Revisiting the empirical fundamental relationship of traffic flow in highways using a causal econometric approach

Anupriya*, Daniel J. Graham, Daniel Hörcher, Prateek Bansal

Transport Strategy Centre
Department of Civil and Environmental Engineering
Imperial College London
South Kensington Campus
London SW7 2AZ

Short Paper for the hEART conference 2020

Abstract

The fundamental diagram of traffic flow, or in other words, the vehicular flow-speed or flow-density relationship, for a highway section is determined by both physical characteristics of the section and attributes of the population of drivers and vehicles travelling through it. The empirical estimation of this relationship by fitting a regression curve to a cloud of observations of traffic variables may not be robust as the relationship may suffer from confounding from observed and unobserved attributes of driving behaviour and vehicular characteristics. This paper adopts a causal econometric approach to obtain a more reproducible characterisation of the fundamental relationship. We relax the strict stationary state assumption and instead focus on the cause-effect interactions between flow and speed to model the spatio-temporal variations in traffic at a macroscopic level. We use traffic data from a highway section in California. We adopt a fully-flexible non-parametric specification for the relationship and apply instrumental variables estimation to control for the aforementioned confounding bias. We deliver some new empirical insights into the macroscopic level dynamics of traffic flow in a highway section alongside verifying some existing ones.

Keywords: fundamental relationship of traffic flow, endogeneity, causal econometric modelling, non-parametric instrumental variables estimation

---

*Corresponding author. Email address: anupriya15@imperial.ac.uk

1. Introduction

The standard engineering relationship between vehicular flow q, that is, the number of vehicles passing a given point per unit time, and density k, that is, the number of vehicles per unit distance in a highway section, as shown in Figure 1, commonly known as the fundamental relationship of traffic flow[1], is based on the assumption that traffic conditions along the section are stationary, which means that the three key traffic variables, q, k and v, are the same at each and every point in the highway section (Daganzo, 1997). This assumption follows from the engineers' interest in a general relationship to characterise the flow of traffic in a given facility (Daganzo, 1997).

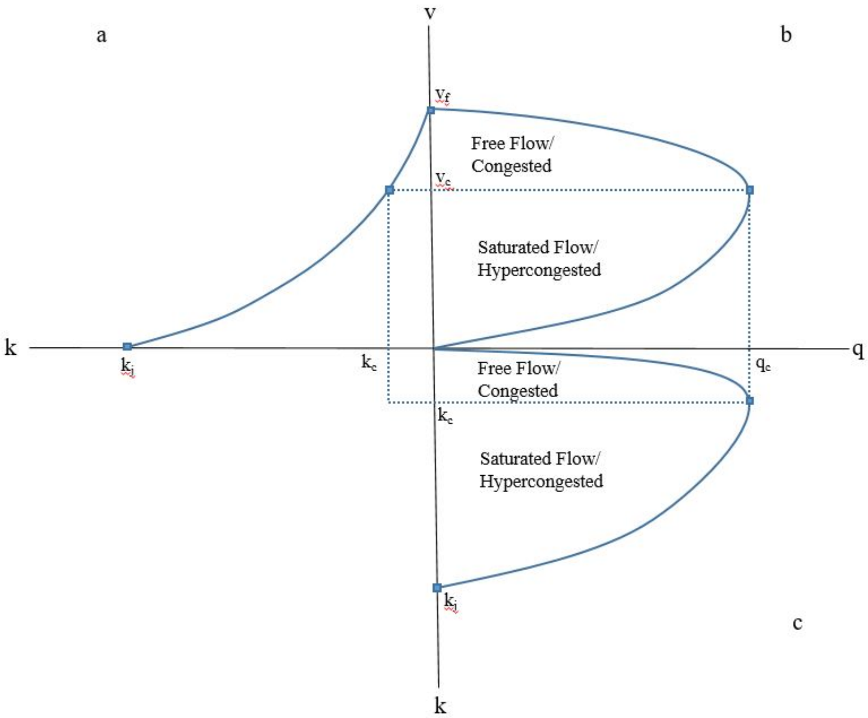

Figure 1: The fundamental diagram of traffic flow (adopted from Small et al. (2007)).

Consequently, the relationship is estimated empirically by pooling observations from different cross-sections along the highway section and from different time-periods and fitting a regression curve to the point cloud. Engineers believe the estimated relationship to be a property of the road section, the environment and the population of travellers as on an average, drivers show the same behaviour under same average conditions (Da-

[1]This relationship can be equivalently expressed as flow-speed or speed-density relationship using the average vehicle speed v in the highway section, as shown in Figure 1.

ganzo, 1997). The estimated relationship, however, is only *associational* and not *causal* due to lack of control for persisting uncertainties about the data generating process, for instance, observed and unobserved time-invariant and time-variant characteristics of the drivers and vehicles. Control for unobserved confounding is important because such characteristics can vary from one person to another and from one situation to another, sometimes in complicated or even unexpected ways, and we are interested in a more robust characterisation of traffic flow in a highway section that is reproducible and is not sensitive to variations in driving characteristics.

In this paper, we adopt a causal econometric framework to estimate the fundamental relationship. Within this framework, we relax the strict stationary state assumption and instead follow a cause-effect approach to model the macroscopic dynamics of traffic flow in a highway section by giving due attention to the direction of causality in this relationship. We use traffic data from a highway section in California and apply a Bayesian non-parametric instrumental variables estimation that allows us to capture the non-linearities in the relationship with a fully flexible non-parametric specification and adjust for any confounding bias via the use of relevant and exogenous instruments as controls.

The main contribution of this research resides in developing a comprehensive understanding of the macroscopic level dynamics of traffic flow in a highway section within a causal econometric framework and determining a novel causal relationship between traffic flow and speed in a highway section.

Our proposed causal framework is different from the existing causal framework in the transportation economics literature, which is based on the demand-supply interpretation of the fundamental relationship. Under stationary state traffic conditions, economists suggest that the fundamental relationship represents the supply curve for travel in the road section (Walters, 1961), where users are suppliers of travel. Based on this supply interpretation of the fundamental relationship, a recent economics literature calls into question the hypercongested part of the fundamental diagram as shown in Figure 1 and in turn questions the applicability of traffic controls and congestion pricing in the absence of evidence indicating any drop in flow with increasing density or demand (Anderson and Davis, 2018). We argue that in developing a causal understanding of the fundamental relationship, the economists' representation of this model as a supply curve may lead to ambiguity. The fundamental relationship of traffic flow can be treated as equivalent to the supply curve for travel only under stationary state traffic conditions, however, existence of such conditions are highly improbable. Therefore, we do not find this economic approach to be the most useful framework for our empirical model.

2. Data

We make use of traffic data from a standard highway bottleneck located in the westbound direction of the California State Route 24 (SR-24) at the Caldecott Tunnel in Oakland, California. A schematic representation of this bottleneck is shown in Figure 2. The high-quality data is collected via a series of loop detectors installed at various locations along the highway, which measure traffic flow and vehicle speed averaged over every 5-minute duration. The data is maintained by the California Department of Transportation (Caltrans) and made publicly available through their Performance Measurement System (PeMS) website[2].

We use observations on the westbound traffic from weekdays in the summer months between June-August in 2005-2010 and between the time period 12:00 hours to 00:00 hours for this study. We only include those periods of observation when no infrastructure related shocks in form of lane closures or traffic incidents are present and when weather conditions are good and favourable for drivers. As the highway section is located well away from any major upstream or downstream intersections, we can assume that this section allows us to study the traffic dynamics arising solely from the presence of the bottleneck, without being affected by any upstream or downstream influences.

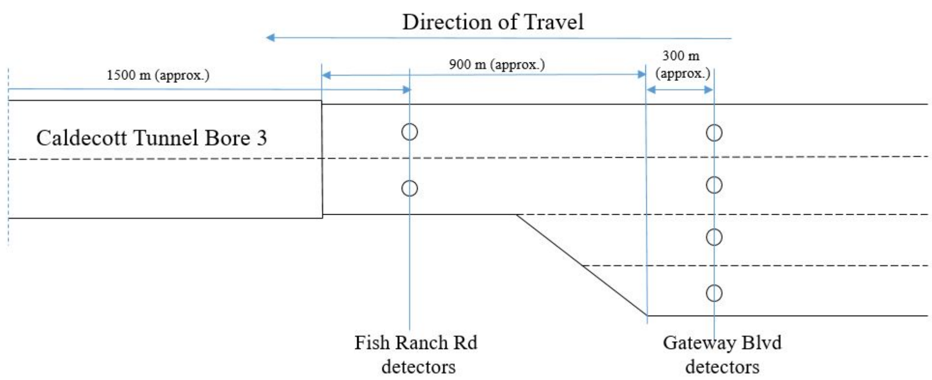

Figure 2: Highway Bottleneck in the westbound SR-24 at Caldecott Tunnel in California.

[2]Performance Measurement System (PeMS) website: `http://pems.dot.ca.gov/`

3. Methodology

3.1. The Direction of Causality in the Fundamental Relationship

According to the engineering literature, the direction of causality in the fundamental relationship, that is whether flow affects speed or vice-versa, could be both ways (Daganzo, 1997). For instance, if traffic enters the upstream end of a highway at rate q_u until a stationary state develops downstream, then the downstream space-mean speed v_d should be a consequence of the input flow q_u and the behaviour of drivers as they interact with one another while passing (Daganzo, 1997). The downstream density k_d would also be a consequence of q_u. In this case, causality comes from upstream (refer figure 3). On the other hand, if a stationary state develops behind a slow-moving obstruction to traffic or any other bottleneck, then through the reproducible behaviour of drivers we would expect the average spacing inside the queue upstream of the obstruction, and therefore k_u and q_u to be a consequence of the obstruction's speed v_o (Daganzo, 1997). In this case, causality comes from downstream (refer figure 4).

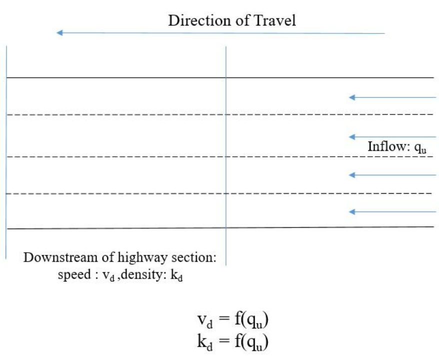

Figure 3: Causality comes from upstream.

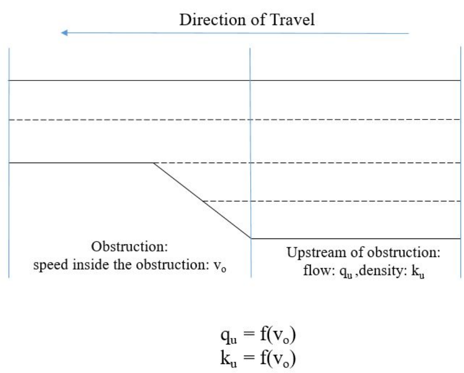

Figure 4: Causality comes from downstream.

For our highway section with a downstream bottleneck, we consider two possible scenarios: first, before activation of the bottleneck, causality becomes from upstream and second, after the bottleneck is triggered, causality comes from downstream.

3.2. Model Specification

3.2.1. Before activation of the bottleneck

In this case, we consider that the causality comes from the upstream and the inflow into the highway section controls the macroscopic traffic dynamics inside the section. We consider the flow through the bottleneck, q^b_{it}, and the speed inside the bottleneck,

v_{it}^b, in the five-minute interval i, $i = 1,..,N$, on a particular day t, $t = 1,...,T$, to be a function of the inflow into the highway section, q_{it}^u, conditional on the properties of the infrastructure, the environmental conditions and the average behaviour of drivers and vehicles.

$$\begin{aligned} q_{it}^b &= f(q_{it}^u) + \delta_{it} + \xi_{it} \\ v_{it}^b &= g(q_{it}^u) + \omega_{it} + \psi_{it} \end{aligned} \tag{1}$$

where δ_{it} and ω_{it} are the unobserved (to researchers) traffic specific behavioural component common to all drivers or any traffic specific vehicular characteristic common to all vehicles or any weather specific component affecting the entire traffic stream and ξ_{it} and ψ_{it} represent normally distributed idiosyncratic error terms representing all random shocks to the dependent variable. The exact structural form of how q_{it}^u enters the two equations is unknown, so we adopt non-parametric specifications $f(.)$ and $g(.)$. We expect δ_{it} and ω_{it} to be correlated with q_{it}^u and introduce an upward bias in the estimated relationships as we expect the direction of correlation between δ_{it} and ω_{it} and q_{it}^u to be the same as that between q_{it}^b or v_{it}^b and δ_{it} and ω_{it}. As an example, for an average population of risk-taking drivers we expect a positive correlation between high risk taking behaviour and speeds as well as high risk taking behaviour and flows.

3.2.2. After activation of the bottleneck

In this case, we consider that the causality comes from the downstream and the speed inside the bottleneck controls the macroscopic traffic dynamics inside the section. We consider both inflow into the highway section, q_{it}^u, and flow through the bottleneck, q_{it}^b to be a function of the speed inside the bottleneck, v_{it}^b, conditional on the properties of the infrastructure, the environmental conditions and the average behaviour of drivers and vehicles.

$$\begin{aligned} q_{it}^b &= f(q_{it}^u) + \delta_{it} + \xi_{it} \\ v_{it}^b &= g(q_{it}^u) + \omega_{it} + \psi_{it} \end{aligned} \tag{2}$$

where δ_{it} and ω_{it} are the unobserved traffic or weather components and ξ_{it} and ψ_{it} are the random shocks to the dependent variable as explained in the previous case. Similar to the previous case, we expect an upward bias in the estimated relationships in the absence of control for the unobserved effects δ_{it} and ω_{it}.

3.3. Empirical Identification

To estimate equations 1 and 2, we adopt a Bayesian non-parametric instrumental variable approach proposed by Wiesenfarth et al. (2014) that allows us to correct for endogeneity bias in regression models where the covariate effects enter the model with unknown functional form. Bias correction relies on a simultaneous equations specification as shown in equation 3 and the joint error distribution is modelled flexibly via a Dirichlet process mixture prior. To account for nonlinear effects of continuous covariates, both the structural and instrumental variable equation (refer 3) are specified in terms of additive predictors comprising penalised splines. Efficient Markov chain Monte Carlo simulation techniques are employed for a fully Bayesian inference. The resulting posterior samples allow us to construct simultaneous credible bands for the non-parametric effects, including data-driven smoothing parameter selection. In addition, improved robustness properties, such as adjustment for outliers and extreme observations, are achieved due to the flexible error distribution specification.

We have a model with a single endogenous covariate, that is,

$$y = s(x) + \epsilon_2, \quad x = h(z) + \epsilon_1 \tag{3}$$

with response y, covariate x, and instrumental variable z and with effects of unknown functional form $s(.)$ and $h(.)$, respectively and random errors ϵ_2 and ϵ_1. Endogeneity bias arises if $E(\epsilon_2|\epsilon_1) \neq 0$. Then assuming the identification restrictions

$$E(\epsilon_1|z) \quad and \quad E(\epsilon_2|\epsilon_1, z) = E(\epsilon_2|\epsilon_1), \tag{4}$$

yields

$$\begin{aligned} E(y_2|y_1, z) &= s(x) + E(\epsilon_2|\epsilon_1, z) = s(x) + E(\epsilon_2|\epsilon_1) \\ &= s(x) + \nu(\epsilon_1), \end{aligned} \tag{5}$$

where the unobserved term in the first equation $\nu(\epsilon_1)$ is the control function.

Due to the absence of suitable exogenous instruments, we use lagged levels of endogenous covariates as their instruments, that is, for an endogenous covariate observed in the five-minute interval i on day t, we consider the observation on the covariate from the same interval i from the previous day $t-1$ as its instrument.

4. Results and Discussion

4.1. Activation of the bottleneck

To understand when the bottleneck gets activated, we study the variation of the upstream speed v_{it}^u over the speed inside the bottleneck v_{it}^b as shown in Figures 5, estimated via the Bayesian approach discussed in Section 3. We assume that the activation of the bottleneck will be followed by a significant fall in the upstream speed indicating the onset of queuing. We identify the value of the bottleneck speed at which the upstream speed drops to a constant speed indicating the movement of queue behind the bottleneck. We also carry out a Regression Kink Design analysis to test for any statistically significant change in slope of the v_{it}^u vs v_{it}^b line at the 99 percent confidence level. The results are presented in Table 1. We note that the bottleneck gets activated approximately when the speeds inside the bottleneck below to 54.9 mph.

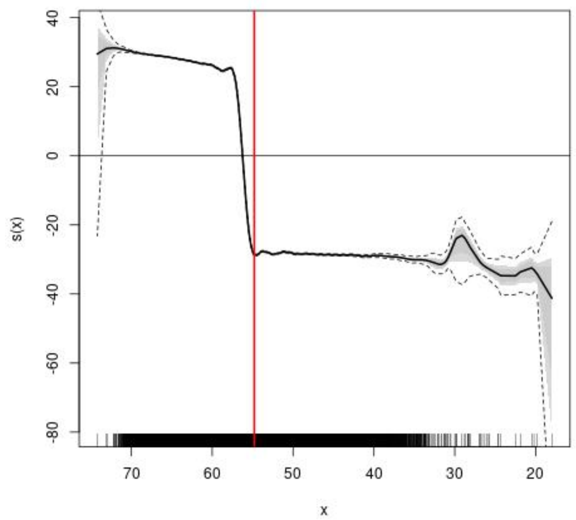

Figure 5: The estimated variation of upstream speed with speed inside the bottleneck [Estimated intercept = 44.10 (0.06)].

Table 1: Summary of results from Regression Kink Design analysis.

Speed inside	Change in slope of v_{it}^u vs v_{it}^b	
Bottleneck (in mph)	Coefficient	Standard Error
55.5	4.18	3.48
55.4	4.81	3.25
55.3	5.74*	3.28
55.2	6.79**	3.09
55.1	7.03**	3.32
55.0	7.02**	3.08
54.9	7.06**	3.01
54.8	8.11***	2.57
54.7	8.76***	2.56
54.6	8.80***	2.55

Significance: (*) 90 percent, (**) 95 percent, (***) 99 percent

4.2. Before activation of the bottleneck

We present the results from estimation of equation 1 using the Bayesian instrumental variables approach discussed in Section 3. We compare these results with a simple Bayesian penalised spline-based non-parametric regression without involving any instrumental variables.

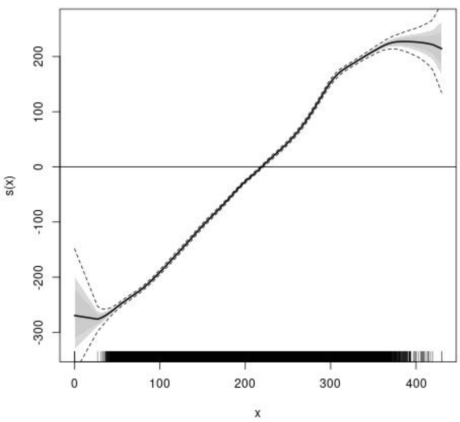

Figure 6: Estimated effect of inflow on flow through the bottleneck using instrumental variables [Estimated intercept = 370.55 (0.32)].

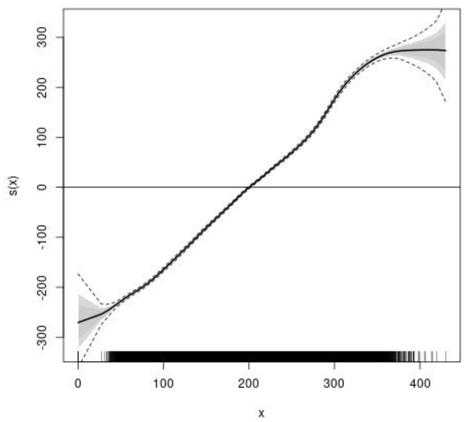

Figure 7: Estimated effect of inflow on flow through the bottleneck without instrumental variables [Estimated intercept = 341.41 (0.36)].

From 6 and 7, we do not note any noticeable differences in the two curves. We find that the flow through the bottleneck increases linearly with inflow. The estimated capacity of the bottleneck is around 640 vehicles per five minutes.

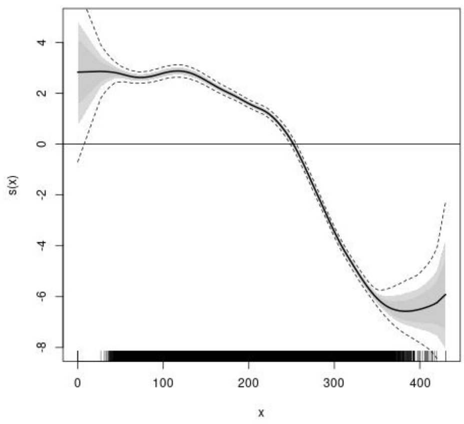

Figure 8: Estimated effect of inflow on speed inside the bottleneck using instrumental variables [Estimated intercept = 63.26 (0.01)].

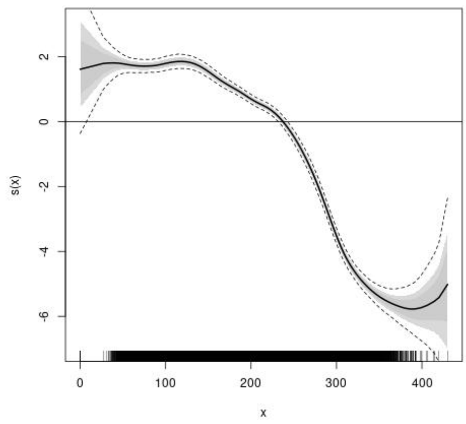

Figure 9: Estimated effect of inflow on speed inside the bottleneck without instrumental variables [Estimated intercept = 64.56 (0.01)].

From 8 and 9, we do not again note any noticeable differences in the two curves. We find that speeds inside the bottleneck fall at an increasing rate with inflow.

4.3. After activation of the bottleneck

We present the results from estimation of equation 2 using the Bayesian instrumental variables approach discussed in Section 3. We compare these results with a simple Bayesian penalised spline-based non-parametric regression without involving any instrumental variables.

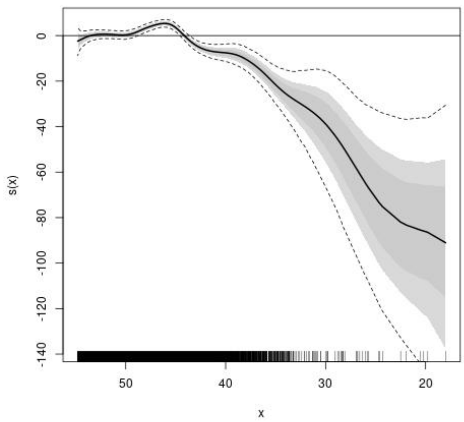

Figure 10: Estimated effect of speed inside the bottleneck on inflow using instrumental variables [Estimated intercept = 296.95 (0.18)].

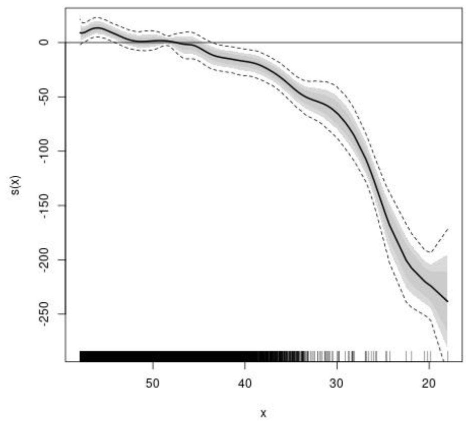

Figure 11: Estimated effect of speed inside the bottleneck on inflow without instrumental variables [Estimated intercept = 298.21 (0.22)].

From Figures 10 and 11, we note that with activation of the bottleneck follows, there is continuous decrease in inflow into the highway section, which primarily occurs due to queuing. Although we do not see any noticeable differences in the shapes of the two curves, we find that use of instrumental variables yields wider creditable bands when speed inside the bottleneck falls approximately below 35 mph. The bar on the X-axis denotes the number of observations for any value of X. We note that the use of instrumental variables allows for appropriate adjustment for any extreme observations or outliers. These extreme observations may be a result of any external shock which may be infrastructure related like incidents, or any weather related shock.

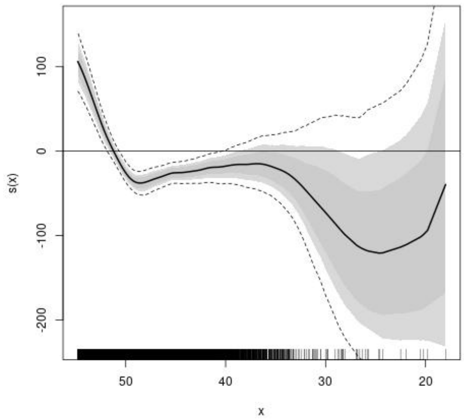

Figure 12: Estimated effect of speed inside the bottleneck on flow through the bottleneck using instrumental variables [Estimated intercept = 540.10 (0.32)].

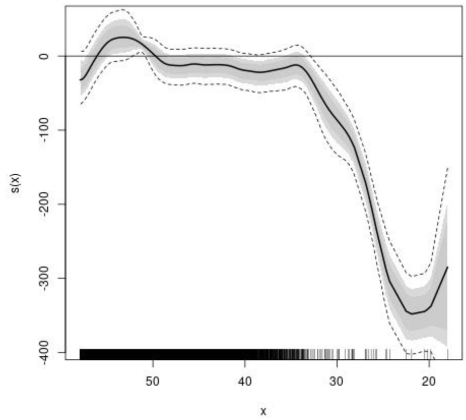

Figure 13: Estimated effect of speed inside the bottleneck on flow through the bottleneck without instrumental variables [Estimated intercept = 538.07 (0.40)].

From Figures 12 and 13, we note significant differences between the two curves. Where, the capacity drop estimated via the regression without instrumental variables is around 9 percent, with instrumental variables we estimate a capacity drop of about 20 percent followed by an eventual rise of around 3 percent. Thus, the net drop in capacity is around 18 percent. Thus, without the use of instrumental variables, there is an upward bias in estimated slope of the speed-flow curve which leads to underestimation of the capacity drop. From Figure 12, we also note the uncertainty in further drop in flows for bottleneck speeds below 35 mph. As noted previously, these observations may be related to infrastructure or weather related shocks. Without instrumental variables, we are unable to quantify this uncertainty. In our full paper, we present a full comparison of these results with the existing engineering literature.

5. Relevance and Future Work

This theme is important as much of the traffic flow theory depends on the existence of a fundamental relationship between flow, density and speed either explicitly as in the LWR hydrodynamic model or implicitly as in car following models. Moreover design of highway sections are based on the fundamental relationship as detailed in standard reference manuals like the HCM or the UK-CoBA. Thus, a more robust characterisation of this relationship is important. Furthermore, as this relationship forms the backbone of the highway pricing literature, therefore, the empirical findings from this study also have a high relevance for transportation economists.

Our future work comprises of finding equivalent parametric estimates to mimic the non-parametric functional forms for the fundamental relationship and to validating the results by applying the proposed framework on a different highway section. In our full paper, we also discuss the implications of the estimated relationships on the transportation economics theory on highway pricing.

References

Anderson, M. L. and Davis, L. W. (2018), 'Two empirical tests of hypercongestion', *National Bureau of Economic Research Program(s):Environment and Energy Economics, Public Economics* (w24469).
URL: *https://www.nber.org/papers/w24469*

Daganzo, C. F. (1997), *Fundamentals of transportation and traffic operations*, Vol. 30, Pergamon Oxford.

Small, K. A., Verhoef, E. T. and Lindsey, R. (2007), *The economics of urban transportation*, Routledge.

Walters, A. A. (1961), 'The theory and measurement of private and social cost of highway congestion', *Econometrica: Journal of the Econometric Society* pp. 676–699.

Wiesenfarth, M., Hisgen, C. M., Kneib, T. and Cadarso-Suarez, C. (2014), 'Bayesian non-parametric instrumental variables regression based on penalized splines and dirichlet process mixtures', *Journal of Business and Economic Statistics* **32**(3), 468–482.