Does Hypercongestion Exist?  
New Evidence Suggests Not

Michael L. Anderson Lucas W. Davis*

October 2018

Abstract

Transportation engineers are taught that as demand for travel goes up, this decreases not only speed but also the capacity of the road system, a phenomenon known as hypercongestion. We revisit this idea. There is no question that road systems experience periods in which capacity falls. However, we point out that capacity is determined by both demand and supply. Road construction, lane closures, stalled vehicles, weather, and other supply shocks provide an alternative explanation for the empirical evidence on hypercongestion. Using high-frequency data on traffic flows and vehicle speeds, we show that a naive regression recovers the standard hypercongestion result in the literature. However, once we use instrumental variables to isolate the effect of travel demand this effect disappears and across specifications we find no evidence that capacity decreases during periods of high demand. This lack of evidence of hypercongestion calls into question long-standing conventional wisdom held by transportation engineers and implies that efficient “Pigouvian” congestion taxes should be substantially lower than implied by supply curves exhibiting hypercongestion.

Key Words: Hypercongestion, Traffic Congestion, Capacity Drop, Speed, Ramp Meters
JEL: C36, H23, R41, R42, R48

* (Anderson) University of California, Berkeley; mlanderson@berkeley.edu (Davis) University of California, Berkeley; lwdavis@berkeley.edu. We are grateful to Gilles Duranton, Jonathan Hughes, Mark Jacobsen, Ian Parry, and Kenneth Small for helpful comments. Neither of us have received any financial compensation for this project, nor do we have any financial relationships that relate to this research.
1 Introduction

The relationship between the number of vehicles on the road and the speed at which they travel is fundamental to transportation and urban economics. To anyone who has driven in traffic, it is clear that traffic congestion decreases speed. But transportation engineers go further, arguing that as demand for travel goes up, this decreases not only speed but also the capacity of the road system, a phenomenon known as hypercongestion. Many studies claim to document what has been called “capacity drop”, “flow breakdown”, or the “two capacity phenomenon” in the context of bottlenecks (Banks, 1990, 1991; Hall and Agyemang-Duah, 1991; Bertini and Malik, 2004; Persaud et al., 1998; Cassidy and Bertini, 1999; Zhang and Levinson, 2004; Chung et al., 2007; Oh and Yeo, 2012).

Our paper revisits this idea. There is no question that road systems experience periods in which capacity falls. However, we point out that capacity is determined by both demand and supply. Road construction, accidents, lane closures, disabled vehicles, weather, and other supply shocks provide an alternative explanation for the empirical phenomenon of hypercongestion. We argue this is equivalent to the standard problem of estimating a supply curve using market data on prices and quantities. Economists have long used instrumental variables in these settings to econometrically separate demand and supply (see, e.g., Angrist and Krueger, 2001). We use the same approach here.

Our empirical analysis uses high-frequency traffic data from the Caldecott Tunnel in Oakland, California. This bottleneck is a near-ideal locations for studying traffic
congestion both because the data are unusually high-quality and because this is a location where traffic delays are common; indeed, transportation engineers have frequently studied this exact location (Chin and May, 1991; Chung and Cassidy, 2002; Chung et al., 2007). We observe traffic speed and vehicle counts at several locations before, during, and after the bottleneck. We first show that naive regressions recover the standard hypercongestion result in the literature, with speed and traffic flows dropping simultaneously. However, once we use instruments to isolate the effect of travel demand, this relationship disappears, and across a variety of specifications we find no evidence that capacity decreases during periods of high demand.

This absence of evidence of hypercongestion calls into question long-standing conventional wisdom in the transportation engineering literature. Hypercongestion has long been, for example, the primary rationale for freeway metering lights and other traffic interventions aimed at regulating demand (Diakaki et al., 2000; Smaragdis et al., 2004; Cassidy and Rudjanakanoknad, 2005), but our evidence raises questions about the applicability of these interventions. Our results also imply that the efficient “Pigouvian” congestion taxes should be substantially lower than implied by supply curves exhibiting hypercongestion. Using a stylized simulation model we show that without hypercongestion the marginal damages from driving are about half as large.

Seminal studies by economists tend to take hypercongestion as given and derive welfare implications. For example, Walters (1961) takes a supply curve with hypercongestion from the transportation engineering literature and uses the parameters to derive efficient congestion prices. Similarly, Keeler and Small (1977) uses a sup-
ply curve with hypercongestion to develop a long-run model of highway pricing and investment. More recently, Hall (2018) uses a hypercongestion model to show how highway pricing can generate a Pareto improvement when agents are homogeneous, even before redistributing toll revenues.

Other studies provide theoretical models that explain how speed and traffic flow can drop simultaneously. Mun (1999) considers a bottleneck model that gives rise to hypercongestion. Small and Chu (2003) notes that the standard “engineering relationship” has a backward-bending region known as hypercongestion, and then presents bottleneck and street network models with this feature. Small (2015) explains that hypercongestion takes the form of “postulating that bottleneck capacity varies inversely with the length of the queue” (p. 115), though the application is in the context of downtown street networks.

Our paper is germane to a growing empirical literature on the economic costs of traffic congestion. Couture et al. (forthcoming) develops an econometric methodology for estimating city-level supply curves for trip travel, and constructs travel speed indices for large U.S. cities. Yang et al. (2018) measures the effect of traffic density on speed in Beijing, using variation in traffic from driving restrictions. Akbar and Duranton (2017) uses travel surveys and other data from Bogotá, Colombia to estimate the deadweight loss of traffic congestion. Adler et al. (2017) examines the effects of public transit strikes in Rome on arterial road congestion and concludes that hypercongestion, while rare, accounts for approximately 30 percent of congestion-related welfare losses.¹

¹Farther afield, there are also a number of studies by economists that examine the effect of
Before proceeding, we note two important caveats. First, our study focuses on freeways, not arterial street networks. Freeways are a vital component of the road network, accounting for the majority of vehicle miles traveled (Lomax et al., 2018). Indeed, all of the transportation engineering papers that we cite above focus on freeways. Freeway geometry, however, differs fundamentally from arterial road geometry because freeways lack intersections with conflicting cross traffic. Our results do not speak to whether hypercongestion occurs on a dense street network with conflicting directions of traffic. Second, our setting is a standard bottleneck in which multiple lanes feed into a lesser number of lanes. In more esoteric settings, a queue from a bottleneck on one route may spill over onto a different route that does not traverse the bottleneck, blocking that route and creating a “triggerneck” (Vickrey, 1969). Our results do not apply to triggernecks.

2 Conceptual Framework

The first panel of Figure 1 comes from the *Highway Capacity Manual*, the standard reference text in transportation engineering published by the Transportation Research Board, a division of the U.S. National Research Council (Transportation Research Board, 2016). The figure shows data on speed and flow rate from a freeway building highways on traffic congestion, suburbanization, and other outcomes (see, e.g. Baum-Snow, 2007; Duranton and Turner, 2011). Another strand of the economics literature on traffic congestion considers “bottleneck” models in which drivers face a tradeoff between time delays and when to leave for a trip (Vickrey, 1969; Small, 1982; Arnott et al., 1990, 1993, 1994). Finally, Hanna et al. (2017) show that elimination of high-occupancy vehicle lanes in Jakarta worsened traffic and Kreindler (2018) uses data from a smartphone app to study traffic congestion in Bangalore, India, finding at the city-level an approximately linear relationship between traffic volume and travel time.
in Los Angeles. With few vehicles on the road, travel is unconstrained, and average speeds are about 70 mph. Travel continues unconstrained at low demand levels, and then speeds modestly decrease as flow reaches a maximum capacity of about 2,000 vehicles per lane per hour.

What is surprising, however, is that the curve then appears to bend backward. The manual explains that further increases in travel demand generate “flow breakdown”, causing severe decreases in speed as well as decreases in freeway capacity, with flow rates falling well below the observed maximum. Transportation engineers are taught that this backward-bending curve is one of the “basic relationships” in traffic, and Panels 2 through 7 of Figure 1 present additional examples of speed-flow curves taken from the transportation literature. The manual attributes these low-speed, low-flow observations to “oversaturated flow” arising from excess demand.

In addition to the Highway Capacity Manual, other reference texts in transportation engineering describe traffic congestion similarly. For example, the chapter on “Congestion Modeling” in the Handbook of Transport Modeling explains that, “at higher densities the speed-density curve drops more rapidly, passing through the point at which flow reaches a maximum, and reaching zero at the jam density, where speed and flow are both zero.” The handbook then explains that this implies that, “The trip cost curve has a positively sloped portion corresponding to the congested branch of the speed-flow curve, and a negatively sloped backward-bending portion corresponding to the hypercongested branch.” (Lindsey and Verhoef, 2008). See also May (1990), p. 117 and Daganzo (1997), p. 87.

Studies have argued that even a small increase in demand can cause flow breakdown, “The reason is that there is a sharp discontinuity in the mode of traffic flow as traffic reaches capacity. A small increase in the traffic flow, typically resulting from an extra inflow of traffic from a junction, causes a transition from relatively free flow to congested flow; speed drops sharply, and so does total flow.” (Newbery, 1989).

In an article titled “Hypercongestion”, Small and Chu (2003) present a series of dynamic models for straight uniform highways and dense street networks in which transient demand surges cause long vehicle queues, resulting in large travel time increases. They stress the importance of studying hypercongestion using dynamic models, “Hypercongestion is a real phenomenon, potentially creating inefficiencies and imposing considerable costs. However, it cannot be understood within a steady-state analysis because it does not in practice persist as a steady state. Rather, hypercongestion occurs as a result of transient demand surges and can be fully analyzed only within a dynamic
We take a simple alternative view. Without question there are periods in which the capacity of a road system decreases. But we point out that supply shocks provide an important alternative explanation for low-speed, low-flow observations. Construction, lane closures, accidents, disabled vehicles, weather, and other supply shocks represent a “shift” in the curve, rather than a movement along the curve. Both demand and supply are shifting, so data on speed and flow rates should be interpreted as a locus of possible equilibria, rather than as a supply curve.

Our goal is to estimate the travel supply curve under congested conditions. Figure 2 a stylized travel supply curve. Following the convention in economics to put “price” on the $y$-axis, we have flipped the curve vertically relative to Figure 1, with the inverse of speed, rather than speed, on the $y$-axis. The dashed extension of the curve bends backward, as road capacity falls after demand reaches a certain level. We refer to a backward-bending supply curve as exhibiting hypercongestion. The solid extension of the curve has no backward-bending region; while travel costs can become high, additional demand does not compound the problem by also reducing capacity. Economists have long recognized that a backward-bending supply curve has important implications for welfare and the potential benefits of congestion pricing (see, for example, Walters (1961), Keeler and Small (1977), and Hall (2018)), and model.”

While the term “hypercongestion” appears frequently in the literature, there is confusion regarding its exact definition. Mun (1999), for example, notes, “There is no commonly used terminology to represent [traffic jams] among disciplines. Although traffic engineers use the term ‘congested flow,’ this is not appropriate here because ‘uncongested flow’ in engineering is still regarded as congested flow in an economic sense. On the other hand, some economists use the term ‘hypercongestion’...”, p. 323. Small and Chu (2003) defines hypercongestion as the region in the speed-flow diagram “in which speed increases with flow” and notes that it is “unsuitable as a supply curve for equilibrium analysis”, p. 319.
we confirm this in our simulation model. Our work thus seeks to identify whether the hypercongested region of the supply curve exists.\textsuperscript{6}

The final panel of Figure 1 plots data on speed and traffic flows using observations from our study location when the bottleneck is in effect. The basic pattern is similar to the preceding panels in the figure; speed falls with traffic flow. However, there are also many low-speed, low-flow observations which make the curve appear to bend backward. In the sections that follow, we show that these observations are the result of supply shocks. Using standard two stage least squares (2SLS) regressions, we empirically separate these “shifts” from movements along the curve. In particular, we instrument for demand using time-of-day. We argue that this instrument is a plausibly exogenous demand shifter — correlated with demand but not with supply — which allows us to hold supply factors fixed while tracing out the causal impact of travel demand on freeway capacity.

3 Empirical Application

3.1 Data

Our empirical application examines westbound traffic along California State Route 24 (SR-24) at the Caldecott Tunnel. SR-24 connects suburban Contra Costa County,

\textsuperscript{6}We contrast the actual travel supply curve with the speed-flow diagrams in Figure 1. It is an empirical fact that low-speed, low-flow observations exist, and in some contexts economists refer to the existence of these observations as “hypercongestion.” Our interest, however, is in the causal relationship between travel demand and capacity, so we focus on whether the actual travel supply curve exhibits a hypercongested region.
to the east, with the urban cities of Oakland and San Francisco, to the west. This location is a classic bottleneck, with the number of lanes decreasing as traffic approaches the tunnel. During the study period the tunnel featured two reversible lanes that operated westbound in the morning and eastbound in the afternoon and evening. We focus on weekday afternoons and evenings from 2005 to 2010, a period and set of hours during which the Caldecott Tunnel was operated such that westbound vehicles had to merge from four lanes to two as they approached the tunnel.\textsuperscript{7}

Figure 3 shows the study location. Approximately 3,000 feet before the tunnel the number of lanes merges from four down to two. This is the key feature of our study location and the place where the vehicle queue typically begins. The figure also shows using small circles the location of 30 “loop detectors”, our primary source of data on both traffic flows and average vehicle speed.\textsuperscript{8} We observe a set of two loop detectors

\textsuperscript{7}Rather than a single wide tunnel, the Caldecott consists of multiple “bores”, each with two lanes carrying traffic in a single direction. Although the tunnel was expanded to four bores (eight total lanes) in 2013, we study the period from 2005 to 2010 when the tunnel still had only three bores and construction had not yet begun on the fourth bore. During the period we study, the middle bore was operated westward during morning hours as commuters drive toward Oakland and San Francisco, and eastward during afternoon and evening hours as commuters drive toward suburban Contra Costa County. We focus on weekday afternoons because this is when the middle bore was operated eastward. The middle bore also operated eastward during some weekend hours. However, we do not have reliable documentation of these hours, so our analysis focuses on weekdays. Afternoon westbound traffic is lighter than eastbound traffic, but with only a single bore open in that direction, the bottleneck was more than sufficient to generate significant traffic delays. Finally, we do not use the eastbound morning bottleneck in our analysis because it features traffic merging from multiple directions, making it impossible to measure queue length, which we use later as a proxy for demand.

\textsuperscript{8}Loop detectors are small insulated electric circuits installed in the middle of traffic lane. Loop detectors measure the rate at which vehicles pass, e.g. vehicles per five-minute period. In addition, loop detectors measure average vehicle speed by sensing how long it takes each vehicle to pass over the detector. These loop detectors are maintained by the California Department of Transportation (Caltrans), and data are made publicly available through the Performance Measurement System.
after the merge but before the tunnel, as well as a series of loop detectors upstream of the merge. For westbound travelers there are no reasonable alternatives to the tunnel.

Before introducing our main empirical models we reproduce the hypercongestion phenomenon from the existing literature using several “naive” regressions. In our data there are clear positive relationships between vehicle speed and traffic flows in both the times series and cross section. We first estimate a time series regression using observations every five minutes at Gateway Blvd during weekday daytime hours. Regressing traffic flow on observed speed we find that a 50% reduction in speed correlates with a 4% drop in total traffic flow across all lanes ($t = 14.8$). In the context of our empirical application, this relationship has little to do with “oversaturated flow” or excess demand. Rather, it is an artifact of the daily supply shock — reversible lane closures — that simultaneously lowers traffic flow and decreases speeds. This is a somewhat extreme case, but it illustrates how a shifting supply curve can easily generate a positive correlation between speed and traffic flows, causing it to appear as if the supply curve bends backwards.

Moreover, even restricting the sample to include only afternoon and evening hours when the bottleneck is always present, there is still a pronounced positive cross-sectional relationship between vehicle speeds and traffic flows. For this next “naive”

(9) For visual clarity the figure does not include freeway exits and entrances. One of the significant advantages of the study location is that there are relatively few exits and entrances nearby. The last freeway entrance prior to the bottleneck is approximately 9,000 feet (1.7 miles) west of the tunnel.

(PeMS) at http://pems.dot.ca.gov/. Subsequent to our sample dates, the Gateway Blvd exit was renamed Wilder Rd.
regression, we take speed and traffic flow measurements from six lanes – two down-
stream (Fish Ranch) and four upstream (Gateway) of the bottleneck, generating six
observations for each five-minute period. When we then regress traffic flows on speed
while including fixed effects for every five-minute period in the sample (thus isolating
cross-sectional variation), we find that a 50% reduction in speed correlates with a
21% decrease in traffic flows ($t = 48.3$). This relationship is nearly mechanical —
each of the four lanes upstream of the bottleneck must process fewer vehicles per
minute, at slower speeds, than each of the two lanes downstream of the bottleneck.
These results are not causal relationships, but they highlight the ease with which a
casual observer might conclude that hypercongestion is present at our location.

These “naive” regressions also highlight that traffic flows and capacity are not the
same thing. Capacity is the maximum possible traffic flow, so it may differ signifi-
cantly from actual traffic flows during off-peak periods. Accordingly, in the analyses
that follow, we restrict the sample to include only observations from periods in
which a queue has formed. By definition, if there is no queue, then traffic flows have
not reached their maximum capacity, and measurements are limited by insufficient
demand rather than maximum capacity. We refer to this condition as “demand star-
vation” — the bottleneck could process more vehicles were they available. Thus, in
the regression analyses below we compare traffic flows between periods with different
lengths of non-zero queues. We refer to traffic flows during these periods as “capac-
ity”, and ask whether increases in travel demand cause capacity to decrease.
3.2 Summary Statistics

The unit of observation in our data is a five-minute period. The sample includes all weekday observations during which a queue is present between 4 pm and 7 pm and June 2005 and June 2010 (Section 3.3 discusses the choice of 4 pm to 7 pm in more detail). We determine whether or not there is a queue and measure the length of the queue using the loop detectors upstream of the bottleneck. The first upstream loop detector, Gateway, is 1,690 feet before the bottleneck, and we observe approximately equidistant detectors all the way to St. Stephens East, which is 15,690 feet (about three miles) away from the bottleneck. For each loop detector we determine whether a queue is present by measuring whether traffic is moving at under 30 miles per hour (mph).10 Thus, we define a queue as present if the average speed at Gateway is below 30 mph. The length of the queue is then measured using the number of consecutive upstream detectors for which we observe a delay.11

Appendix Table A1 reports summary statistics for our estimation sample.12 We measure speed and capacity downstream of the bottleneck using the loop detector at Fish Ranch Road. This loop detector is after traffic has merged from four to two lanes and cleared the bottleneck, but it is still east of the Caldecott Tunnel. We measure speed upstream of the bottleneck using the loop detector at Gateway.

---

10This threshold is arbitrary, but we show the results are robust to alternative definitions.
11For example, if the measured speed is below 30 mph at the first detector (Gateway), but not at the second (Orinda West), then we conclude that the queue is 1,690 feet in length. “Broken” queues (e.g. a case in which traffic moves below 30 mph at Gateway, above 30 mph at Orinda West, and below 30 mph at Camino Pablo West) are rare in our estimation sample, and our results are robust to their inclusion or exclusion.
12Appendix Table A2 reports summary statistics for the full weekday sample, including periods with no queue.
It is worth making several observations about the summary statistics. First, as expected, average speed declines significantly at the bottleneck. Downstream of the bottleneck the average speed is over 40 mph, but upstream the average speed is only 11 mph. Second, the average measured queue length is about 4,000 feet (0.8 miles), but the range is very wide, and at times the queue reaches as long as 15,700 feet. Third, there is also wide variation in capacity measured downstream of the bottleneck, ranging from zero to 408 vehicles per five minutes. The low-capacity observations are striking given that these are all periods with a non-zero queue, and, we argue, are due primarily to supply shocks (e.g. lane closures from accidents).

3.3 Estimation Sample

Avoiding observations with demand starvation is a key statistical challenge — it is impossible to measure a bottleneck’s capacity if it has not reached capacity. A simple strategy is to condition the estimation sample on the presence of a queue, regardless of time of day. This sample selection criterion, however, can introduce subtle but meaningful bias into an analysis that leverages exogenous variation in demand.

Consider periods of low demand. These periods typically have no queues, and if a queue is present it tends to be short. But if demand is low, why does a queue form?

---

13 Even during the most congested times of day, the queue achieves a length of three miles less than one percent of the time.
14 Observations with zero vehicles in a five-minute period account for less than 0.3% of observations in our main estimation sample. Dropping these zero-vehicle observations has no impact on our estimates.
One possibility is a negative supply shock, such as an accident or poor weather. When conditioning on a queue existing, negative supply shocks will thus be more common during periods of low demand than during periods of high demand, as supply shocks are necessary for the formation of queues during low demand periods. Thus, the sample selection criterion itself can generate a spurious positive relationship between capacity and queue length.

To mitigate this bias we focus on periods when demand should be sufficient to generate a queue even in the absence of a negative supply shock. Appendix Figure A1 plots the share of weekdays that a queue is present at each time of day. Visual inspection suggests that we focus on the period from 4 pm to 7 pm, which comprises our baseline estimation sample. To further narrow our sample, in some specifications we focus on times of day when there is at least a 75% chance of a queue (4:30 pm to 6:15 pm) and days of week with the highest frequency of early evening queues (Wednesday, Thursday, and Friday).\textsuperscript{15}

In our most restrictive samples we further limit the sample to include only days predicted to be high demand using a machine-learning model. Specifically, we predict whether there is a continuous queue from 4:30 pm to 6:15 pm as a function of month-of-sample indicators, year-by-day-of-week indicators, calendar-month-by-day-of-week indicators, and third-order polynomials in hourly inflows (well upstream of the bottleneck) for each hour from 12 pm to 4 pm. In total we have 105 predictors. Given the large number of predictors, overfitting is a serious concern, so we esti-

\textsuperscript{15}At 5 pm, there is a 92% chance of observing a queue on a Wednesday, Thursday, or Friday. In comparison, there is only a 32% chance of a queue at 5 pm on Mondays, and a 76% chance of a queue at 5 pm on Tuesdays.
mate our regression using LASSO. The LASSO estimates non-zero coefficients for 17 predictors, the most important of which is, unsurprisingly, inflows at 4 pm.

Figure 4 presents a visual depiction of our empirical strategy using data from our baseline estimation sample. The solid line plots average queue length (for non-zero queues) by time-of-day from 4 pm to 7 pm. The increase in queue length in the early evening and the subsequent decrease later strongly suggest that demand varies with time-of-day. The solid line is thus visual evidence of a strong “first-stage” relationship between queue length and time-of-day. The dashed line plots average capacity (traffic flows) by time-of-day (again conditional on a non-zero queue). Its flat profile is evidence of no “reduced-form” relationship between capacity and time-of-day. Taken together the two lines suggest that increases in traffic demand are not accompanied by decreases capacity, calling into question the hypercongestion hypothesis.

3.4 Main Regression Estimates

Table 1 reports our main results. We show estimates and standard errors from four two stage least squares (2SLS) regressions of the following form,

\[ \text{capacity}_t = \alpha + \beta \cdot \text{queue length}_t + \varepsilon_t. \]  

(1)

The dependent variable in all regressions is capacity, measured downstream of the bottleneck in vehicles per five minutes. The independent variable of interest is queue
length, measured in thousands of feet. The coefficient of interest is \( \beta \), which is the change in capacity associated with a one-unit increase in queue length. Motivated by Figure 4, we instrument for queue length using a third-order polynomial in time-of-day. Time-of-day is highly predictive of travel demand and the instrument \( F \)-statistics from the first-stage regression are large across the four columns — 61, 71, 87, and 23, respectively — indicating that we do not have a weak instruments problem.

We consider four different specifications with increasingly stringent criteria for selecting the sample. The main sample in Column (1) includes all five-minute periods on weekdays with a non-zero queue between 4 pm and 7 pm and June 2005 and June 2010. In Column (2) we restrict the sample to focus on Wednesdays through Fridays between 4:30 pm and 6:15 pm. We then further restrict the sample in Column (3) to include only days on which our LASSO model predicts a queue that persists from 4:30 pm to 6:15 pm with greater than 95% probability. Finally, in Column (4) we restrict the sample to summer months only, when traffic is least likely to be affected by adverse weather and lighting conditions.

The 2SLS procedure is equivalent to first regressing queue length on a third-order polynomial in time-of-day and then using the fitted values from that first-stage rather than queue length itself as the independent variable. Thus, the 2SLS regression is estimated using the predictable time-of-day variation in queue length rather than idiosyncratic day-to-day variation. This is desirable because the predictable time-of-day variation in queue length is mostly driven by differences in travel demand, while the day-to-day variation is also affected by construction delays, accidents, and
other supply shocks. The only supply shock evident to us that might correlate with time-of-day is lighting conditions. To check the sensitivity of our results to lighting conditions or inclement weather we restrict the sample to summer months only in the last column. In these months sunset generally occurs after 8 pm, total monthly precipitation never exceeds 0.1 inches, and fog is rarely present on summer afternoons.

Across specifications, there is no evidence of hypercongestion. If hypercongestion were present, we would expect the estimates of $\beta$ to be negative, with increased demand leading to a decrease in capacity. Instead, all four estimates of $\beta$ are positive and economically insignificant. In the specification in Column (1), each additional 1,000 feet of queue is associated with a 6.1 vehicle per five minute increase in capacity. This coefficient is small compared to mean capacity (350) but precisely estimated. Column (2) focuses on periods when demand is likely to be sufficient to generate a queue. The coefficient estimate drops to 2.7 and remains statistically significant (but still modest in comparison to mean capacity). Further restricting the sample to only include periods when our LASSO model predicts a queue or during summer months — Columns (3) and (4) — reduces the coefficient estimates to 1.2 and 2.3 and eliminates statistical significance. These estimates are precise enough to rule out

---

16An alternative to this instrumental variables strategy would be to exclude observations affected by supply shocks. This ends up being impractical, however, because many supply shocks are not observed. For example, the California Department of Transportation’s Performance Measurement System (PeMS) tracks major incidents but many smaller incidents are unreported.

17Fog in this area forms overnight and burns off by the late morning. The median sunset from June to August at our location occurs at 8:26 pm, and the earliest sunset occurs around 7:38 pm. Furthermore, the hill through which the Caldecott Tunnel bores has a peak of 1,400 feet, which is at least 500 feet higher than the summit of SR-24. Thus drivers never experience sunlight shining directly in their eyes, as the sun gets blocked by the hill at least 90 minutes prior to sunset.
capacity drops of more than 0.3% from each additional 1,000 feet of queue.

3.5 Additional Specifications and Robustness Checks

In the online appendix we present results from several alternative specifications and robustness checks. First, when we estimate these regressions using least squares, rather than 2SLS, the estimates are much smaller, close to zero, and not statistically significant. This pattern, reported in Appendix Table A3, is not surprising given that queue length is driven by both demand and supply shocks. Appendix Table A4 verifies that our results are not sensitive to the speed threshold for defining a queue or the presence of zero-flow observations.

One factor we are cannot directly control for is the composition of vehicles. In particular, the fraction of heavy trucks might vary sharply with time of day. While we lack high-frequency truck count data, we did obtain hourly estimates of truck counts for 2008, our sample midpoint, from the CalTrans PeMS site. We find that trucks comprise less than 5% of vehicles on SR-24.\(^\text{18}\) Furthermore, the share of trucks per hour is stable at 4 pm, 5 pm, and 6 pm (comprising 4.9%, 4.8%, and 4.6% of vehicles respectively).

Next we check whether large outliers, in either the positive or negative direction, can explain our results. We are specifically concerned about outliers in the negative direction, as these may represent accidents; if accidents occur disproportionally

\(^{18}\)The low truck counts are expected because trucks were restricted in response to a catastrophic tanker-truck accident in the tunnel in 1982. Furthermore, SR-24 is not a logical route for long-distance truckers.
around 4 pm or 7 pm, when queues are shorter, they could bias our 2SLS estimates. To test whether outliers drive our results we estimate least absolute values (LAV) regressions instead of least squares regressions. LAV regressions estimate the conditional median of the dependent variable, whereas OLS estimates the conditional mean. Median regressions are, by definition, insensitive to outliers. Appendix Table A5 reports estimates from two conditional quantile instrumental variables regressions in Columns (1) and (2) and two LAV regressions in Columns (3) and (4). In all cases the coefficient estimates are close to zero and statistically insignificant, providing no evidence of hypercongestion.

Finally we look for evidence of any discrete drop in capacity at the onset of a queue; perhaps it is the mere formation of a queue, not its length, that matters. Figure 5 plots an “event study” analysis in which time is normalized so that zero represents the queue’s approximate onset on each day. The figure includes both observed average capacity and simulated capacity under the assumption of a 10% capacity drop around the queue’s onset (a value consistent with the existing literature). The discordance between the two series indicates that the formation of a queue does not by itself generate a significant capacity drop.

4 Policy Implications

Thus, across specifications we find no evidence that capacity decreases during periods of high demand. This absence of evidence of hypercongestion calls into question long-standing conventional wisdom in the transportation engineering literature. In
this section, we argue that this finding also has important policy implications. In particular, we use a stylized simulation model to show how the presence (or lack) of hypercongestion affects the marginal damages from driving.

Traffic congestion is a negative externality (Pigou, 1920; Vickrey, 1963, 1969; Newbery, 1990). That is, when a motorist drives on a congested road, she decreases the average speed of all drivers on the road. Since drivers do not internalize the delays they impose on others, the efficient “Pigouvian” tax would tax drivers an amount equal to the value of the marginal delay they impose on other drivers.

We parameterize a stylized simulation model using data from our empirical application. Our objective is to describe traffic, queuing, and marginal damages of traffic congestion for a typical weekday afternoon and evening. As in the empirical analysis, we consider five-minute periods. We measure demand using traffic inflows observed on a representative day well upstream of the bottleneck at the Orinda West loop detector. The model determines queue lengths and delays for all drivers based on the maximum capacity of the bottleneck. For the simulation without hypercongestion, we set a maximum capacity of 352.2 vehicles per five minutes. This value calibrates the model so that the late afternoon queue starts around 3 pm, the maximum average queue length over a one-hour period reaches approximately 930 vehicles, and the queue dissipates around 7 pm. These features are consistent with the observed data for our representative day.

---

19 We choose the representative day to be as close as possible to the sample averages on three dimensions: fraction of time that a queue was present; average length of queue (when present); and average inflow of vehicles. This procedure selects February 7, 2006. However, our qualitative conclusions hold for any day on which there are nontrivial, but not interminable, queues.
We calculate marginal damages by introducing an additional 10 vehicles during a given hour of the day.\textsuperscript{20} One feature that the literature on bottleneck models emphasizes is that drivers may face a tradeoff between total delays and the time they choose to cross the bottleneck. To accommodate this tradeoff, following the introduction of additional vehicles we allow drivers to change their departure times in response to the change in the queue length. To determine whether it is optimal for a given driver to change her departure time, we follow Arnott et al. (1993) and assume that the cost of arriving one minute early is 60\% of the value of travel time, and the cost of arriving one minute late is 240\% of the value of travel time.\textsuperscript{21}

We calculate total delay (in minutes) both with and without the additional vehicles, so the difference gives us marginal damages in minutes. For the simulation, we repeat this exercise for 4 pm, 5 pm, and 6 pm, and then calculate the average marginal damages in minutes. It is a large number. For example, each additional vehicle introduced at 5 pm increases delays for other vehicles by a total of 149 minutes. If we value travel time at half the local median wage, this translates to damages of $26.67 (Parry and Small, 2009; Bureau of Labor Statistics, 2018).

To simulate hypercongestion we rely on the literature that has tended to find a “capacity drop” of approximately 10\% during periods of high demand (see the list of studies in the introduction as well as those summarized by Hall (2018)). In particular, we assume that capacity during each hour is approximately 366.3 multiplied by

\textsuperscript{20}10 vehicles corresponds to less than 10 seconds of traffic flow during normal conditions. We get similar results (per vehicle) if we introduce 100 vehicles, or just one vehicle.

\textsuperscript{21}Our results are robust to any reasonable choices for these parameter values, i.e., choices that are not trivially close to zero.
1 – (queue/9,000), where queue is measured in number of vehicles. The maximum queue size that the model, so parameterized, generates over a one-hour period is about 1,000 vehicles, which reduces the capacity to 366.3 * (0.89) = 326 vehicles per five minutes. The “intercept” of approximately 366.3 generates similar average queue length as a model with a capacity of 352.2 and no capacity drop (our baseline model). The comparison of the two models thus addresses the question of how introducing hypercongestion affects marginal damages while holding other features of the model approximately fixed.22

Table 2 summarizes the results. In the scenario described above, in which the maximum capacity drop reaches 10% with hypercongestion, marginal damages without hypercongestion are 53% smaller than marginal damages with hypercongestion.23 In a scenario in which the capacity drop reaches 15% with hypercongestion — a value that several transportation engineering studies have found — marginal damages without hypercongestion are 67% smaller than marginal damages with hypercongestion. That is, marginal damages are only about one-third as large. These stylized simulations could be refined further. For example, one could introduce heterogeneity in the value of time. The simulation nevertheless illustrates that welfare impacts depend

---

22To equate average queue length over the day, the intercept must be 366.3. If we instead wish to equate maximum queue length over a one-hour period, the intercept must be 367.6, and if we wish to equate queue length at 5 pm, the intercept must be 359.9. All of these parameterizations generate qualitatively similar conclusions, however; in all three cases the marginal damages without hypercongestion are half as large or less as marginal damages calculated under hypercongestion.

23Inserting additional vehicles at 4 pm instead of 5 pm increases the share of marginal damages generated by hypercongestion from 53% to 64%. Inserting additional vehicles at 6 pm instead of 5 pm reduces the share of marginal damages generated by hypercongestion from 53% to 42%. Inserting additional vehicles earlier in the queue increases the total number of vehicles affected (i.e. the total number of vehicles that follow the marginal vehicle), and inserting them later in the queue decreases the total number of vehicles affected.
strongly on whether hypercongestion exists.\textsuperscript{24}

\textbf{5 Conclusion}

This paper revisits the idea of hypercongestion. When we naively examine average speeds and traffic flows we find evidence consistent with a large literature on hypercongestion. But once we use standard econometric techniques to disentangle demand and supply, this backward-bending supply curve disappears. Across a variety of specifications we find no evidence of hypercongestion.

How can this be? To anyone who has been stuck in heavy traffic, it certainly feels as if the capacity of the roadway is being restricted in these moments. We suspect, however, that this feeling is largely about speed rather than capacity. There is no question that as more vehicles crowd onto the road, speed decreases. But speed and capacity are not equivalent. Speed is readily apparent to drivers, but capacity requires careful measurement.

In addition, on freeways the feeling of being trapped in heavy traffic usually occurs in a queue, waiting to pass a bottleneck. By definition the capacity \textit{per lane} when approaching a bottleneck must drop, as the number of lanes drops at the bottleneck. Nevertheless, we find that the capacity of the bottleneck itself — the rate at which vehicles pass through the bottleneck — does not drop with the length of the queue. In

\textsuperscript{24}This finding that welfare impacts depend strongly on whether hypercongestion exists is consistent with Fosgerau and Small (2013), who develop a variable-capacity bottleneck model with hypercongestion.
short, no matter how much travel demand increases, the number of vehicles passing through the bottleneck does not decrease.

Our findings imply that congestion taxes should be much lower than implied by supply curves exhibiting hypercongestion and that tolling or other demand management strategies, such as ramp meters, are unlikely to increase the capacity of the freeway system. Nevertheless, congestion taxes should not be zero. To the contrary, our simulation estimates show that, even without hypercongestion, the marginal damages from traffic congestion can be very large. Starting from zero — the level at which most roadways are currently taxed — leaves considerable headroom for increases.
References


Chin, Hong C and Adolf D May, “Examination of the Speed-Flow Relationship at the Caldecott Tunnel,” Transportation Research Record, 1991, 1320, 75–82.


Figure 1: Travel Supply Curves?

Sources: Transportation Research Board (2016), Drake and Schofer (1966), Keeler and Small (1977), Allen et al. (1985), Newbery (1989), May (1990), Ni (2015), and authors’ own calculations, respectively.
Figure 2: Stylized Travel Supply Curve

Cost per trip (1/speed)

Traffic flow

Notes: Dashed extension exhibits hypercongested (backward-bending) region; solid extension does not.

Figure 3: Traffic Merge Heading into Caldecott Tunnel

Notes: Figure approximately to scale.
Figure 4: Queue Length and Freeway Capacity by Time of Day

![Queue Length and Freeway Capacity by Time of Day](image)

Figure 5: Traffic Flows by Time of Queue Onset

![Traffic Flows by Time of Queue Onset](image)

Notes: Figure 5 plots average vehicle flows in the minutes before and after a queue begins on 118 days. Time is normalized so that the afternoon queue begins at time 0 on each day. The solid line plots average observed capacity (with dotted lines representing a 95% confidence interval), and the dashed line plots what average observed capacity would look like if there were a capacity drop of 10% at queue onset.
Table 1: The Effect of Travel Demand on Freeway Capacity, 2SLS

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Capacity</th>
<th>(2) Capacity</th>
<th>(3) Capacity</th>
<th>(4) Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queue length (1000s of feet)</td>
<td>6.051***</td>
<td>2.665***</td>
<td>1.212</td>
<td>2.253</td>
</tr>
<tr>
<td></td>
<td>(0.712)</td>
<td>(0.714)</td>
<td>(0.675)</td>
<td>(1.201)</td>
</tr>
<tr>
<td>Observations</td>
<td>11,594</td>
<td>5,213</td>
<td>2,936</td>
<td>732</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>349.7</td>
<td>353.5</td>
<td>357.1</td>
<td>359.4</td>
</tr>
<tr>
<td>Days</td>
<td>411</td>
<td>251</td>
<td>137</td>
<td>34</td>
</tr>
<tr>
<td>Days of week</td>
<td>Weekdays</td>
<td>Wed - Fri</td>
<td>Wed - Fri</td>
<td>Wed - Fri</td>
</tr>
<tr>
<td>Time of day</td>
<td>4 - 7pm</td>
<td>4:30 - 6:15pm</td>
<td>4:30 - 6:15pm</td>
<td>4:30 - 6:15pm</td>
</tr>
<tr>
<td>Require that queue predicted</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Months</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>Jun - Aug</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates and standard errors from four separate regressions, all estimated using two stage least squares (2SLS) with a third-order polynomial in time-of-day as the instrumental variables. The dependent variable in all regressions is freeway capacity, measured in vehicles per five minutes downstream of the bottleneck at Fish Ranch Road. The sample includes five-minute periods between June 2005 and June 2010 during which there is a queue. “Queue predicted” implies that the LASSO model predicted greater than 95% probability of a continuous queue from 4:30 to 6:15 pm. Queue length is measured in thousands of feet. Standard errors are clustered by day.
Table 2: Simulation Model Results

<table>
<thead>
<tr>
<th></th>
<th>Marginal damages w/ hypercongestion</th>
<th>Marginal damages w/o hypercongestion</th>
<th>Percent reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline model</td>
<td>315 mins</td>
<td>149 mins</td>
<td>53%</td>
</tr>
<tr>
<td>w/10% capacity drop</td>
<td>[$56.39]</td>
<td>[$26.67]</td>
<td></td>
</tr>
<tr>
<td>Alternative model</td>
<td>449 mins</td>
<td>149 mins</td>
<td>67%</td>
</tr>
<tr>
<td>w/15% capacity drop</td>
<td>[$80.37]</td>
<td>[$26.67]</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports the results from our stylized simulation model of the Caldecott Tunnel. The first and second columns report the marginal damages in minutes imposed by an additional vehicle at 5 pm. In the first column the capacity of the bottleneck is assumed to be fixed and constant, whereas in the second column the capacity is assumed to drop by up to 10% or 15% due to hypercongestion. Finally, the last column reports the percent reduction in damages without hypercongestion. Brackets contain the dollar value of marginal damages when valuing travel time at half the local median wage (2010 $). See the text for details.
Figure A1: Queue Presence by Time of Day

Queue frequency by time of day

Queue present

12pm 2pm 4pm 6pm 8pm

Queue frequency by time of day
Table A1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>11,594</td>
<td>2008</td>
<td>1.800</td>
<td>2005</td>
<td>2010</td>
</tr>
<tr>
<td>Queue length (1000s of feet)</td>
<td>11,594</td>
<td>4.064</td>
<td>2.991</td>
<td>1.690</td>
<td>15.68</td>
</tr>
<tr>
<td>Capacity downstream of bottleneck (vehicles per 5 mins)</td>
<td>11,594</td>
<td>349.7</td>
<td>26.17</td>
<td>0</td>
<td>408</td>
</tr>
<tr>
<td>Speed upstream of bottleneck (mph)</td>
<td>11,594</td>
<td>11.00</td>
<td>6.643</td>
<td>3</td>
<td>45.90</td>
</tr>
<tr>
<td>Speed downstream of bottleneck (mph)</td>
<td>11,594</td>
<td>41.16</td>
<td>4.693</td>
<td>5.850</td>
<td>54</td>
</tr>
</tbody>
</table>

Notes: The sample includes all five-minute weekday periods between 4 pm and 7 pm and June 2005 and June 2010 during which a queue has formed. We define a queue as present at a detector during a five-minute period if the average speed over that detector falls below 30 mph. Speed upstream of the bottleneck is measured at Gateway Boulevard, averaged across four lanes. Speed downstream of the bottleneck is measured at Fish Ranch Road, averaged across two lanes.
Table A2: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>272,571</td>
<td>2008</td>
<td>1.509</td>
<td>2005</td>
<td>2010</td>
</tr>
<tr>
<td>Queue present</td>
<td>272,571</td>
<td>0.088</td>
<td>0.284</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Queue length (1000s of feet)</td>
<td>272,571</td>
<td>0.294</td>
<td>1.251</td>
<td>0</td>
<td>15.68</td>
</tr>
<tr>
<td>Capacity downstream of bottleneck (vehicles per 5 mins)</td>
<td>272,571</td>
<td>195.2</td>
<td>114.7</td>
<td>0</td>
<td>408</td>
</tr>
<tr>
<td>Speed upstream of bottleneck (mph)</td>
<td>267,205</td>
<td>59.34</td>
<td>16.22</td>
<td>3</td>
<td>76.28</td>
</tr>
<tr>
<td>Speed downstream of bottleneck (mph)</td>
<td>272,571</td>
<td>57.13</td>
<td>8.661</td>
<td>4.200</td>
<td>75.10</td>
</tr>
</tbody>
</table>

Notes: The sample includes all five-minute weekday periods with non-missing data between June 2005 and June 2010. We define a queue as present at a detector during a five-minute period if the average speed over that detector falls below 30 mph. Speed upstream of the bottleneck is measured at Gateway Boulevard, averaged across four lanes. Speed downstream of the bottleneck is measured at Fish Ranch Road, averaged across two lanes.
Table A3: The Effect of Travel Demand on Capacity, Least Squares

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>-0.435</td>
<td>-0.696</td>
<td>-0.203</td>
<td>0.351</td>
</tr>
<tr>
<td></td>
<td>(0.411)</td>
<td>(0.374)</td>
<td>(0.282)</td>
<td>(0.432)</td>
</tr>
<tr>
<td>Queue length (1000s of feet)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>11,594</td>
<td>5,213</td>
<td>2,936</td>
<td>732</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>349.7</td>
<td>353.5</td>
<td>357.1</td>
<td>359.4</td>
</tr>
<tr>
<td>Days</td>
<td>411</td>
<td>251</td>
<td>137</td>
<td>34</td>
</tr>
<tr>
<td>Days of week</td>
<td>Weekdays</td>
<td>Wed - Fri</td>
<td>Wed - Fri</td>
<td>Wed - Fri</td>
</tr>
<tr>
<td>Time of day</td>
<td>4 - 7pm</td>
<td>4:30 - 6:15pm</td>
<td>4:30 - 6:15pm</td>
<td>4:30 - 6:15pm</td>
</tr>
<tr>
<td>Require that queue predicted</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Months</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>Jun - Aug</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates and standard errors from four separate least squares regressions. The dependent variable in all regressions is freeway capacity, measured in vehicles per five minutes downstream of the bottleneck at Fish Ranch Road. The sample includes five-minute periods between June 2005 and June 2010 during which there is a queue. “Queue predicted” implies that the LASSO model predicted greater than 95% probability of a continuous queue from 4:30 to 6:15 pm. Queue length is measured in thousands of feet. Standard errors are clustered by day.
Online Appendix

Table A4: Robustness Checks

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Capacity</th>
<th>(2) Capacity</th>
<th>(3) Capacity</th>
<th>(4) Capacity</th>
<th>(5) Capacity</th>
<th>(6) Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queue length (1000s of feet)</td>
<td>2.744***</td>
<td>-0.763*</td>
<td>2.594***</td>
<td>-0.625</td>
<td>2.580***</td>
<td>-0.720</td>
</tr>
<tr>
<td></td>
<td>(0.736)</td>
<td>(0.387)</td>
<td>(0.698)</td>
<td>(0.366)</td>
<td>(0.710)</td>
<td>(0.373)</td>
</tr>
<tr>
<td>Observations</td>
<td>5,164</td>
<td>5,164</td>
<td>5,261</td>
<td>5,261</td>
<td>5,200</td>
<td>5,200</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>353.6</td>
<td>353.6</td>
<td>353.4</td>
<td>353.4</td>
<td>353.6</td>
<td>353.6</td>
</tr>
<tr>
<td>Days</td>
<td>249</td>
<td>249</td>
<td>252</td>
<td>252</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Model</td>
<td>2SLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>OLS</td>
</tr>
<tr>
<td>Drop zero-flow observations</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Queue speed threshold</td>
<td>20 mph</td>
<td>20 mph</td>
<td>40 mph</td>
<td>40 mph</td>
<td>30 mph</td>
<td>30 mph</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates and standard errors from six separate regressions, estimated using either least squares or two stage least squares (2SLS) with a third-order polynomial in time-of-day as the instrumental variables. The dependent variable in all regressions is freeway capacity, measured in vehicles per five minutes downstream of the bottleneck at Fish Ranch Road. The sample includes five-minute periods between 4:30 pm and 6:15 pm, Wednesday and Friday, and June 2005 and June 2010 during which there is a queue (identical to Columns (2) in Tables 1 and A3). Queue length is measured in thousands of feet. Standard errors are clustered by day.
Table A5: The Effect of Travel Demand on Capacity, Median Regression

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queue length (1000s of feet)</td>
<td>0.691</td>
<td>0.952</td>
<td>-0.162</td>
<td>-0.324</td>
</tr>
<tr>
<td></td>
<td>(0.736)</td>
<td>(0.681)</td>
<td>(0.267)</td>
<td>(0.249)</td>
</tr>
<tr>
<td>Observations</td>
<td>5,213</td>
<td>2,936</td>
<td>5,213</td>
<td>2,936</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>353.5</td>
<td>357.1</td>
<td>353.5</td>
<td>357.1</td>
</tr>
<tr>
<td>Days</td>
<td>251</td>
<td>137</td>
<td>251</td>
<td>137</td>
</tr>
<tr>
<td>Model</td>
<td>Median IV</td>
<td>Median IV</td>
<td>Median Reg</td>
<td>Median Reg</td>
</tr>
<tr>
<td>Require that queue predicted</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates and standard errors from four separate median regressions, estimated using least absolute values (Median Reg) or conditional quantile instrumental variables (Median IV) with a third-order polynomial in time-of-day as the instrumental variables. The dependent variable in all regressions is freeway capacity, measured in vehicles per five minutes downstream of the bottleneck at Fish Ranch Road. The sample includes five-minute periods between 4:30 pm and 6:15 pm, Wednesday and Friday, and June 2005 and June 2010 during which there is a queue (identical to Columns (2) in Tables 1 and A3), with additional restrictions in Columns (2) and (4) (which are analogous to Columns (3) in Tables 1 and A3)). Queue length is measured in thousands of feet. Standard errors are clustered by day.