{P}[D = 1 \mid \boldsymbol{X} = \boldsymbol{x}] < 1$ for all $\boldsymbol{x}$. This assumption implies that there exists no strata of $\boldsymbol{X}$ such that treatment assignment is uniquely determined. To illustrate this point, suppose that for a given value of $\boldsymbol{X}$, say $\boldsymbol{x}^*$, $\mathbb{P}[D = 0 \mid \boldsymbol{X} = \boldsymbol{x}^*] = 1$. Then there will be no triplets of observed data $(Y, D, \boldsymbol{X})$ such that $D = 1$ and $\boldsymbol{X} = \boldsymbol{x}^*$; thus, any estimates of treatment effects, say a conditional average treatment effect such as $\mathbb{E}[Y \mid D = 1, \boldsymbol{X} = \boldsymbol{x}^*] - \mathbb{E}[Y \mid D = 0, \boldsymbol{X} = \boldsymbol{x}^*]$ 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 summarised 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 percentiles) of the potential outcomes (Díaz, 2017; Zhang et al., 2012). 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 τ_0th $(0 < \tau_0 < 1)$ quantile is defined as

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

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 vs. 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

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

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

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

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. counterfractual distribution) is given by

$$\begin{aligned} F_t(y) &= \mathbb{E}\big[P(Y(t) \le y \mid \boldsymbol{X})\big] \\ &= \mathbb{E}\big[P(Y(t) \le y \mid D = t, \boldsymbol{X})\big] \\ &= \mathbb{E}\big[P(Y \le y \mid D = t, \boldsymbol{X})\big] \\ &= \int G_{t,x}(y) dH(\boldsymbol{x}), \end{aligned} \tag{3}$$

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

$$F_{t|1}(y) = E\{P(Y(t) \le y \mid \boldsymbol{X}) \mid D = 1\} = \int G_{t,x}(y) dH_1(\boldsymbol{x}), \tag{4}$$

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

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

$$\hat{F}_t(y) = n^{-1} \sum_{i=1}^{n} \hat{G}_{t,\boldsymbol{X}_i}(y), \tag{5}$$

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 τ_0th 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 equation (1) to obtain the estimate of the quantile treatment effect for the population. This is known as outcome regression-based (OR) estimate. Similarly, one can estimate $F_{t|1}(\cdot)$, given in equation (4), by

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

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 ζ_{τ_0} defined in equation (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 $\boldsymbol{X} = \boldsymbol{x}$ can be obtained from the conditional quantile function $Q_{Y|t,\boldsymbol{x}}(\tau) = G^{-1}_{t,\boldsymbol{x}}(\tau)$ of the outcome Y for all $\tau \in (0, 1)$ (Chernozhukov et al., 2013). Let $U \sim Unif(0, 1)$. Then we can write

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

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 $G_{t,x}$ with some heavy-tailed

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

components. To model the bulk and the extreme component of $G_{t,\boldsymbol{x}}(y)$ efficiently, we choose a suitable transition probability level $\tau_u \in (0, 1)$ and decompose equation (7) as

$$G_{t,\boldsymbol{x}}(y) = \int_0^{\tau_u} \mathbb{I}\left[Q_{Y|t,\boldsymbol{x}}(\tau) \le y\right] d\tau + \int_{\tau_u}^1 \mathbb{I}\left[Q_{Y|t,\boldsymbol{x}}(\tau) \le y\right] d\tau. \tag{8}$$

The transition point τ_u is assumed to represent the probability level of quantile above which the distribution (second part of equation (8)) exhibits heavy-tailed nature. At the same time a very high probability level as a choice of τ_u, may prevent enough number of observations being generated from the extreme part of the distribution. In practice, τ_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 τ_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 equation (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 $G_{t,\boldsymbol{x}}(\cdot)$ as

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

where the integer u is determined from τ_u, and $\hat{Q}_{Y|t,\boldsymbol{x}}(\tau_j)$ and $\tilde{Q}_{Y|t,\boldsymbol{x}}(\tau_j)$ are estimates of $Q_{Y|t,\boldsymbol{x}}(\tau_j)$ 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 and 3.2. The estimated conditional distribution function is then plugged into equation (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 $\hat{q}_{Y(t)}(\tau_0) = \inf\{z : \hat{F}_t(z) \ge \tau_0\}$ for $t = 0, 1$, and finally the estimate of quantile causal effect η_{τ_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 $\hat{q}_{Y(t)|D=1}(\tau_0) = \inf\{z : \hat{F}_{t|1}(z \mid D = 1) \ge \tau_0\}$ from equation (6), and the estimate of ζ_{τ_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 τ_uth quantile in equation (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 Ψ with right end point y_Ψ. Then by the Pickands-Balkema-de Haan Theorem (Balkema & de Haan, 1974; Pickands, 1975), the conditional distribution of $Y - U$ above an excess $U < y_\Psi$ is well approximated by the generalised Pareto distribution (GPD) for large U under some general regularity conditions. That is,

$$\lim_{y \uparrow y_\Psi} P\left[Y - U \le y \mid Y > U\right] \to H(y)$$

and H is the distribution function of GPD, given as

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

where $y \ge 0$ if $\xi \ge 0$, and $0 \le y \le -\sigma/\xi$ if $\xi < 0$. Here σ and ξ are the scale and shape parameters, respectively. The value of the shape parameter $\xi = 0$, interpreted as $\xi \to 0$, leads H to be the

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

exponential distribution with mean σ. The negative values of ξ 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 ξ 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 exceedances 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 equation (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 (Coles, 2001; Davison & Smith, 1990; Drees et al., 2000). The second approach involves methods that minimise 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 (Beirlant et al., 2004; Hall, 1990; Langousis et al., 2016). The third category of methods are based on the goodness-of-fit (gof) test of the GPD (Bader et al., 2018; Davison & Smith, 1990), 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 τ_u

To select the transition point τ_u in equation (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 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 gof 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 $U(\tau \mid \boldsymbol{w}) = \boldsymbol{w}^T \hat{\boldsymbol{\beta}}(\tau)$, where $\boldsymbol{v} = (t, \boldsymbol{x})$ and $\boldsymbol{w} = (1, \boldsymbol{v})^T$ is a covariate vector of dimension $(m + 1)$ including 1, and $\hat{\boldsymbol{\beta}}(\tau) = (\hat{\beta}_0(\tau), \ldots, \hat{\beta}_m(\tau))^T$ are the estimates of the regression parameter vector at $\tau = \tau_1, \ldots, \tau_l$. Therefore, for a given set of l probability levels, the method produces thresholds $U(\tau_1 \mid \boldsymbol{w}) \leq \cdots \leq U(\tau_l \mid \boldsymbol{w})$. We then use each of the thresholds, and fit a GPD to the exceedences 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 exceedences corresponding to the kth threshold, for $k = 1, \ldots, l$. Then the sequence of null hypotheses to be tested are $H_0^{(k)}$: the distribution of n_k observations above $U(\tau_k \mid \boldsymbol{w})$ 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 $H_0^{(1)}, \ldots, H_0^{(l)}$ when the tests are dependent. The method uses the p-values corresponding to the tests of l hypotheses as $p_1, \ldots, p_l \in [0, 1]$. 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 cut-off $\hat{k}$ such that $H_0^{(1)}, \ldots, H_0^{(\hat{k})}$ are rejected. They prescribe that if no $\hat{k} \in 1, \ldots, l$ exists, then no rejection is made. ForwardStop is given by

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

where λ is a pre-specified level. In this sequential testing, stopping at k implies that the gof test of the GPD to the exceedances at the first k thresholds $U(\tau_l \mid \boldsymbol{w}), \ldots, U(\tau_k \mid \boldsymbol{w})$ are rejected. In other

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

words, the set of first k null hypotheses $H_0^{(1)}, \ldots, H_0^{(k)}$ are rejected. Let us denote $U(\tau_u \mid \boldsymbol{w})$ 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 equation (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 & Yuan, 2008). Simultaneous estimation borrows information from all the neighbouring quantiles that may improve the efficiency of individual quantile estimates (Liu & 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 equation (9)). Let $\boldsymbol{C} \subset \mathbb{R}^m$, be a closed convex polytope, represented as the convex hull of n points in m-dimensions. Recall that $\boldsymbol{v} = (D, \boldsymbol{x})$ and $\boldsymbol{w} = (1, \boldsymbol{v})^T$, and we are interested in ensuring that the quantile curves do not cross for all values of the covariate $\boldsymbol{v} \in \boldsymbol{C}$. For estimating multiple quantiles at probability levels $\tau_1 <, \ldots, < \tau_u$, Bondell et al. (2010) devised a simple and computationally efficient procedure using a version of constrained optimisation that gureentees non-crossing of the estimated quantiles for all values of the covariate $\boldsymbol{v} \in \boldsymbol{C}$. This goal is achieved by imposing appropriate dominance in every consecutive pair of quantile regression planes. Under this set-up, the parameters associated with the quantile function are estimated by solving the following optimisation problem:

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

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 $\boldsymbol{X} = \boldsymbol{x}$ in the first summand of equation (9) is given by $\hat{Q}_{Y|t,\boldsymbol{x}}(\tau) = \hat{\beta}_0(\tau) + t\hat{\beta}_1(\tau) + \boldsymbol{x}^T \hat{\boldsymbol{\beta}}^*(\tau)$, where $\hat{\boldsymbol{\beta}}^*(\tau)$ is the $(m-1)$ dimensional vector of estimated coefficients corresponding to the covariate vector $\boldsymbol{x}$, for $t = 0, 1$, $\tau = \tau_1, \ldots, \tau_u$.

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 $U(\tau_u \mid \boldsymbol{w})$ provided in Subsection 3.1 guarantees the suitability of GPD to model the exceedences $Y - U(\tau_u \mid \boldsymbol{w})$, where $U(\tau_u \mid \boldsymbol{w}) = \hat{\beta}_0(\tau_u) + t\hat{\beta}_1(\tau_u) + \boldsymbol{w}^T \hat{\boldsymbol{\beta}}^*(\tau_u)$ 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 $\boldsymbol{X} = \boldsymbol{x}$ at probability

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

level $\tau = \tau_{u+1}, \ldots, \tau_J$, in the second summand of equation (9) is given by the approximation (see Coles, 2001, Section 4.3.3):

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

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

Note that, under the assumption of a heavy-tailed response distribution (with shape parameter $\xi > 0$), the above estimated quantile $\tilde{Q}_{Y|t,\boldsymbol{w}}(\tau)$ in equation (11) is an increasing function of τ. This fact ensures no crossings for the fitted quantiles at extreme tails. Also, each of the quantile estimate $\tilde{Q}_{Y|t,\boldsymbol{w}}(\tau)$ is larger than the threshold $U(\tau_u \mid \boldsymbol{w})$ 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 and Bansal (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 bootstap considering the following steps to compute the standard error and confidence interval of the proposed estimator.

Step 1. Sample $(Y_i^*, D_i^*, \boldsymbol{X}_i^*)$ from the observed data $(Y_i, D_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$.
Step 2. Based on the data $(Y_i^*, D_i^*, \boldsymbol{X}_i^*)$, $i = 1, \ldots, n$, 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 1,000.
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 & 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

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

distribution with both mean and standard deviation 1, and specify the following relationships between the response variable Y and a binary treatment D:

$$\text{logit}[\mathbb{E}(D \mid X_1, X_2, X_3)] = -3 + 0.1X_1 + 0.1X_2 + 0.2X_3$$
$$Y = 10 + 15D + X_1 + 3X_2 + 2X_1D + (1 + 4X_2 + 3D)\epsilon,$$

where error ϵ is generated from two different symmetric distributions with mean 0. In the first case, we consider ϵ following the normal distribution with mean 0 and standard deviation 10 and in the second case ϵ 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 $n = 1{,}000$.

The transition point representing the shift from the bulk to the extreme part (τ_u in equation (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 τ 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 gof 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 τ_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 equation (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 equation (9), we adopt the simultaneous quantile regression approach of Bondell et al. (2010) and use the associated R package. In the extreme part of equation (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 sub-population. As a set of competing estimators, we consider the 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 1,000 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:

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

where $\hat{\eta}_{\tau_0}^{(i)}$ and $\hat{\hat{\eta}}_{\tau_0}^{(i)}$ are the estimates of the population quantile effect at τ_0 based on the proposed method and a competitor in the ith iteration, respectively, with $L = 1{,}000$. Similarly, we compute these metrics for the treated sub-population (i.e. for ζ_{τ_0}). The results for the quantile effects on population are summarised graphically in Figures 3 and 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 sub-population and the findings are presented in Figures 5 and 6.

First, we discuss the results corresponding to quantile causal effects for the population (Figures 3 and 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

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

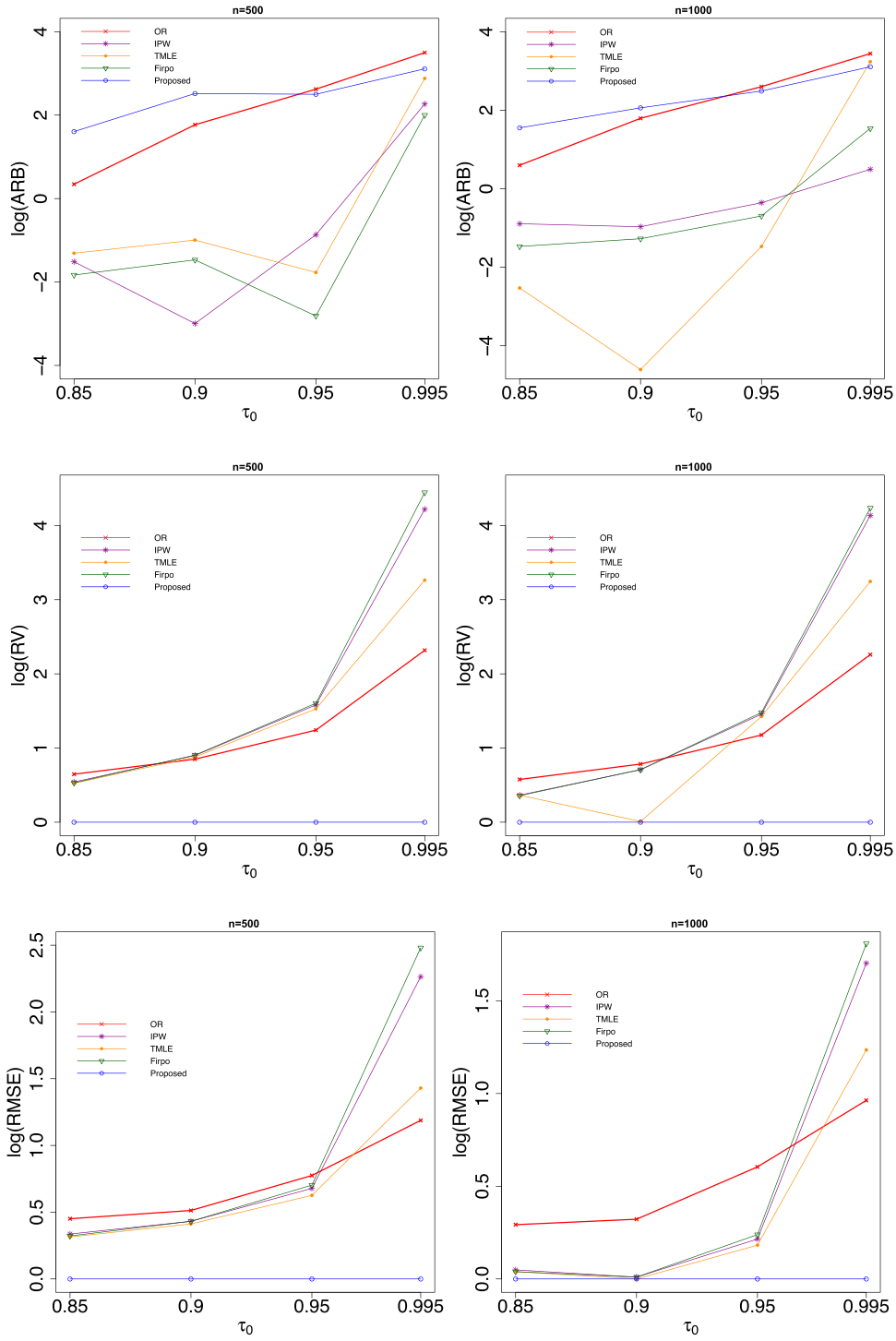


Figure 3. Comparison of proposed estimator, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the population at different probability levels based on the data generating process with Gaussian error. The top, middle, and bottom rows show the comparison through three statistics: ARB, RV, and RMSE, respectively, for two different sample sizes 500 and 1,000. 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. OR = outcome regression; IPW = inverse propensity weighting; TMLE = targeted maximum likelihood estimate; ARB = absolute relative bias; RV = relative variance; RMSE = relative mean squared error.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

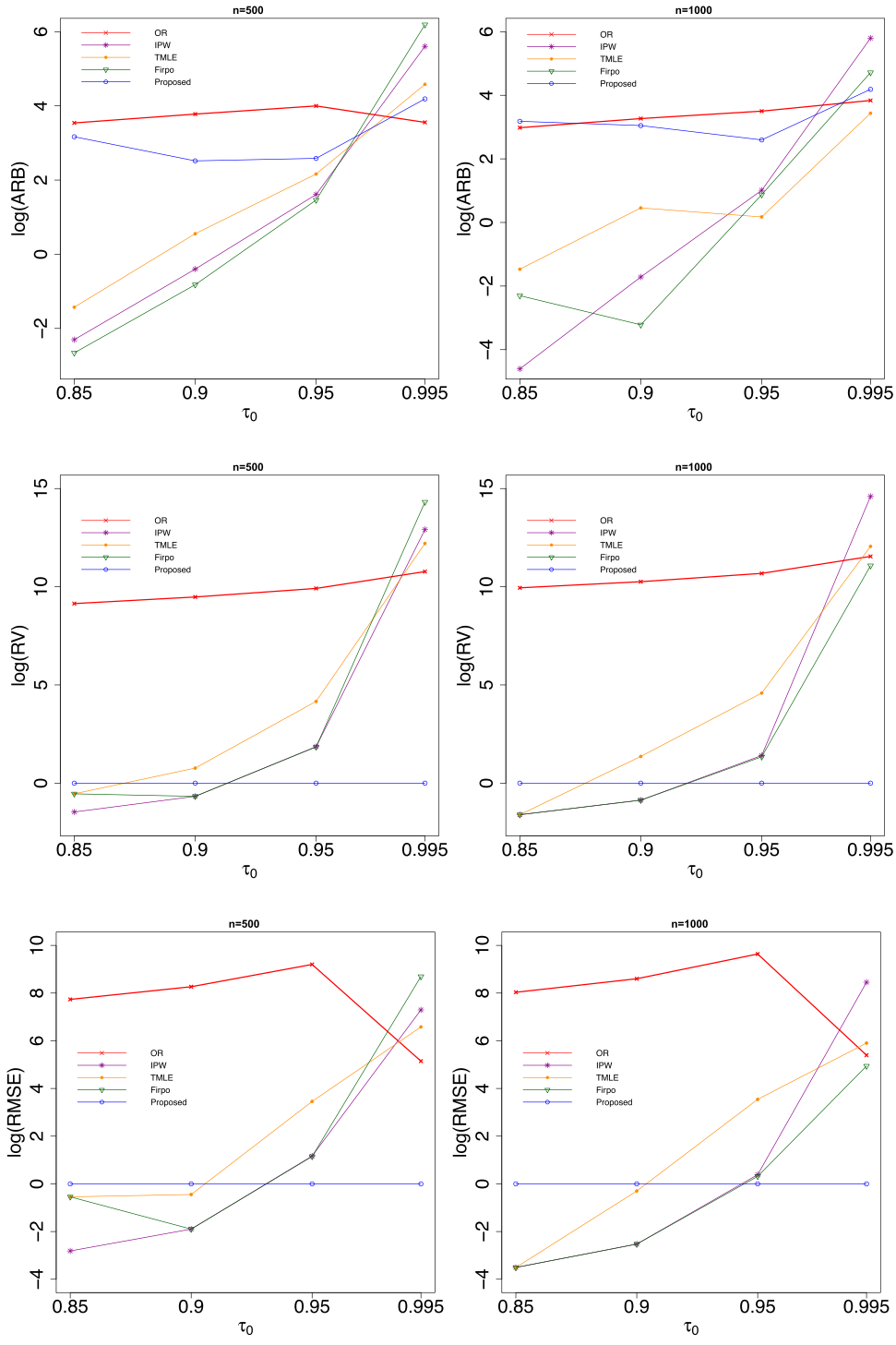


Figure 4. Comparison of proposed estimator, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the population at different probability levels based on the data generating process with heavy-tailed error from *t*-distribution. The top, middle, and bottom rows show the comparison through three statistics: ARB, RV, and RMSE, respectively, for two different sample sizes 500 and 1,000. 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. OR = outcome regression; IPW = inverse propensity weighting; TMLE = targeted maximum likelihood estimate; ARB = absolute relative bias; RV = relative variance; RMSE = relative mean squared error.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

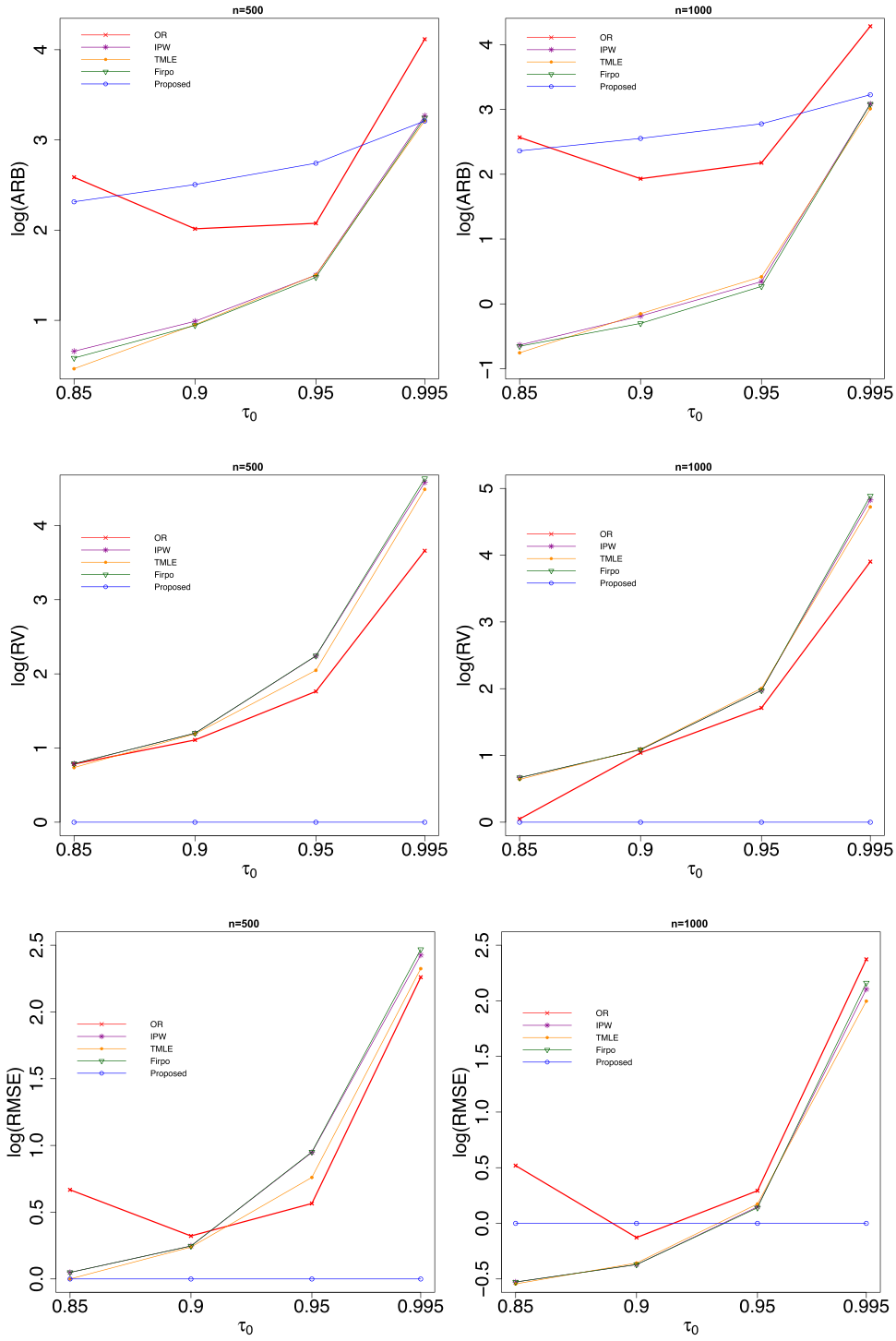


Figure 5. Comparison of proposed semi-parametric, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the treated at different probability levels based on the data generating process with Gaussian error. The top, middle, and bottom rows show the comparison through three statistics: ARB, RV, and RMSE, respectively, for two different sample sizes 500 and 1,000. 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. OR = outcome regression; IPW = inverse propensity weighting; TMLE = targeted maximum likelihood estimate; ARB = absolute relative bias; RV = relative variance; RMSE = relative mean squared error.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

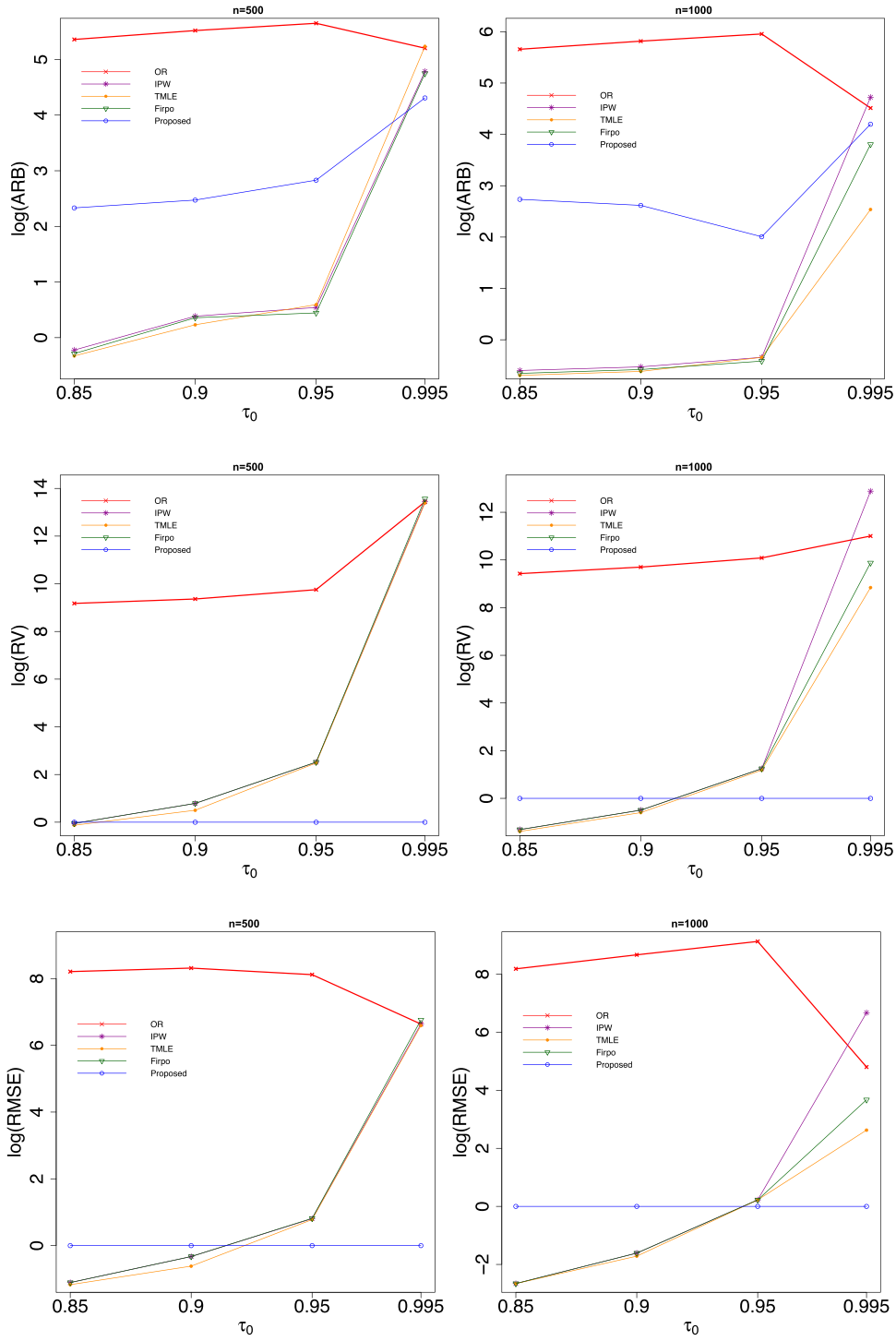


Figure 6. Comparison of proposed semi-parametric, OR, IPW, TMLE, and Firpo estimators for quantile treatment effect for the treated at different probability levels based on the data generating process with heavy-tailed error from *t*-distribution. The top, middle, and bottom rows show the comparison through three statistics: ARB, RV, and RMSE, respectively, for two different sample sizes 500 and 1,000. 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. OR = outcome regression; IPW = inverse propensity weighting; TMLE = targeted maximum likelihood estimate; ARB = absolute relative bias; RV = relative variance; RMSE = relative mean squared error.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

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 set-up, 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 and 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 CS effects

The primary focus of this study is to evaluate the effect of London CS (intervention) to contain extreme congestion. The data that are available and the rationale to construct the response [annual average daily traffic (AADT)] 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 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 urbanisation. 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.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

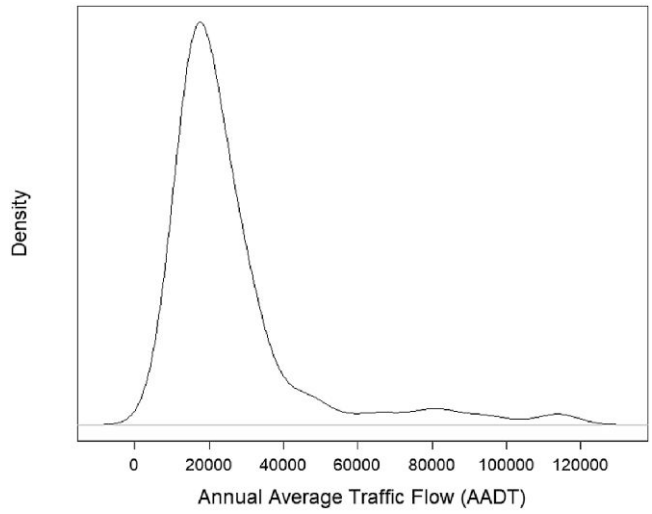


Figure 7. Density plot of annual average traffic flow during 2011–2014 in London metropolitan area.

(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 & Wallis, 1987). Using the proposed method (with 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 AADT relative to the respective quantile of AADT in the pre-intervention period are presented in Tables 1 and 2. The pre-intervention (2007–2010) 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 1,000 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 and Bansal (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

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

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 τ_0

τ_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)

Note. OR = outcome regression; IPW = inverse propensity weighting; TMLE = targeted maximum likelihood estimate; CS = Cycle Superhighways; AADT = annual average daily traffic.

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.

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 OR model to combine a semi-parametric quantile

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

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 τ_0

τ_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)

Note. OR = outcome regression; IPW = inverse propensity weighting; TMLE = targeted maximum likelihood estimate; CS = Cycle Superhighways; AADT = annual average daily traffic.

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 κ such that $\kappa \to \infty$ and $\kappa/n \to 0$. 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 does not 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 OR model are correct. An alternative approach is to develop a doubly robust (DR) estimator. Several DR estimation methodologies are proposed in the literature (Kang & 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 $\phi(\pi(D_i \mid x_i, \hat{\gamma}))$ as an additional covariate in equations (10) and (11). Although empirical validation of the DR property seems possible, the study of theoretical properties is a challenging problem under the current

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

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 CS. Our results suggest that the introduction of CS can increase extreme traffic flow, encouraging further research and cautious approaches for any possible extension of CS 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. (2022) for unconditional extreme quantile estimates can be extended to assess estimates of causal effects at extreme tails.

Acknowledgments

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.

Conflict of interest: The authors have declared that no competing interests exist.

Funding

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.

Data availability

The data that support the findings of this study are available on request from Dr. Haojie Li. The data are not publicly available due to privacy or ethical restrictions.

References

Bader B., Yan J., & Zhang X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Annals of Applied Statistics*, *12*(1), 310–329. https://doi.org/10.1214/17-AOAS1092

Badoe D. A., & 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. https://doi.org/10.1016/S1361-9209(99)00036-X

Balkema A. A., & de Haan L. (1974). Residual life time at great age. *Annals of Probability*, *2*(5), 792–804. https://doi.org/10.1214/aop/1176996548

Barkley B. G., Hudgens M. G., Clemens J. D., Ali M., & 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. https://doi.org/10.1214/19-AOAS1314

Beirlant J., Goegebeur Y., Segers J., Teugels J., De Waal D., & Ferro C. (2004). *Statistics of extremes: Theory and applications*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Beirlant J., Goegebeur Y., Teugels J., & Segers J. (2005). *Statistics of extremes* (1st ed.). Wiley-Blackwell.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

Benjamini Y., & 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. https://doi.org/10.1111/rssb.1995.57.issue-1

Bhuyan P., McCoy E. J., Li H., & Graham D. J. (2021). Analysing the causal effect of London cycle superhighways on traffic congestion. *The Annals of Applied Statistics*, *15*(4), 1999–2022. https://doi.org/10.1214/21-AOAS1450

Blunden M. (2016). Cycle superhighways make traffic worse in the city, report reveals. *EveningStandard*. https://www.standard.co.uk/news/transport/cycle-superhighways-make-traffic-worse-in-the-city-report-reveals-a3361216.html

Bondell H. D., Reich B. J., & Wang H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, *97*(4), 825–838. https://doi.org/10.1093/biomet/asq048

Chen C., Jia Z., & Varaiya P. (2001). Causes and cures of highway congestion. *IEEE Control Systems Magazine*, *21*(6), 26–32. https://doi.org/10.1109/37.969132

Chernozhukov V., Fernández-Val I., & Melly B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268. https://doi.org/10.3982/ECTA10582

Chernozhukov V., Fernández-Val I., & Melly B. (2020). Fast algorithms for the quantile regression process. *Empirical Economics*, *62*(1), 7–33. https://doi.org/10.1007/s00181-020-01898-0

Coles S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Davison A. C., & Smith R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, *52*(3), 393–442. https://doi.org/10.1111/rssb.1990.52.issue-3

Deuber D., Li J., Engelke S., & Maathuis M. H. (2021). Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions. arXiv:2110.06627, preprint: not peer reviewed.

Díaz I. (2017). Efficient estimation of quantiles in missing data models. *Journal of Statistical Planning and Inference*, *190*, 39–51. https://doi.org/10.1016/j.jspi.2017.05.001

Drees H., de Haan L., & Resnick S. (2000). How to make a hill plot. *Annals of Statistics*, *28*(1), 254–274. https://doi.org/10.1214/aos/1016120372

Embrechts P., Klüppelberg C., & Mikosch T. (1997). *Modelling extremal events*. Springer.

Engelke S., & Hitz S. A. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(4), 871–932. https://doi.org/10.1111/rssb.12355

FarahaCarlos H., & Azevedo L. (2017). Safety analysis of passing maneuvers using extreme value theory. *IATSS Research*, *41*(1), 12–21. https://doi.org/10.1016/j.iatssr.2016.07.001

Firpo S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276. https://doi.org/10.1111/ecta.2007.75.issue-1

Fisher R. A., & 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. https://doi.org/10.1017/S0305004100015681

Freedman D. A. (2010). *Statistical models and causal inference: A dialogue with the social sciences*. Cambridge University Press.

Frölich M., & Melly B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, *31*(3), 346–357. https://doi.org/10.1080/07350015.2013.803869

Gandy A., Jana K., & Veraart A. E. D. (2022). Scoring predictions at extreme quantiles. *AStA Advances in Statistical Analysis*, *106*, 527–544. https://doi.org/10.1007/s10182-021-00421-9

Gangl M. (2010). Causal inference in sociological research. *Annual Review of Sociology*, *36*(1), 21–47. https://doi.org/10.1146/soc.2010.36.issue-1

George K. (1970). Transportation compatible land uses and bus stop location. *WIT Transactions on the Built Environment*, *41*, 459–468. https://doi.org/10.2495/UT990431

Gissibl N., & Klüppelberg C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, *24*(4A), 2693–2720. https://doi.org/10.3150/17-BEJ941

Gissibl N., Klüppelberg C., & Otto M. (2017). Tail dependence of recursive max-linear models with regularly varying noise variables. arXiv:1701.07351, preprint: not peer reviewed.

Gnecco N., Meinshausen N., Peters J., & Engelke S. (2021). Causal discovery in heavy-tailed models. *Annals of Statistics*, *49*(3), 1755–1778. https://doi.org/10.1214/20-AOS2021

Gomes M. I., & Guillou A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, *83*(2), 263–292. https://doi.org/10.1111/insr.v83.2

G'Sell M. G., Wager S., Chouldechova A., & 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. https://doi.org/10.1111/rssb.12122

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. https://doi.org/10.1016/0047-259X(90)90080-2

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

Hannart A., & Naveau P. (2018). Probabilities of causation of climate changes. *Journal of Climate*, *31*(14), 5507–5524. https://doi.org/10.1175/JCLI-D-17-0304.1

Hosking J. R. M., & Wallis J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, *29*(3), 339–349. https://doi.org/10.1080/00401706.1987.10488243

Hudgens M. G., & Halloran M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, *103*(482), 832–842. https://doi.org/10.1198/016214508000000292

Imbens G. W., & Rubin D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jin J., & Rafferty P. (2017). Does congestion negatively affect income growth and employment growth? Empirical evidence from us metropolitan regions. *Transport Policy*, *55*, 1–8. https://doi.org/10.1016/j.tranpol.2016.12.003

Kang J. D. Y., & 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. http://doi.org/10.1214/07-STS227

Karmakar N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, *4*(4), 373–395. https://doi.org/10.1007/BF02579150

Koenker R. (2005). *Quantile regression*. Cambridge University Press.

Langousis A., Mamalakis A., Puliga M., & 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. https://doi.org/10.1002/2015WR018502

Li H., Graham D. J., & Liu P. (2017). Safety effects of the London cycle superhighways on cycle collisions. *Accident Analysis and Prevention*, *99*, 90–101. https://doi.org/10.1016/j.aap.2016.11.016

Litvinova S., & Silvapulle M. J. (2020). Consistency of full-sample bootstrap for estimating high-quantile, tail probability, and tail index. arXiv:2004.12639, preprint: not peer reviewed.

Liu Y., & Wu Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics*, *23*(2), 415–437. https://doi.org/10.1080/10485252.2010.537336

Mhalla L., Chavez-Demoulin V., & Dupuis D. J. (2020). Causal mechanism of extreme river discharges in the upper Danube basin network. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *69*(4), 741–764. https://doi.org/10.1111/rssc.12415

Moodie E. E. M., Saarela O., & Stephens D. A. (2018). A doubly robust weighting estimator of the average treatment effect on the treated. *Stat*, *7*(1), e205 https://doi.org/10.1002/sta4.205

Moodie E. E. M., & Stephens D. A. (2022). Causal inference: Critical developments, past and future. *The Canadian Journal of Statistics*, *50*(4), 1299–1320. https://doi.org/10.1002/cjs.11718

Naveau P., Hannart A., & Ribes A. (2020). Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and its Application*, *7*(1), 89–110. https://doi.org/10.1146/statistics.2020.7.issue-1

Norman W. (2017). Bike lanes don't clog up our roads, they keep London moving. *The Gaurdian*. https://www.theguardian.com/environment/bike-blog/2017/dec/01/bike-lanes-dont-clog-up-our-roads-they-keep-london-moving#:~:text=What%20the%20new%20report%20affirms,experience%20over%20the%20coming%20decades

Pickands J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*(1), 119–131. https://doi.org/10.1214/aos/1176343003

Reid C. (2021). None of top five congested roads in U.K. feature adjacent cycleways. *Forbes*. https://www.forbes.com/sites/carltonreid/2021/12/07/none-of-top-five-congested-roads-in-uk-feature-adjacent-cycleways/?sh=757d25032d1f

Retallack A., & Ostendorf B. (2019). Current understanding of the effects of congestion on traffic accidents. *International Journal of Environmental Research and Public Health*, *16*(18). 3400. http://doi.org/10.3390/ijerph16183400

Ribes A., Thao S., & Cattiaux J. (2020). Describing the relationship between a weather event and climate change: A new statistical approach. *Journal of Climate*. *33*(15), 6297. http://doi.org/10.1175/JCLI-D-19-0217.1

Rosenbaum P., & Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effec. *Biometrika*, *40*(1), 41–55. https://doi.org/10.2307/2335942

Scarrott C., & MacDonald A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal*, *10*(1), 33–60. https://doi.org/10.57805/revstat.v10i1.110

Schnabel S. K., & Eilers P. H. C. (2013). Simultaneous estimation of quantile curves using quantile sheets. *Advances in Statistical Analysis*, *97*(1), 77–87. https://doi.org/10.1007/s10182-012-0198-1

Slawson N. (2017). Traffic jams on major UK roads cost economy around £9 bn. *The Gaurdian*. https://www.theguardian.com/world/2017/oct/18/traffic-jams-on-major-uk-roads-cost-economy-around-9bn

Transport for London. (2010). *Cycling revolution London*.

Transport for London. (2011). *Barclays cycle superhighways evaluation of pilot routes 3 and 7*.

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025

UN Environment Programme. (2019). *Cycling, the better mode of transport*. https://www.unep.org/news-and-stories/story/cycling-better-mode-transport.
United Nations. (2021). *World bicycle day*. https://www.un.org/en/observances/bicycle-day.
Xian X., Ye H., Wang X., & Lie K. (2021). Spatiotemporal modeling and real-time prediction of origin-destination traffic demand. *Technometrics*, *63*(1), 77–89. https://doi.org/10.1080/00401706.2019.1704887
Xu D., Daniels M. J., & Winterstein A. G. (2018). A Bayesian nonparametric approach to causal inference on quantiles. *Biometrics*, *74*(3), 259–276. http://doi.org/10.1111/biom.12863
Xu L., & Nusholtz G. (2017). Application of extreme value theory to crash data analysis. *The Stapp Car Crash Journal*, *61*, 287–298. https://doi.org/10.4271/2017-22-0011
Zhang K., Sun D. J., Shen S., & 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. https://doi.org/10.5198/jtlu.2017.954
Zhang N., & Bansal P. (2021). A causal inference approach to measure the vulnerability of urban metro systems. *Transportation*. *48*(6), 3269. http://doi.org/10.1007/s11116-020-10152-6
Zhang Y. (2018). Extremal quantile treatment effects. *Annals of Statistics*, *46*(6B), 3707–3740. https://doi.org/10.1214/17-AOS1673
Zhang Z., Chen Z., Troendle J. F., & Zhang J. (2012). Causal inference on quantiles with an obstetric application. *Biometrics*, *68*(3), 697–706. https://doi.org/10.1111/j.1541-0420.2011.01712.x
Zheng L., & Sayed T. (2019). Application of extreme value theory for before-after road safety analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(4), 1001. http://doi.org/10.1177/0361198119841555
Zou H., & Yuan M. (2008). Regularized simultaneous model selection in multiple quantiles regression. *Computational Statistics & Data Analysis*, *52*(12), 5296–5304. https://doi.org/10.1016/j.csda.2008.05.013

Downloaded from https://academic.oup.com/jrsssc/article/72/5/1452/7263180 by Ahmedabad University user on 09 January 2025