Heterogeneous Speed, Reliability, and Traffic Externalities

Ian Herzog*

May 2023

Abstract

This paper estimates heterogeneous congestion externalities on a broad sample of roads in England's capital region. A fixed effects approach that compares across times of day finds that 10% more traffic increases travel time by 4%, erodes reliability, and finds smallest marginal effects on high capacity roads. A static model and travel data find that Central London's Congestion Charge exceeds the travel time externality from a representative tolled driver, that public transit's congestion relief benefit exceeds its per-rider subsidy, and vehicle-miles-travelled fees capture over half the benefit from congestion charges that vary across time and space. JEL codes: R41, D62, H23

*Department of Economics, Huron University College, 1349 Western Rd, London, Ontario, Canada, N6G 1H3. E-mail iherzog2@uwo.ca. Thanks to Rohan Vedprakash Khaneja for outstanding research assistance as well as Nathaniel Baum-Snow, Jonathan Hall, Stephan Heblich, Ruben Gaetani, Peter Morrow, Daniel Sturm, Prottoy Akbar, and other participants at the 2021 virtual UEA meeting for comments on precursors to this paper.

1 Introduction

Road congestion is a major concern in cities worldwide and policy makers have proposed solutions ranging from transit investments to congestion charges. For any policy proposal, predicting benefits and informing design requires quantifying congestion costs and identifying how they vary throughout the road network. To shed light on this issue, this paper uses spatially detailed data from London, England to estimate congestion externalities, examine their variation across roads and travelers, and quantify the benefits of policies that internalize these costs.

I estimate congestion costs using data describing regular traffic flow and travel times on a broad cross-section of London's roads. London is an ideal setting for this study since it is a globally important city with data covering an extensive set of roads from Outer London's motorways to the historic city centre. Traffic data come from the region's extensive traffic monitoring network and I use Uber Movement Speeds to measure typical travel times, day-to-day reliability, and how they vary throughout the day on each stretch of road. London has also demonstrated an appetite for road tolls with Central London's Congestion Charge, an early example of a second best tax on downtown drivers, and I shed light on the potential gains from more ambitious policies.

As a first step towards quantifying congestion externalities, I follow Akbar and Duranton (2017), Koch et al. (2021), and Kreindler (2022) and exploit regular travel demand shifts across rush hour and mid day to estimate causal effects of traffic on travel times and their reliability. Econometric estimates suggest that increasing a road's regular rush hour traffic by ten percent increases typical travel time by four percent; within the range of comparable estimates from other settings.[1] I also find that a driver's marginal effect on travel times is smallest on high capacity roads, while a road's location in the city has little impact on its congestion elasticity given observed structural characteristics. In terms of reliability, I find that ten percent more regular rush hour traffic widens the gap between 90^{th} and 50^{th} percentiles of a road's cross-day travel time distribution by 1.9 percentage points. Based on these findings, I argue that travel supply depends on road characteristics and that the impact of traffic on typical travel times can be partially attributed to consistently high travel demand causing particularly slow driving on the worst days.

I incorporate econometric results into a travel supply and demand model and identify significant social costs of time lost to congestion. To quantify externalities from typical trips, I aggregate across roads on driving routes to find that a representative car commuter's marginal external cost is between £0.66 and £0.89 and that public transit creates a congestion relief benefit of between between £0.90 to £1.19 for each peak-hour rider shifted out of a car. Notably, this congestion relief benefit is several times larger than the average subsidy for transit trips in the region.

A particularly salient result is that travel time externalities are too small to justify the £15 congestion charge drivers pay to enter the city centre. For example, the highest external cost

[1] My travel time elasticity estimates are smaller than those implied by speed-flow curves in early engineering studies (Small et al., 2007; Geroliminis and Daganzo, 2008) but in the range implied by recent micro-data studies that directly address endogeneity (Yang et al., 2020; Russo et al., 2021; Koch et al., 2021) and typically imply a relatively inelastic relationship. For example, Koch et al. (2021) and Akbar and Duranton (2017) both yield cost elasticities significantly under 0.5 using different data and methods.

computed is £5.38 for driving downtown from Heathrow Airport and census data show that this falls to £2.15 on average for commuters driving downtown. Since a marginal driver should pay a toll equal to their external cost, justifying Central London's large prevailing congestion charge requires additional external costs of reliability, reduced pollution (Green et al., 2020), traffic accidents (Green et al., 2016), and residential amenity externalities (Tang, 2021).

Deadweight loss estimates suggest that congestion's ongoing social cost is £178.5 million per year in London's untolled areas. I also document substantial variation in optimal tolls across roads that reflects differences in structural characteristics and travel demand. Despite this variation, I find that a relatively simple vehicle-miles-travelled charge—a fixed fee per distance driven regardless of congestion or location—can capture 57% of the benefits of an optimal toll.

This paper contributes new estimates to a literature that uses traffic's effect on travel times to measure congestion costs. Most prior work estimates time costs of traffic where government speed data are available (Anderson, 2014; Yang et al., 2020; Russo et al., 2021). Others use self-reported travel surveys directly (Couture et al., 2018) or as a supplement to Google Maps (Akbar and Duranton, 2017; Akbar et al., 2021) in city-level analysis. Recent work embraces heterogeneity in congestion elasticities and external costs with detailed GPS data from taxis and fleet vehicles (Molnar and Mangrum, 2018; Koch et al., 2021). My data are a unique combination of geographic disaggregation and extensive network coverage that I use to estimate heterogeneous externalities and to compute route-level externalities that speak to the benefits of policies targeting particular trips and travelers.

I also provide new evidence that a road's regular rush-hour traffic affects its day-to-day reliability. An ongoing literature addresses reliability's effect on travel decisions and often finds that drivers will pay more to improve reliability than expected travel times (Brownstone and Small, 2005; Carrion and Levinson, 2012; Brent and Gross, 2018; Bento et al., 2021). However, there has been less progress directly estimating traffic's effect on reliability and connecting reliability to driving's external costs. Prior work in this direction includes highway-corridor case studies that document correlations between traffic flows and upper-tail travel times (Tu et al., 2007) and highlight accidents as a mechanism (Wright et al., 2015). My reliability results speak to the mechanisms through which regular congestion affects drivers; increasing travel demand not only slows a road's average day, it also makes travel times less predictable on any given day and makes the worst days especially slow.

This paper continues by describing traffic data and new measures of speed and reliability from Uber Movement in Section 2. Section 3 then sheds light on travel supply by estimating traffic's effect on travel time, reliability, and examining heterogeneity, and Section 4 uses these estimates to quantify congestion externalities, optimal corrective tolls, and deadweight loss in the city's untolled areas. Section 4 also estimates the potential gains from second best policies and Section 5 concludes with a discussion of implications, caveats, and potential directions for future work.

2 Data and setting

I measure traffic and speed along 1,405 road segments in the London Region, a 1,572 km^2 area including 32 boroughs and the City of London. Since 2003, Central London has been designated as a Congestion Charge Zone (CCZ) and the city currently charges drivers £15 to enter the CCZ on weekdays.[2] Econometric estimation accounts for this using data collected after the policy's introduction and my identification strategy controls for the policy's heterogeneous spatial effects.

My primary data set separates roads into directions of travel, calls directed segments road links, and measures link-level regular traffic, speed, and day-to-day reliability at three times of day. I also observe road type and adjacent traffic signals from OpenStreetMap (OSM), a representative point from which Department for Transport (DfT) enumerates traffic, and sub-region definitions from the Office for National Statistics (ONS).[3] I supplement these data with routing software and origin-destination flows to compute trip-level externalities.

Figure 1 maps the sample's spatial distribution, showing that it provides good coverage of the region's road network and that travel is slower (time costs are higher) in Inner London (the dark grey area) than Outer London. Sample roads are also highly heterogeneous, Table 1 Panel A shows that 10% of observations are in Central London's CCZ and 35% are on motorways or trunk roads. Mean length is 1,040 metres and length's standard deviation is large because of a high frequency of short links; median length is 800 metres and 25% are shorter than 400 metres.

Traffic counts: DfT staff enumerate each link's traffic on a non-winter weekday either annually or on 2, 4, or 8 year cycles. On each visit, enumerators record the number of vehicles crossing in each hour from 07:00 to 19:00 in each direction. My raw data are hourly traffic counts on a panel of road links on Greater London's major roads covering all years from 2003 to 2019. For each link, the primary traffic statistic is a regular flow that captures typical levels in morning, mid day, and evening rush hours. I measure regular flow by averaging each link-hour's counts across years from 2003 to 2019. Econometric analysis then aggregates to hourly flows during each of the morning rush (07:00 to 10:00), mid day (11:00 to 14:00), and evening rush (16:00 to 19:00) at each link. Three-hour windows are chosen so that morning and evening periods capture all passing rush hour traffic even if the road is blocked for part of the period. In the final analysis sample, regular traffic flow draws from 4.1 count days on average and 81% of links are observed at least one time after 2014 (others are observed at least once before 2015). Section 3.6 describes an additional real-time sample that uses raw hourly counts for robustness tests.

Travel times and reliability: My outcome variables use Uber Movement to measure link-level driving speeds during select hours of a representative weekday in October 2019. Uber Movement measures speed using high-frequency Global Positioning System (GPS) pings from the app that

[2]The CCZ is active 07:00-18:00 Monday-Friday and 12:00-18:00 Sat-Sun and bank holidays and my data exclude weekends.

[3]Each road segment is a traffic count point that is regularly visited by DfT enumerators. I average across coordinates when a count-point's representative point moves over time.

tracks and dispatches drivers for Uber's ride-hailing service.[4] This data series matches GPS pings to the OSM road network, drops pickup and drop-off instances, and reports mean speeds on link-hours covered by at least five trips (Uber Movement, 2022).

For each road link provided by Uber, I average across weekdays in October 2019 to compute mean speed during each hour of the day. I also compute deciles and standard deviations of each link-hour's daily speed distribution in October 2019 to produce three reliability measures. Upper-tail unreliability is percent difference between 90^{th} and 50^{th} percentiles of cross-day travel time (Brownstone and Small, 2005) and symmetric unreliability is measured using 80^{th} to 20^{th} percentile differences and the cross-day standard deviation (Carrion and Levinson, 2012).[5] I build an additional sample for real-time analysis that measures speed on a subset of links with traffic counted in 2018 or 2019 and retains hour-day variation that exactly matches DfT count times; this sample is used to validate travel time elasticities, but reliability cannot be measured in real time.

I match Uber links to traffic data using location and travel direction. For each DfT count, I draw a 20 metre buffer around its representative point and identify all Uber links with matching travel directions. I average speed measures in cases where a DfT count matches multiple Uber links. To confirm that underlying speed data are rich enough to reflect daily variation and measure reliability, note that 37% of link-hours are observed on all 23 possible weekdays and half are observed on at least 20 days.[6]

Road characteristics: Before linking with DfT roads, I match Uber segments to all traffic signals within 10 metres and extract speed limits and road types from OSM. A link's speed limit is the average of attached Uber segments, is observed for 93% of link-hours, and I impute missing speed limits using the sample mean on roads of the same class (M, 1-digit A, 2-digit A, or other). For road types, I define a link as a trunk road if OSM classifies at least one of its matched Uber segments as a trunk road or motorway.

Routes and flows: I compute route-level externalities based on time-minimizing paths through the city from the OSM Project's Open Source Routing Machine (OSRM) version 5.22.0 (Luxen and Vetter, 2011). I use the procedure described by Herzog (2022) to define a trip's route as the set of links within 50 metres of the driving path between the centroids of its origin and destination accounting for direction of travel at the point of intersection. I also obtain counts of commuters by mode of transportation between all combinations of 967 middle super output areas (MSOA) tiling Greater London from the 2011 UK Decennial Census Interaction Data.[7] Finally, I obtain trip counts between all station pairs on the London Underground Limited Network from the 2017

[4]Public data releases began in January of 2017 with neighbourhood-pair travel times for select cities. In May of 2019, Uber Movement Speeds launched to publish hourly speeds on road segments five cities including London England.

[5]I compute percent difference reliability measures using the midpoint method.

[6]Appendix Figure A.1 demonstrates an example of this match, highlighting that Uber segments are chosen to measure speed at each road link's representative point.

[7]Commuting data are the number of individuals either (a) driving a car or van or (b) using some other mode or combination of modes for the majority of their trip to/from work as reported in UK Data Service table WU03BUK, which I aggregate from output areas and workplace zones to MSOAs.

Transport for London Rolling Origin and Destination Survey (RODS) which reports rider-flow matrices by time of day.[8]

Descriptive statistics: Table 1 Panel B summarises road link traffic and travel times in each period of the day. Row 1 shows that mean traffic flow falls from a high of 971 vehicles per hour in the evening to 811 vehicles per hour in mid day, when the standard deviation also falls. Mean speed also grows from a 29.6 Km/Hr low in the evening to 33.7 Km/Hr during mid-day and Row 5 confirms substantial deviations in time cost (inverse speed in seconds per metre travelled) from mid day to rush hour at the link level. One point that is omitted from Table 1 but relevant for functional form in econometric analysis is that both traffic and time costs are approximately log-normally distributed; taking logarithms reduces traffic's skewness from 2.7 to 0.2 and time cost's skewness from 2.9 to 0.3.

In terms of reliability, Table 1 Rows 6 and 8 show that both 20^{th} and 80^{th} percentile speeds are fastest in mid-day but that the daily speed distribution's 20^{th} percentile (which is travel time's 80^{th} percentile) varies much more throughout the day. This is reflected in the final three rows which show that travel times are more reliable in mid day than rush hour. This is true of upper-tail unreliability (Row 9) and symmetric unreliability (Rows 10 and 11). Quantitatively, an average road link is most reliable in mid day when there is a 17.4% difference between 90^{th} and 50^{th} percentile travel times and least reliable in the morning when this gap grows to 24 percent. Comparing Panel B's first and last rows gives a sense of speed's substantial cross-day variation, the average standard deviation across day within links is about 30% as large the standard deviation of mean speed across links despite the sample's substantial heterogeneity.

3 Traffic's effect on travel times and reliability

3.1 Empirical strategy

I estimate effects of a road's traffic on per-metre travel time cost, measured as inverse speed, and reliability, measured with the daily speed distribution's dispersion, using fixed effects regressions of the form

$$ln(cost_{rdh}) = \beta ln(traffic_{rdh}) + \alpha_{rh} + \alpha_{rd} + u_{rdh} \tag{1}$$

where $traffic_{rdh}$ measures regular traffic flow along the road link at segment r in direction d at hour h and $cost_{rdh}$ is inverse speed in seconds per metre required to cross that link. Some specifications replace the dependent variable with reliability measures to estimate effects on day-to-day extremes in travel time. Link fixed effects α_{rd} capture fixed road qualities such as width,

[8] 2017 RODS data obtained from `rods.zip` from `http://crowding.data.tfl.gov.uk/` as of April 13, 2021. Rider flows come from the RODS file `Origin-destination matrix by station-zone-time of day 2017.xls` which separately reports station to station London Underground flows in the morning (07:00 to 10:00), mid-day (11:00 to 14:00), and evening (16:00-19:00). The GitHub repository `https://github.com/oobrien/vis/blob/master/tube/data` (also accessed in April 2021) offers a complementary list of station locations.

speed limits, parking, and are central to my identification strategy. Segment by hour fixed effects α_{rh} capture systematic changes across rush-hours of a typical day and the residual u_{rdh} contains other factors that I assume cluster by link but are uncorrelated with traffic.

Equation 1 assumes a constant-elasticity supply function rather than the convex relationship that is conceptually appealing in many contexts (Small et al., 2007). In concept, if traffic is sufficiently low, a road will be in free flow and the marginal vehicle will not affect travel time. Effects on travel occur when traffic grows to the point where the road is congested, and increased interaction between drivers at higher flow can make travel supply convex. I accommodate this using an exponential function $cost_{rdh} = \exp(\beta' traffic_{rdh} + \alpha'_{rh} + \alpha'_{rd} + u'_{rdh})$ which yields the semi-log estimating equation

$$ln(cost_{rdh}) = \beta' traffic_{rdh} + \alpha'_{rh} + \alpha'_{rd} + u'_{rdh}. \quad (2)$$

Equation 2 implicitly subsumes free flow travel times into link fixed effects α'_{rd} and captures traffic's effect with a constant semi-elasticity β'. This parsimony is particularly useful when estimating heterogeneous elasticities and modeling social costs and I assess more flexible specifications as robustness tests.[9]

In both cases, the goal is to estimate the travel time cost elasticity β or semi-elasticity β' and capture the marginal effect of traffic on travel time or reliability. Since primary results use regular speed and traffic (averaged across days) this parameter reflects the effect of typical rush hour surges on driving conditions. Robustness tests apply the same identification strategy to real-time traffic and travel times, but this drops roads from the sample, hampering heterogeneity analysis, and is incompatible with reliability measures. I expect positive time cost elasticities, reflecting falling expected speeds, and adverse effects on reliability that reflect regular rush hour traffic increasing the probability of irregular incidents, such as collisions, that dramatically slow travel.

The main threat to identification is that unobserved road qualities, such as the number of lanes, affect both speed and traffic. Conceptually, these qualities vary across segments and capture shifts in the travel supply curve and movement along travel demand, negatively biasing cost elasticity estimates. Link fixed effects address this concern by forcing identification based on changes across morning, mid day, and evening rush hours. Then, segment by hour fixed effects control for city-wide variation across hours and provide additional robustness by forcing comparisons across two-way roads' directions at each time of day. The identifying assumption is that traffic differences across hours of the day are unrelated to unobserved determinants of a road's speed.

Before continuing, it is worth noting that reverse causality can occur if slow speed constrains the amount of traffic that can flow across a road at a point in time (Small and Chu, 2003). This issue is particularly acute in high-frequency data, where slow speed and low flow coincide whenever a road

[9]While functional form matters for large changes in traffic, such as comparisons between midnight and rush hour, it is important to note that my identification strategy isolates marginal changes in traffic caused by demand shifts across the morning rush hour, evening rush hour, and mid day so that empirical estimates are ultimately local approximations to the true cost function's congested portion.

hits capacity; this is often called hypercongestion and makes instantaneous speed-flow relationships bend backwards. I argue that this issue is less important in my data, which aggregates over entire rush hour periods so that my traffic flow measure inherits properties of what is called density in the context of an instantaneous model of congestion. Appendix A.1 provides additional discussion of concerns about model specification and reverse causality.

3.2 Graphical evidence

Figure 2 visualizes my identification strategy and shows how comparisons across hours within links identify a travel supply function. Panel A plots log time cost against log traffic to find a slight negative correlation, possibly because wider roads accommodate more traffic at higher speeds. This plot's colour gradient shows that fast road links tend to be far from London's central business district, evidence that speed and traffic are also correlated with unmeasured road characteristics. Panel B plots residuals from regressions of log time cost and log traffic on link fixed effects, giving the expected positive relationship.

3.3 Effects on travel time

Table 2 presents ordinary least squares (OLS) estimates of traffic's effect on travel time. Each cell estimates Equation 1 (Panel A) or Equation 2 (Panel B) with link-clustered standard errors in parentheses. Column 1 presents simple OLS estimates, Column 2 adds link fixed effects, Column 3 includes link and hour effects, and Column 4 replaces hour effects with hour by segment effects.

Simple OLS estimates in Column 1 reflect the negative correlation between time costs and traffic visualized in Figure 2, which I argue is caused by high capacity roads accommodating more traffic at higher speeds on average.[10] Column 2 shows that controlling for link fixed effects gives the expected positive sign, suggesting that this specification can identify a weighted average cost elasticity. Controlling for hour of the day in Columns 3 and 4 attenuates effects. This suggests that high-traffic rush hours are (in part) slow because of time varying unobservables such as peak hour increases in pedestrian crossings or lane blockages.

Quantitatively, the elasticity estimate in Column 4 of Table 2 Panel A suggests that adding 10% more vehicles to a road during an average rush hour increases expected travel time by 4%, equivalent to a 4% decrease in speed. Panel B estimates Equation 2's convex relationship and Column 4 finds that adding 100 vehicles per hour increases time cost by 4%, which implies a cost elasticity of 0.35 evaluated at mean traffic.[11] This is smaller than Panel A, but 95% confidence intervals of each model's elasticities overlap, suggesting that the log-log and exponential models are both viable local approximations to the travel supply relationship.[12]

[10] OLS gives a near-zero traffic elasticity ($\beta = -0.01$ and $SE = 0.017$) if fixed effects are replaced with controls for observed road type, traffic signal presence, traffic count year and month dummies, and link length, length squared, and a second order polynomial of longitude and latitude.

[11] The precise semi-elasticity $100 \times (e^{0.394 \frac{100}{1,000}} - 1) = 3.8\%$ rounds to 4%.

[12] In Column 4, 95% confidence intervals give a lower bound of 0.36 for Panel A's log-log elasticity, and evaluating the upper bound of Panel B's semi-elasticity at mean traffic gives an elasticity of 0.41.

3.4 Effects on reliability

Table 3 presents OLS estimates of traffic's effect on travel time reliability, each column is estimated using a variant of Equation 1 with link fixed effects, hour-segment fixed effects, and a different dependent variable. The dependent variable in Column 1 measures upper-tail unreliability as the percent difference between 90^{th} and 50^{th} percentiles of travel time costs across days. Columns 2 and 3 estimate effects on symmetric unreliability; the dependent variable is the percent difference between 80^{th} and 20^{th} percentile time cost in Column 2 and the inverse hyperbolic sine of speed's cross day standard deviation in Column 3. Each dependent variable tracks unreliability with larger values on link-hours with more variation in speed across days.

Results indicate that increasing a road's regular rush-hour traffic flow erodes the day-to-day reliability of its travel times. Specifically, Column 1 associates a 10% increase in regular traffic with a 1.9 percentage point larger difference between a link-hour's 90^{th} and 50^{th} percentile days. Assuming that traffic increases median travel time with an elasticity of 0.405 (as Table 2, Column 4 suggests is true of expected time) this translates to 90^{th} percentile travel times having a traffic elasticity of 0.6.[13] Columns 2 and 3 show that traffic also increases symmetric measures of unreliability and normalizing by the sample standard deviation of each dependent variable shows that the effect of a standard deviation increase in log traffic is relatively large for the 80-20 gap (0.68 of an S.D.) and small for the cross day standard deviation (0.47 of an S.D.).[14]

To confirm that effects on reliability reflect an erosion of the worst travel days, Appendix Table A.1 independently estimates traffic's effects on time cost's 20^{th}, 50^{th}, and 80^{th} cross-day percentiles. Results show that 20^{th} percentile time costs are least responsive with a traffic elasticity of 0.31 and 80^{th} time costs are most responsive with an elasticity of 0.56. Overall, these results support the argument that persistent increases in travel demand make travel less predictable on any given day and, at least in part, increase travel times by making the worst days particularly slow.

3.5 Heterogeneity

I estimate heterogeneous supply functions by interacting traffic with location and road characteristics. Specifically, I modify Equation 2 to be:

$$ln(cost_{rdh}) = \beta_r traffic_{rdh} + \alpha_{rh} + \alpha_{rd} + u_{rdh} \tag{3a}$$

$$\text{s.t. } \beta_r = \beta_0 + \beta_1 signal_r + \beta_2 limit_r + \beta_3 trunk_r + \beta_4 centre_r + \beta_5 outer_r + \beta_6 lat_r + \beta_7 lon_r + \beta_8 lat_r \times lon_r \tag{3b}$$

[13] Suppressing subscripts, the dependent variable is $y = 100 \times \frac{Q90-Q50}{(Q90+Q50)/2}$. Differentiating and simplifying yields $dy = 100 \times \left(\frac{Q50\times d(Q90) - Q90 \times d(Q50)}{((Q90+Q50)/2)^2}\right)$. Isolating for $d(Q90)$ and substituting $dy = 1.9$, an assumption for $d(Q50)$, and $Q50 = 0.113$ Sec/Metre, and $Q90 = 0.135$ Sec/Metre from sample means gives quoted results.

[14] The standard deviation of $asinh(CrossDayStdDev)$ is 0.583 and this variable is zero for 1.9% of observations. Replacing the inverse hyperbolic sine with the natural log and dropping zeros gives similar results.

This model estimates link-specific exponential supply functions with semi-elasticities that depend on whether the link has a traffic signal, is on a trunk road or motorway, its speed limit, whether it is in Central London, in Outer London, its latitude, its longitude, and their interaction.[15] Traffic signals, trunk roads, and speed limits are observable predictors of capacity and location characteristics are included to capture unobserved features that vary across space. I combine equations 3a and 3b and estimate coefficients by OLS with link and hour by segment fixed effects before imputing β_r for each segment.

Appendix Figure A.2a evaluates Equation 3 to compute cost elasticities at observed morning traffic (averaged across directions at each segment) and maps them with thicker and darker lines corresponding to larger increases in time costs from a marginal increase in traffic. I find substantial variation across roads, imputed morning elasticities range from 0.21 at the 10^{th} percentile to 0.80 at the 90^{th} percentile, and the map suggests relatively large elasticities on some radial roads and London's North Circular, but other geographic patterns are subtle. It is also unclear if differences reflect road characteristics, modeled by Equation 3b, or convexity interacting with observed traffic flows in Equation 3a.

Table 4 evaluates Equation 3 to estimate supply functions for four road scenarios (chosen to reflect progressively decreasing road capacity) and computes traffic elasticities by evaluating at traffic's 25^{th}, 50^{th}, and 75^{th} percentiles. Column 1 simulates a highway as a North Circular trunk road with a speed limit of 50 miles per hour, Column 2 stays at that location but moves to a non-trunk road with a 30 mile per hour limit, Column 3 adds a traffic signal, and Column 4 moves to Central London.

Comparing across Table 4's columns suggests that, all else equal, a vehicle's marginal effect on travel times is smallest on high capacity roads. Focusing on the first row's low-traffic case suggests a cost elasticity of 0.129 on a highway, a slightly larger elasticity of 0.256 on a nearby city street, and an even larger elasticity of 0.291 if that street has a traffic signal. Column 4 shows that moving to Central London hardly changes cost elasticities relative to Column 3's suburban counterpart and the next two rows show similar patterns at higher traffic flow. Overall, an analysis of variance finds that signals, trunk roads, and speed limits jointly explain 92.8% of variation in β_r.[16] Together, these results suggest that these three observed road characteristics leave little variation in semi elasticities for location to capture.[17]

Table 4's fourth row shows that estimated supply functions are more convex on city streets (Columns 2, 3, and 4) than on highways (Column 1). This could reflect the fact that rows reflect city-wide traffic percentiles rather than a link-specific traffic distribution. Indeed, traffic flow on

[15]Central London is defined by the original boundaries of London's CCZ and the rest of Inner London is the sub-region dummy's base case.

[16]Extending Equation 3b to include a third-order polynomial in lat_r and lon_r still leaves 57.7% of variation generated by this 13-parameter model captured by direct effects of signals, trunk roads, and speed limits. Adding all interactions of these three characteristics increases their share of explained variation to 66%.

[17]By definition, road characteristics and location jointly explain all variation in estimated β_r and this analysis does leave the possibility of unexplained heterogeneity in true supply elasticities ending up in u_{rdh}. Unfortunately, I do not have sufficient within-segment variation to non-parametrically identify segment-specific β_r and so leave a more thorough analysis of unexplained heterogeneity for future work.

a high capacity highway modeled in Column 1 might frequently exceed the 75^{th} percentile used in Table 4, suggesting that an even larger increase is required to capture a similar proximity to that road's capacity. The data provide some support for this hypothesis; evaluating the highway scenario (Column 1) at the sample's 95^{th} percentile traffic flow yields a cost elasticity of 0.63.

Appendix Table A.2 shows that results are qualitatively similar if Equation 3a is a more flexible quadratic function of log traffic. However, different patterns emerge in a non-convex alternative using log traffic without quadratic terms. Specifically, the non-convex alternative suggests that the highway scenario has the largest cost elasticity and that this elasticity depends on unobserved geographic factors, falling substantially between scenarios in Columns 3 and 4. This disagreement could result from mis-specification of the simple model, and I prefer Equation 3a's convex specification because it captures the same patterns as the more flexible log-quadratic model with a parsimonious functional form.[18]

3.6 Sensitivity analysis

Alternative regression specifications: Appendix Table A.3 shows that estimates of traffic's effect on travel times are robust to alternative regression specifications. Column 1 controls for traffic quartile dummies to force identification based on small changes in traffic. In concept, this addition reduces the influence of links with large changes in traffic throughout the day. In practice, this hardly affects the constant elasticity estimate but reduces the exponential model's semi-elasticity from 0.394 to 0.343. To address time-varying omitted variables, Column 2 adds controls for the observables specified in Equation 3b interacted with hour dummies and finds no substantial changes.

Column 3 weights link-hours by the number of times traffic is counted by DfT enumerators in the raw data and finds that this has opposite effects on the constant elasticity and exponential models. The weighted least squares constant elasticity estimate is somewhat larger than the baseline result (0.511 rather than 0.405) but the exponential model's weighted semi-elasticity estimate is slightly smaller than baseline (0.355 rather than 0.394). Column 4 tests outlier influence by dropping observations in the top 1% of the traffic distribution and finds little effect on the log-log elasticity estimate. However, dropping outliers increases the exponential model's semi-elasticity to 0.45, equivalent to a time cost elasticity of 0.41 at mean traffic. This highlights that the convex estimating equation is more sensitive to outliers than its constant elasticity counterpart, which can contribute to differences in their results.

In terms of travel's reliability, Appendix Table A.4 shows that estimates of traffic's effect remain positive and significant (albeit smaller) when controlling for log expected travel time per metre.[19] This result reassures that effects on reliability are not mechanical reflections of travel time effects.

Speed-traffic match quality: Approximately half of the sample is at DfT traffic count locations

[18] Traffic flow on trunk roads is 119% higher than other sampled roads on average. It is therefore plausible that the true supply function is convex and high traffic makes a log-linear approximation yield supply functions that are a locally steepest on highways.

[19] Attenuated effects on reliability given expected costs are unsurprising since these regressions condition on an endogenous control (Angrist and Pischke, 2008).

that average speed across more than one Uber link. This is expected (Uber links can be shorter than DfT links and can abut near DfT's representative point) but it can generate measurement error if intersecting roads are erroneously matched to the same traffic count. Appendix Table A.5 Column 1 shows that dropping these one-to-many matches only slightly increases traffic's effect in both log-log and exponential models. Column 2 then extracts road name descriptions from both DfT and Uber links to identify cases where a DfT count on a small street (3-digit A or smaller) is matched to an Uber link that OSM classifies as a trunk road or motorway; dropping these cases only slightly attenuates the log-log elasticity and does not affect the exponential model's semi-elasticity.

Real time estimates: Appendix Table A.6 presents effects on travel time estimated using real time traffic and speed data. These estimates restrict the sample to a subset of links with traffic observed in 2018/19 and matches traffic counts with speeds by calendar day and hour. This sample includes twelve hourly observations of traffic on each link-day, but is an unbalanced panel because Uber suppresses speeds for link-hours with less than five trips. All specifications include link fixed effects so that elasticities are identified by variation across hours of the day and some specifications average traffic and cost within each rush hour to focus on variation across (rather than within) morning, mid day, and evening as in the main specification.

Overall, real-time estimates resemble primary results. If anything, results that exploit all available hourly variation give slightly smaller effects than baseline results. In contrast, real-time estimates that focus on variation across rush hours give slightly larger elasticities than baseline results, but semi-elasticities estimated with a real-time exponential model remain attenuated.[20]

Flexible supply functions: Appendix Table A.7 estimates convex effects of traffic on travel times using flexible variants of Equation 1. Column 1 interacts log traffic with four quartile dummies to estimate a simple four part spline. Column 2 then adds controls for dummies for each quartile of log traffic. This has the advantage of forcing identification based on small changes in traffic, but imposes a more modest interpretation by estimating local elasticities within each traffic quartile rather than a smooth supply function. Columns 3 and 4 take an alternative approach by adding quadratic and cubic log traffic to the estimating equation. Column 5 adds a quadratic traffic term to Equation 2's exponential function and Column 6 tests for increasing marginal effects directly by dropping logarithms and regressing time cost on traffic and its square. All regressions include link and segment-hour fixed effects and Columns 3 though 6 demean traffic variables so that main effects reflect average partial effects.

Column 1 of Appendix Table A.7 shows that without including quantile dummies, a four-part spline finds no evidence of a convex travel supply function. However, this could be because comparisons across quartiles partially reflect large changes in daily traffic. Column 2 shows that conditional on quartile dummies, estimated traffic elasticities grow monotonically in traffic; going from 0.17 in the bottom quarter of traffic flow to 0.75 in the top quarter. Column 3 goes on to show that specifying a quadratic function of log traffic also finds an increasing elasticity and Column 4

[20]Trunk roads' share of the real-time sample (33%) is similar to the main sample, suggesting heterogeneous treatment effects in this dimension do not explain differences between regular-traffic and real-time supply estimates.

finds no effect of adding a cubic term.

While Appendix Table A.7 shows evidence of increasing cost elasticities, this does not imply a convex cost function. Indeed, Column 5 removes logarithms from traffic to find a negative coefficient on squared traffic, indicating a diminishing semi-elasticity. Column 6 also removes logarithms from time costs and again finds a negative coefficient on squared traffic, indicating a diminishing marginal effect. All told, I find no conclusive evidence in favour of any given functional form. With this in mind, the next section's social cost calculations examine implications of alternative functional form assumptions but favour the exponential model for its intuitive appeal.

4 The social cost of congestion

I estimate congestion externalities and optimal tolls in a standard Pigouvian model (e.g. Small et al. (2007)) where traffic's total time cost to all drivers on road link rd at hour h is

$$TC_{rdh} = w \cdot len_r \cdot traffic_{rdh} \cdot cost_{rdh}$$

which is the product of the opportunity cost of time w, number of vehicles travelling $traffic_{rdh}$, road length len_r, and expected time per metre $cost_{rdh}$ which increases in traffic flow as drivers create congestion externalities. I examine implications of travel supply's functional form by alternating between constant elasticity (log-log), exponential, and heterogeneous cost functions. In each case, congestion externalities increase in time costs and I measure congestion's social cost as its deadweight loss.[21]

4.1 Marginal external costs

The marginal social cost of adding a vehicle to road link rd at hour h is $MSC_{rdh} = [1 + \beta_{rdh}] \cdot w \cdot len_r \cdot cost_{rdh}$ where β_{rdh} is the link-specific elasticity of time cost with respect to traffic, which equals $\beta'_r traffic_{rdh}$ in the exponential case.[22] Drivers internalize average cost $AC_{rdh} = w \cdot len_r \cdot cost_{rdh}$ and link-hour marginal external cost is

$$MEC_{rdh} = MSC_{rdh} - AC_{rdh} = \beta_{rdh} \cdot w \cdot len_r \cdot cost_{rdh}. \quad (4)$$

I quantify congestion externalities with w equal to half of the London Region's median wage (Small et al., 2007) and alternate between a constant elasticity $\beta = 0.41$, exponential scale $\beta' = 0.39/1,000$,

[21] The model implicitly assumes that reducing travel time reliability has no direct costs other than as a mechanism that increases typical time costs. This simplification makes the model consistent with prior work and and facilitates adopting a standard calibration for w.

[22] Differentiating $w \cdot len_r \cdot cost_{rdh} \cdot traffic_{rdh}$ with respect to $traffic_{rdh}$ gives $w \cdot len_r \cdot \left(cost_{rdh} + traffic_{rdh} \cdot cost'(traffic_{rdh})\right)$ which simplifies to $w \cdot len_r \cdot cost_{rdh} \cdot [1 + \beta]$ when the cost function has a constant traffic-elasticity β. Even with a constant cost elasticity, MSC_{rdh} implicitly depends on traffic flow because $\beta \cdot cost_{rdh} = traffic_{rdh} \cdot cost'(traffic_{rdh})$.

and heterogeneous effects that depend on location and road characteristics as in Section 3.5.[23] Looking forward, Section 4.2 defines optimal tolls by evaluating Equation 4 at traffic and time costs that equate marginal social benefits and costs of travel.

Figure 3 depicts equilibrium and optimal outcomes with exponential supply. The marginal external cost at any point is the vertical difference between MSC and AC and labelled MEC at the observed equilibrium. The optimal toll τ^* equals optimal traffic's marginal external cost and depends on travel demand, which Figure 3 depicts as the downward sloping MB curve. Figure 3 also demonstrates that, even if demand is unknown, observed MEC is an upper bound on τ^* whenever observed traffic (and time cost) is higher than optimal.

I use a simple heuristic to assess Central London's Congestion Charge, which is a second-best cordon fee policy that currently charges £15 to drivers entering the city centre. This heuristic assumes that there is a representative trip into the tolled CCZ, defined so that the marginal driver taking this trip would create an MEC equal to the congestion charge if this charge were set optimally. If set too low, the charge would shift private travel costs below Figure 3's blue curve, which would then intersect MB to the right of $Traffic^*$ and yield an equilibrium MEC above the toll. Conversely, if the charge is too high, private cost would shift up too much, traffic would fall to the left of $Traffic^*$, and MEC would be lower than the toll. With this in mind, I compute trip-level MEC to test different definitions of a representative tolled trip and argue that Central London's Congestion Charge is too high (or low) if tolled drivers' MEC is above (or below) the existing congestion charge.[24]

I estimate external costs of trips through the city where a driver's external cost is the sum of the externalities imposed on each road used. To this end, define a route R_{ij} as the set of road links traversed between origin i and destination j. Aggregating Equation 4 across links on a route gives the marginal external cost of driving from i to j in hour h as

$$MEC_{ijh} = w \sum_{rd} 1\{(c,d) \in R_{ij}\} \cdot \beta_r \cdot len_r \cdot cost_{rdh} \tag{5}$$

where the $1\{\cdot\}$ is the indicator function and R_{ij} is defined using the Open Source Routing Machine. For example, I compute the external cost of driving from Heathrow Airport to the city centre as a rough upper bound on the optimal toll for a long and salient trip that would pay Central London's congestion charge.[25]

I also compute the average car commuter's marginal external cost by defining routes between the

[23]The value of time is half of the £18.3 full-time hourly wage reported by Office for National Statistics (2020) and cost functions are based on Table 2.

[24]While useful, this heuristic adapts Figure 3 to describe a route rather than an isolated road, ignores overlapping trips, and contradicts the standard notion of link level travel demand that I use in Section 4.2. Overall, defining an optimal cordon fee requires a more complete model of travel demand and is outside of this paper's scope.

[25]The Heathrow to city centre trip begins at the Heathrow Airport MSOA's centroid (ONS code E02000524) and ends at the City of London's centroid (E02000001). OSRM selects a route via M4 to Great West Road (A4).

centroids of all MSOA pairs in Greater London and using origin-destination census data compute

$$\overline{MEC}_h^{work} = \sum_{ij} \frac{n_{ij}^c}{\sum_{rs} n_{rs}^c} MEC_{ijh} \tag{6}$$

where weights n_{ijh}^c are the number of commuters who live in i, work in j, and drive for the majority of their commute, and MEC_{ijh} combines routes, time costs, and parameters using Equation 5.[26] I also evaluate Equation 6 using only commutes into London's CCZ to create a benchmark for evaluating Central London's Congestion Charge.

Finally, I assess public transit's congestion relief benefit by re-weighting Equation 6 to reflect the hypothetical external cost of forcing the average non-car commuter to drive.[27] This census-based measure has the advantage of including multi-mode trips and reflecting the full door to door trips of actual commuters, but includes walking and cycling and abstracts from imperfect substitution between modes of transportation. I also use driving routes between all Underground station pairs and station-station flow data to simulate moving the average London Underground trip into a car.[28] Despite this measure's limitations—weights are based on a sample which does not extend onto London Overground, Light Rail, or Bus networks—it has the advantage of reflecting all trip purposes while highlighting the London Underground's role in alleviating congestion.

Quantitative results

Table 5 presents marginal external costs of simulated trips averaged across morning and evening rush hours. Row 1 assumes a constant elasticity travel supply function, Row 2 assumes exponential supply with a constant semi-elasticity, Row 3 returns to the log-log supply function and estimates heterogeneous elasticities across roads using characteristics in Equation 3b, and Row 4 uses a road-varying exponential supply function (as in Equation 3).

Column 1 shows that the average car commute's external cost ranges from £0.66 to £0.89 for a one-way trip depending on the supply function. Column 2 shows that the marginal driver from Heathrow Airport to Central London creates between £2.32 and £5.38 in externalities. The Heathrow to Centre trip's large external cost is expected because it is much longer than the average car-based commute and traverses Inner London where roads are particularly congested.[29]

More notably, the Heathrow to Centre trip is subject to London's Congestion Charge and

[26] I compute evening commuters' externalities by switching origins and destinations to compute return routes $R_{ij,eve} = R_{ji}$ so that $\overline{MEC}_{eve}^{work} = \sum_{ij} \frac{n_{ij}^c}{\sum_{rs} n_{sr}^c} MEC_{ji,eve}$.

[27] The census-based measure of transit's congestion relief benefit replaces n_{ijh}^c with the number of commuters who live in i, work in j, and use any mode (or combination of modes) other than driving for the majority of their commute and continues using OSRM driving routes to reflect the external costs of each commute's driving option.

[28] The Underground-based congestion relief benefit computes routes between station pairs rather than MSOAs and weights them with Underground rider counts in each hour from RODS data containing 268 London Underground Limited stations (Bank and Monument are combined and Paddington counts as a single station). RODS contains 16 stations that are outside of the London region but still act as origins and destinations.

[29] The Heathrow to Central London route crosses 18 km of in-sample roads (22 links) while the mean car commute crosses 5.5 km of in-sample roads (4.7 links on average).

external cost estimates are all substantially smaller than the £15 toll this driver would pay to enter the CCZ in practice.[30] Column 3 shows that the same is true of tolled commuters; on average, accounting for all roads on their route, a commuter who drives into the CCZ for work creates an external cost of between £1.44 and £2.15 per direction. Overall, these results suggests that if travel time externalities are the only target, London's Congestion Charge is set above its optimal level. However, this does not account for additional benefits of reliability, reduced pollution (Green et al., 2020), traffic accidents (Green et al., 2016), and residential amenity externalities (Tang, 2021), which all justify larger tolls.

Column 4 of Table 5 suggests that forcing the average non-car commuter to drive would create an external cost of between £0.90 and £1.19 per trip. These magnitudes exceed those in Column 1 because transit commutes are longer than car commutes and more often head to central locations using particularly busy roads.[31] This suggests that public transit creates substantial congestion relief benefits by targeting areas where marginal costs are high because of large numbers of affected drivers (in all models) and convex supply (in the exponential case). Comparing columns 4 and 5 shows that the driving alternative to the average London Underground trip would create an even larger external cost than the average non-car commute's driving alternative. For context, congestion relief benefit estimates all substantially exceed the bus and underground network's 2018/19 subsidy per passenger of £0.18, suggesting that congestion externalities alone justify increasing subsidies for London's public transit riders.[32]

Comparing across rows in Table 5 shows that exponential supply functions imply larger externalities than their log-log counterparts and that allowing cross-road heterogeneity also increases external costs. And while differences across supply models are largest for the Heathrow to Centre trip, these patterns hold across simulations. I argue that external costs are large under heterogeneous supply because least squares estimates of homogeneous supply functions overweight high capacity roads where traffic varies most throughout the day and where cost functions are relatively flat.[33]

4.2 Optimal tolls, deadweight loss, and gains from second policies

Measuring optimal tolls and deadweight loss (DWL) requires imposing additional structure on travel supply and demand functions. Specifically, I assume inverse travel supply functions $AC_{rdh} =$

[30] Assuming a two-way trip doubles MEC, but external costs remain below Central London's Congestion Charge. However, a driver who chains trips within the CCZ could conceivably be undercharged for their effect on travel times.

[31] The average non-car commute would alternatively cross 6.4 km of monitored roads (7.5 links on average) on their driving option while actual drivers cross 4.73 km of monitored roads (and 5.5 links) on average.

[32] In the 2018/19 fiscal year, gross expenditure net of passenger income per passenger-journey was £0.18 according to calculations based on Transport for London (2019). It is also worth noting that, in isolation, the London Underground actually turns a profit and effectively cross-subsidizes buses and other TfL operations.

[33] Section 3.5 documents trunk roads' flat supply functions and decomposing the variance of residuals from a regression of traffic on link fixed effects reveals that 77% of variance in residual traffic occurs on trunk roads. Interestingly, both of these facts are flipped in the log-log model: trunk roads account for 34% of variance in residual log traffic and Appendix Table A.2 shows that the heterogeneous log-log model estimates large elasticities on trunk roads.

$w \cdot len_r \cdot \alpha_{rdh} \cdot traffic_r^{\beta}$ in the log-log case and $AC_{rdh} = w \cdot len_r e^{(\alpha'_{rdh} + \beta'_r traffic)}$ in the exponential case. Both include link-level supply shifters and assume constant elasticity demand so that marginal willingness to pay to drive across link rd in hour h is $MB_{rdh} = w \cdot len_r \cdot \gamma_{rdh} \cdot traffic^{\eta}$. I test a range of inverse demand elasticities, but note existing evidence that travel demand is approximately perfectly elastic in the long run (Duranton and Turner, 2011; Garcia-Lopez et al., 2020; Kim, 2022) and so $\eta = 0$ best reflects a long-run view, which matches my static model and measure of regularly experienced rush hour traffic.

For any parametrization, each link-hour's optimal traffic sets $MB_{rdh} = MSC_{rdh}$ while observed traffic is higher and sets $MB_{rdh} = AC_{rdh}$. Optimal road user fees are determined by solving for optimal traffic, corresponding optimal costs, and substituting into Equation 4 for tolls equal to MEC^*_{rdh} per kilometre on each link in each hour. Figure 3 depicts optimal tolls as τ^* and shades corresponding welfare gains, which equal unregulated DWL. I also evaluate second best policies that set tolls using simplified supply models and use heterogeneous exponential functions to compute the change in DWL achieved by each policy.[34]

Quantitative results

Optimal tolls: Table 6 presents means and interquartile ranges of optimal tolls in GBP per kilometre. All estimates exclude the CCZ and weight links by length to reflect an average stretch of road in the city's untolled area.[35] Panel A averages across morning and evening rush hours and Panel B reports percent change from mid-day to morning rush hour peak. Panel B only reports a single range for the log-log model since this supply function always equates ratios of a link's hourly optimal tolls with ratios of time costs.[36]

Panel A of Table 6 shows that mean optimal tolls range from £0.05 to £0.12 per kilometre depending on the shapes of supply and demand functions. The exponential supply model tends to suggest smaller optimal tolls than the log-log model because log-log cost functions tend to call for larger traffic reductions in practice. Given modeling assumptions, there is substantial variation across links and tolls' interquartile range maxes out at a £0.10 difference (83% of the mean) between 75^{th} and 25^{th} percentiles with a heterogeneous log-log supply function and the steepest marginal benefit curve.

Comparing across columns shows that changing the assumed demand elasticity is particularly important for the exponential model, where moving from $\eta = -2$ to $\eta = 0$ decreases the mean optimal toll by 44% in Row 2 and 45% in Row 4. This occurs because optimal traffic is always smallest when demand is most elastic (η near zero) and because MEC increases steeply in traffic with exponential externalities. The upshot is that if travel demand is most elastic in the long run,

[34] Appendix A.2 solves for optimal tolls, DWL, and describes numerical implementation.

[35] I focus on areas outside of the CCZ because it is unclear how to apportion the area's entry-based congestion charge to individual road links when bringing the model to data to determine optimal traffic and tolls.

[36] When $cost_{rdh} = \alpha_{rdh} \cdot traffic_{rdh}^{\beta_r}$, optimal tolls satisfy $\frac{\tau_{rdh}}{\tau_{rdh'}} = \frac{cost^*_{rdh}}{cost^*_{rdh'}} = \frac{cost_{rdh}}{cost_{rdh'}}$ where $cost^*_{rdh}$ is optimal, the first equality follows from Equation 4, and the second follows from $cost^*_{rdh} = \left(\frac{1}{1+\beta_r}\right)^{\frac{\beta_r}{\beta_r - \eta}} cost_{rdh}$ as shown in Appendix A.2.

the exponential model suggests that a stable long run congestion charge can be set much lower than an uncommitted short run toll.[37] Comparing rows 3 and 4 shows that exponential supply is an important piece of this argument; assuming log-log supply gives a toll cut of just 25% when moving from very inelastic to perfectly elastic demand.

Panel B of Table 6 shows that there is also important variation in optimal tolls throughout the day. This is particularly true for the exponential model which suggests that an average road's toll should grow by somewhere between 24.4% to 27.0% depending on supply and demand functions used. Daily variation also has large interquartile ranges and negative 25^{th} percentiles reflect roads where traffic peaks outside of rush hour, motivating higher tolls in mid-day than morning. Finally, the log-log model recommends far less intra-day variation in tolls than the exponential model. Combined with Panel A, this shows that tolls from the log-log model are generally higher in rush-hour but less dynamic than the exponential model.

Deadweight loss and second best tolls: Table 7 presents annual DWL estimates assuming heterogeneous exponential supply, excluding the CCZ, aggregating over entire morning, mid-day, and evening three-hour periods, and extrapolating to 260 workdays per year. Row 1 is untolled DWL, which measures the efficiency gain from optimal tolls that vary across time and space according to congestion and road characteristics. Remaining rows simulate DWL under second best policies and percentages are improvements over Row 1's status quo. Row 2 assumes a fixed toll per kilometre travelled, Row 3 assumes tolls only vary across rush hours, and Row 4 applies congestion-based toll schedules that remain blind to road structure.[38]

Row 1 of Table 7 indicates that DWL estimates range from £178.5 million per year with a constant marginal benefit (perfectly elastic demand) to £67.7 million with the steepest marginal benefit curve. For context, the largest DWL estimate exceeds Central London's £146.7 million in net revenue from congestion charging in the 2018/19 fiscal year.[39] Appendix Table A.8 computes untolled DWL using alternative supply functions to find that magnitudes are comparable across models, that assuming homogeneous roads decreases DWL estimates, and that assuming constant elasticity supply increases DWL with perfectly elastic demand but decreases DWL otherwise. Overall, these results suggest rush hour road congestion creates substantial social costs, particularly in the long run when travel demand tends towards perfectly elastic.

Table 7 also shows that simple second best congestion charging policies yield substantial welfare gains. Focusing on Column 1, with perfectly elastic demand, Row 2 shows that a constant toll per kilometre travelled reduces DWL from £178.5 million to £77.2 million. In other words, this simple vehicle-miles-travelled fee captures 57% of the gains from an optimal toll despite ignoring variation

[37] Kim (2022) studies road improvements and finds that freeway demand is nearly perfectly elastic within a span of about two years. However, in my context the short/long run distinction should be interpreted cautiously given that I do not model dynamics.

[38] Tolls are set by evaluating the homogeneous exponential supply model at observed traffic and time cost on each link hour outside the CCZ (Row 4), averaging by demand elasticity (Row 2), and averaging by demand elasticity and hour (Row 3). The data exclude nights and weekends, so all tolls and DWL are implicitly zero at these times.

[39] Transport for London (2019) reports £229.9 million in gross congestion charging revenues, £66.5 million in facilities and traffic management costs, and £16.7 million in other administrative costs and depreciation.

in roads' structural characteristics and travel demand. Row 3 shows only modest gains from moving to tolls that vary across hours of the day. Finally, Row 4 shows that setting tolls that ignore road structure but dynamically adjust to reflect congestion captures 80% of the long-run gains from optimal congestion charging. And comparing across columns shows that when travel demand is less elastic, such as in the short run, second best tolls tend to perform even better.

5 Conclusion

This paper estimates the time cost of road traffic in England's capital region. To this end, I estimate traffic's effects on travel time and reliability on a broad cross-section of roads, document substantial heterogeneity in travel supply functions, and compute traffic externalities, optimal tolls, and deadweight loss in a standard model of congestion externalities.

I find that a simple distance based congestion charge can have substantial benefits, capturing over half of potential welfare gains even if tolls fail to vary across time and space to reflect road capacity, where people want to go, and when they want to get there. This is particularly encouraging since this policy would simply be an urban vehicle-miles-travelled charge, which is more familiar to drivers and technologically simpler than full-on dynamic congestion charging (Clements et al., 2021). With that said, future work could do more to examine the political economy and privacy concerns tied to high- or low-tech policies that track and toll individual travel. Future work could also assess the external validity of my findings since second best pricing might yield smaller benefits in cities with different travel needs or more car-centric road systems.

In terms of existing policies, I find that travel time externalities justify increasing transit subsidies, but are too small to justify the current rate of Central London's Congestion Charge. This could be taken as a criticism of this area-based fee, perhaps in favour of a more flexible and nuanced congestion charge, but this would inappropriately compare a real second-best policy with an idealized first-best alternative. Instead, I stress that an appropriate road user fee might exceed my external cost estimates if it also reflects effects on travel's reliability, traffic accidents, noise, and pollution. So, policy makers considering the costs and benefits of congestion charging should account for the multiple correlated externalities it can address.

References

Akbar, P. and Duranton, G. (2017). Measuring the cost of congestion in highly congested city: Bogotá. Technical report.

Akbar, P. A., Couture, V., Duranton, G., and Storeygard, A. (2021). Mobility and congestion in urban india. Technical report, National Bureau of Economic Research.

Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104(9):2763–96.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Bento, A., Roth, K., and Waxman, A. R. (2021). The value of urgency. Technical report.

Brent, D. A. and Gross, A. (2018). Dynamic road pricing and the value of time and reliability. *Journal of Regional Science*, 58(2):330–349.

Brownstone, D. and Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4):279–293.

Carrion, C. and Levinson, D. (2012). Value of travel time reliability: A review of current evidence. *Transportation research part A: policy and practice*, 46(4):720–741.

Clements, L. M., Kockelman, K. M., and Alexander, W. (2021). Technologies for congestion pricing. *Research in Transportation Economics*, 90:100863.

Couture, V., Duranton, G., and Turner, M. A. (2018). Speed. *Review of Economics and Statistics*, 100(4):725–739.

Duranton, G. and Turner, M. A. (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, 101(6):2616–52.

Garcia-Lopez, M.-A., Pasidis, I., and Viladecans-Marsal, E. (2020). Congestion in highways when tolls and railroads matter: Evidence from european cities. Technical report.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770.

Green, C. P., Heywood, J. S., and Navarro, M. (2016). Traffic accidents and the London congestion charge. *Journal of Public Economics*, 133:11–22.

Green, C. P., Heywood, J. S., and Paniagua, M. N. (2020). Did the london congestion charge reduce pollution? *Regional Science and Urban Economics*, 84:103573.

Herzog, I. (2022). The city-wide effects of tolling downtown drivers: Evidence from london's congestion charge. Technical report.

Kim, J. (2022). Does roadwork improve road speed? evidence from urban freeways in california. *Regional Science and Urban Economics*, 93:103773.

Koch, N., Ritter, N., Rohlf, A., and Thies, B. (2021). Machine learning from big gps data about the heterogeneous costs of congestion. Technical report.

Kreindler, G. E. (2022). The welfare effect of road congestion pricing: Experimental evidence and equilibrium implications. Technical report.

Luxen, D. and Vetter, C. (2011). Real-time routing with OpenStreetMap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.*

Molnar, A. and Mangrum, D. (2018). The marginal congestion of a taxi in new york city. Technical report.

Office for National Statistics (2020). Annual survey of hours and earnings: Earnings by place of residence, borough. Technical report.

Russo, A., Adler, M. W., Liberini, F., and van Ommeren, J. N. (2021). Welfare losses of road congestion: Evidence from rome. *Regional Science and Urban Economics*, 89:103692.

Small, K. A. and Chu, X. (2003). Hypercongestion. *Journal of Transport Economics and Policy (JTEP)*, 37(3):319–352.

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

Tang, C. K. (2021). The cost of traffic: evidence from the london congestion charge. Technical report.

Transport for London (2019). Annual report and statement of accounts 2018/19. Technical report.

Tu, H., Van Lint, J., and Van Zuylen, H. J. (2007). Impact of traffic flow on travel time variability of freeway corridors. *Transportation Research Record*, 1993(1):59–66.

Uber Movement (2022). Historical speeds, hourly time series. Technical report. (accessed June 2022).

Wright, B., Zou, Y., and Wang, Y. (2015). Impact of traffic incidents on reliability of freeway travel times. *Transportation Research Record*, 2484(1):90–98.

Yang, J., Purevjav, A.-O., and Li, S. (2020). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in Beijing. *American Economic Journal: Economic Policy*, 12(1):418–53.

6 Tables

Table 1: Summary statistics

Panel A: Link characteristics					
Central	Trunk	Length (m)	Counts	Counted since 2015	Uber matches
0.10 (0.30)	0.35 (0.48)	1,039.81 (900.60)	4.07 (2.78)	0.81 (0.39)	1.75 (0.94)

Panel B: Traffic flows, speed, and reliability				
	Full sample	Morning	Mid day	Evening
(1) Traffic	902 (709)	924 (734)	811 (607)	971 (766)
(2) log(Traffic)	6.59 (0.62)	6.61 (0.64)	6.50 (0.60)	6.67 (0.62)
(3) Speed (Km/Hr)	31.48 (14.04)	31.22 (13.64)	33.67 (14.58)	29.56 (13.56)
(4) log(Cost)	-2.07 (0.44)	-4.63 (0.42)	-4.71 (0.44)	-4.57 (0.45)
(5) log($Cost/Cost_{mid}$)	0.07 (0.20)	0.07 (0.22)	0.00 (0.00)	0.14 (0.26)
(6) Cross-day Q20	28.45 (14.58)	27.78 (14.23)	31.18 (15.12)	26.41 (13.95)
(7) Cross-day Q50	31.73 (14.39)	31.44 (14.11)	33.89 (14.78)	29.86 (14.01)
(8) Cross-day Q80	34.69 (14.12)	34.87 (13.86)	36.34 (14.38)	32.87 (13.90)
(9) Q90-Q50 % Diff	21.79 (20.33)	24.01 (21.43)	17.38 (17.29)	23.99 (21.30)
(10) Q80-Q20 % Diff	24.48 (21.23)	27.30 (22.58)	19.64 (18.21)	26.48 (21.80)
(11) Cross-day Std.Dev.	4.32 (2.76)	4.80 (3.04)	3.70 (2.21)	4.46 (2.85)

Cells present means and standard deviations are in parenthesises. Rows 6 to 8 describe percentiles of cross-day speed distributions and Rows 9 and 10 are midpoint method percent differences in time cost percentiles, and Row 11 is speed's cross day standard deviation.

Table 2: Traffic's effect on travel times

	log(Seconds per metre)			
	(1)	(2)	(3)	(4)
Panel A: Constant elasticity				
log(Traffic)	-0.191***	0.467***	0.344***	0.405***
	(0.014)	(0.017)	(0.018)	(0.025)
Observations	7,347	7,347	7,347	7,347
Within Adj. R^2	0.073	0.159	0.073	0.138
Panel B: Exponential				
Traffic/1,000	-0.205***	0.389***	0.285***	0.394***
	(0.010)	(0.021)	(0.021)	(0.033)
Observations	7,347	7,347	7,347	7,347
Within Adj. R^2	0.108	0.144	0.071	0.150
Link FE	N	Y	Y	Y
Hour FE	N	N	Y	N
Hour-Seg FE	N	N	N	Y

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

OLS estimates of β (Panel A) and β' (Panel B) based on Equations 1 and 2, traffic levels are divided by 1,000 for scale, and link-clustered standard errors are in parentheses. Dependent variables are natural logarithms of time cost. Within R^2 is computed using the standard degrees of freedom adjustment $(N.Obs - N.FE)/(N.Obs - N.FE - K)$ with K estimated parameters, conditional on fixed effects in Columns 2 to 4.

Table 3: Traffic's effect on reliability

	(1) Q90-Q50 % Diff.	(2) Q80-Q20 % Diff.	(3) Inv.Hyp.Sine Std.Dev.
log(Traffic)	19.26***	23.21***	0.443***
	(2.257)	(2.246)	(0.061)
Observations	7,347	7,347	7,347

*p<0.1; **p<0.05; ***p<0.01

Each cell presents an OLS estimate of β based on Equation 1 with full sets of fixed effects and link-clustered standard errors are in parentheses. Dependent variables are percent difference between 90^{th} and 50^{th} percentiles, 80^{th} and 20^{th} percentiles, the inverse hyperbolic sine of speed's standard deviation across weekdays in October 2019.

Table 4: Heterogeneous supply functions

	log(Seconds per metre)			
	(1)	(2)	(3)	(4)
Traffic at Q25	0.129***	0.256***	0.291***	0.286***
	(0.030)	(0.036)	(0.042)	(0.059)
Traffic at Q50	0.177***	0.351***	0.400***	0.393***
	(0.042)	(0.049)	(0.058)	(0.081)
Traffic at Q75	0.260***	0.517***	0.588***	0.578***
	(0.061)	(0.073)	(0.085)	(0.119)
Elasticity difference	0.131***	0.261***	0.297***	0.292***
Q75 − Q25	(0.031)	(0.037)	(0.043)	(0.060)
Speed limit (MpH)	50	30	30	30
Trunk	Y	N	N	N
Signal	N	N	Y	Y
Location	North Circ.	North Circ.	North Circ.	CBD

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Each column uses OLS estimates of equation 3 to impute traffic elasticities at 25^{th}, 50^{th}, and 75^{th} percentiles of the traffic distribution at four structural road scenarios. North Circ. is coordinates returned by a Google Maps search for "North Circular" in December 2022 (on the A406 near Bounds Green station) and CBD is Charing Cross. Each cell presents an elasticity computed using numerical derivatives and link-clustered delta method standard errors are in parentheses.

Table 5: Marginal external costs of simulated trips (GBP)

Supply model	(1) Car commute	(2) Heathrow to Centre	(3) Tolled commute	(4) Non-car commute	(5) Underground trip
(1) Homog. log-log	0.66	2.32	1.44	0.90	0.92
(2) Homog. exp.	0.78	4.27	1.73	0.94	1.09
(3) Heterog. log-log	0.78	3.36	1.78	1.10	1.11
(4) Heterog. exp.	0.89	5.38	2.15	1.19	1.39

Trip marginal external costs average across morning and evening rush hours and swap origin and destination across hours to account for road features that cause differences between inbound and outbound routes (e.g. one-way streets). Doubling each cell gives an estimate the round trip's external cost.

Table 6: Optimal tolls

	Inverse demand elasticity (η)			
	0	-0.67	-1.5	-2
Panel A: Optimal rush hour tolls (GBP/KM)				
(1) Homog. log-log	0.09	0.12	0.12	0.12
	[0.07, 0.12]	[0.09, 0.15]	[0.09, 0.16]	[0.10, 0.16]
(2) Homog. exp.	0.05	0.08	0.09	0.09
	[0.03, 0.06]	[0.05, 0.10]	[0.05, 0.12]	[0.06, 0.12]
(3) Heterog. log-log	0.09	0.11	0.12	0.12
	[0.06, 0.13]	[0.08, 0.16]	[0.08, 0.17]	[0.08, 0.18]
(4) Heterog. exp.	0.06	0.09	0.11	0.11
	[0.04, 0.08]	[0.06, 0.13]	[0.07, 0.15]	[0.07, 0.16]
Panel B: Morn/Mid percent change				
(5) Homog./Heterog.	11.9	–	–	–
log-log	[-1.8, 17.6]	–	–	–
(6) Homog. exp.	27.0	25.0	26.1	26.6
	[-0.5, 43.5]	[-0.7, 40.1]	[-0.4, 42.1]	[-0.4, 42.8]
(7) Heterog. exp.	26.5	24.4	25.6	26.1
	[-0.4, 42.7]	[-0.7, 39]	[-0.6, 41.1]	[-0.5, 41.8]

Panel A is length-weighted mean optimal tolls excluding mid-day and Panel B is mean percent change between mid-day and morning rush. Interquartile ranges are in parentheses. Both exclude 247 links (10%) in the CCZ and 8 links (0.3%) with negative imputed cost elasticities in any model.

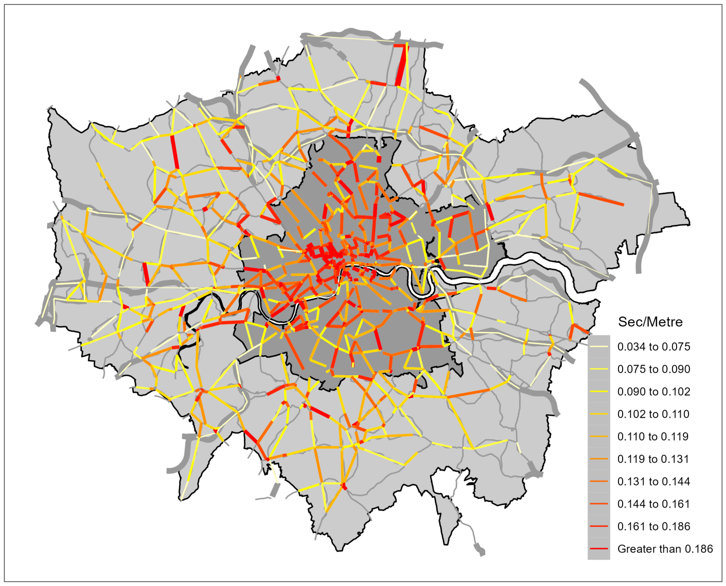
Table 7: Annual deadweight loss under alternative toll regimes (million £)

Tolls	Inverse demand elasticity (η) 0	-0.67	-1.5	-2
(1) None	178.5	111.1	79.0	67.7
(2) Fixed	77.2	27.1	20.1	17.7
per KM	57%	76%	75%	74%
(3) Time	76.1	26.0	19.3	17.0
varying	57%	77%	76%	75%
(4) Congestion	35.9	6.4	5.0	4.5
varying	80%	94%	94%	93%

DWL estimates sum across link-hours and exclude links in the CCZ or with negative imputed cost elasticities in any model and assume 260 nine-hour workdays per year. Percentages indicate each policy's reduction in deadweight loss.

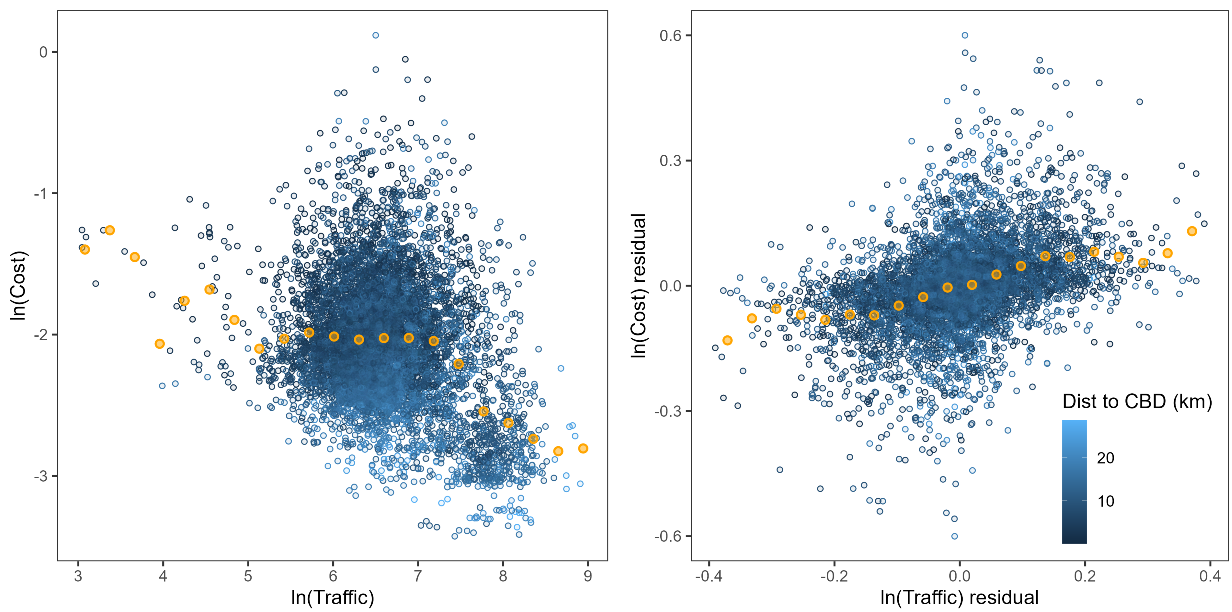
7 Figures

Figure 1: Sample coverage and time cost distribution



The dark shaded area is Inner London and Outer London is light grey. Grey paths are a simplified representation of region's major roads taken from OpenStreetMaps. Thick grey paths have speed limits > 80 Km/Hr, thin grey paths have speed limits ≤ 80 Km/Hr, and side roads are not shown. Straight lines are road segments with thickness and color corresponding to time cost averaged across hours of day and directions of travel.

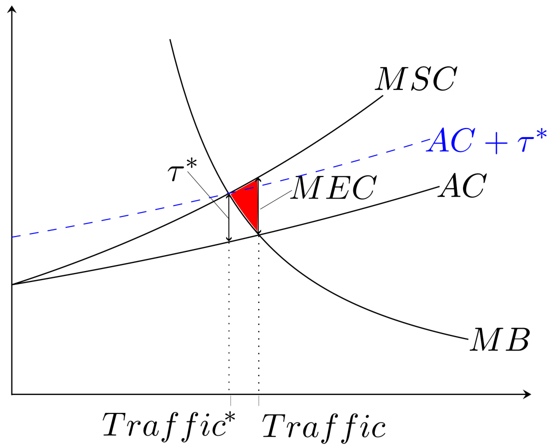
Figure 2: Time costs and traffic



(a) Raw data (b) Within-link residuals

Large orange points are binned means at ventiles of log traffic.

Figure 3: Congestion costs and optimal tolls



Vertical axis measured in GBP, horizontal in number of vehicles, and AC is an exponential travel supply curve. Unregulated DWL is shaded in red and observed $Traffic$ exceeds optimal $Traffic^*$.