

The City-Wide Effects of Tolling Downtown Drivers: Evidence from London's Congestion Charge.

Ian Herzog*

December 2020

(Job Market Paper)

[Click here for latest version](#)

Abstract

I study the effects of London England's Congestion Charge on regional traffic, commuting, and economic activity's spatial distribution. In 2003, London began tolling drivers to enter its central business district on workdays and this policy immediately reduced downtown traffic. I measure exposure to tolled traffic throughout London's regional road network and show that the policy also reduced rush-hour traffic on radial roads leading downtown. I then use London's Congestion Charge as a natural experiment to identify ensuing effects on commuting by car and public transit. Reduced form estimates show that traffic decreases both the number of commuters and their driving rates at the margin. In a quantitative spatial model, the effect on the number of commuters suggests that endogenous location choices, wages, and rents influence the congestion charge's equilibrium effects, so I use a model with endogenous traffic externalities and mode choice for policy analysis. Simulations suggest that London's Congestion Charge increases driving rates among untolled commuters, disproportionately benefits low-skill workers, and gives the region's commuters £32 million in annual benefits. These effects on traffic and welfare are limited by commuters' location and mode choices, which are sufficiently traffic-elastic that London's Congestion Charge induced demand for untolled suburban roads.

*Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario, M5S 3G7, Canada. E-mail ian.herzog@mail.utoronto.ca. Thanks to Nathaniel Baum-Snow, Jonathan Hall, and Stephan Heblich for their guidance and support. I also thank Rosa Sanchis-Guarner for invaluable assistance and seminar participants at the 2020 Virtual Meeting of the Urban Economics Association, the 2020 UEA PhD workshop, and the University of Toronto for comments and suggestions. This research is supported in part by funding from the Social Sciences and Humanities Research Council of Canada.

1 Introduction

Traffic congestion costs time, causes collisions, exacerbates air pollution, erodes drivers' health, and reduces life satisfaction.¹ Economists have proposed congestion charging schemes that internalize these externalities and estimate the annual value of potential time savings at over 30 billion dollars in the United States alone (Couture et al., 2018). Cities increasingly turn to cordon fees, tolling entry to physically small but economically important areas, but many caution that these policies also tax productive employment clusters (Zhang and Kockelman, 2016; Brinkman, 2016). In practice, understanding how this trade-off between congestion and agglomeration externalities affects regional traffic and commuting patterns can answer open questions about a cordon fee's welfare implications.

This paper sheds light on these issues using a new identification strategy and spatial equilibrium model to study the regional effects of Central London's Congestion Charge. This policy became a formative example of a cordon fee in 2003 when London began charging cars and trucks £5 to enter its central business district and had an immediate impact on downtown road congestion. I measure the policy's ripple effects on suburban roads and identify ensuing effects of road traffic on commuters' location and mode choices. These patterns then calibrate a structural model that accounts for endogenous wages, rents, and traffic externalities between overlapping commutes for welfare analysis.

I find that by discouraging driving on tolled routes, London's Congestion Charge reallocates roadspace to cross-town commuters, impacts decisions between driving and taking public transit, and affects employment throughout the city. Taking a regional view, Central London's Congestion Charge gives commuters revenue-neutral gains of £32 million per

¹See Molnar and Mangrum (2018), Russo et al. (2019), Yang et al. (2020), and Herzog (2020) for evidence traffic's effect on travel time costs; Green et al. (2016) for collisions; Zhang and Batterman (2013), Simeonova et al. (2018), and Green et al. (2018) for air pollution; Christian (2012), Hilbrecht et al. (2014), and Kunn-Nelen (2016) for indicators of physical health and well-being; and Kahneman and Krueger (2006), Office of National Statistics (2014), and Adam et al. (2018) for subjective well-being and life satisfaction.

year, equivalent to about ten percent of annual toll revenue (Transportation for London, 2012) or seven percent of the remaining £477 million annual value of time lost to congestion (Herzog, 2020). I also find that these gains were progressive because low-skill commuters benefit from suburban traffic reductions and have loose ties to tolled workplaces. However, London’s Congestion Charge also encouraged driving on untolled routes that share roads with tolled traffic and my estimates suggest that the policy’s benefit would be several times larger had it not induced demand for suburban roads.

I build to these conclusions in three steps. First, I create an Index of Congestion Charge Exposure measuring the effects of Central London’s toll on regional traffic patterns. In the process, I develop a new method to link commutes between neighbourhood pairs with traffic counts at points in Greater London’s road network. Second, I establish causal effects of road traffic externalities on commuting by car and public transit and present a model where they are critical inputs to welfare and policy analysis. Third, I extend the model to a general equilibrium setting with endogenous road traffic to quantify effects of Central London’s Congestion Charge on regional employment, population, mode choice, and welfare.

For my empirical analysis, I link decennial census commute flow data to London’s dense traffic monitor network. I use the OpenStreetMap project’s Open Source Routing Machine to compute optimal routes through the city and attribute traffic counts to particular commutes. Routing data also reveal where commutes intersect, allowing me to identify how tolled traffic spills over onto untolled commutes. The result is the first data set of long-run changes in commuting and traffic patterns within a city that is suitable for estimating traffic congestion’s effect on commuting and employment.

I use these data to develop an Index of Congestion Charge Exposure measuring how Central London’s toll affects traffic throughout the city. This index breaks Greater London’s road network into links and measures pre-toll traffic that would be tolled after 2003.

Congestion Charge Exposure's spatial distribution reveals that radial roads are most affected because they are frequently used by initially high-traffic routes to Central London. I then show that Congestion Charge Exposure predicts declining traffic after 2003 and argue that this relationship is causal, establishing that Central London's Congestion Charge affected traffic on untolled commutes throughout the region.

The second step develops a model of commuters' travel demand and estimates central parameters. The model features heterogeneous commuters solving a nested discrete choice problem; first choosing where to live and work based on endogenous wages, rents, and transportation opportunities, then choosing whether to commute by car or public transit between given locations. Travel demand aggregates across utility-maximizing commuters who pay tolls to drive on certain commutes and incur a utility cost from driving in traffic. Since traffic is costly, increasing it makes commuters reduce driving rates or choose different residences and workplaces to the extent that they can substitute to public transit or sort towards locations linked by low-traffic routes.

Commuting's traffic elasticities are critical inputs to policy analysis and speak to the necessity of studying London's Congestion Charge in a general equilibrium framework. However, correlations between traffic growth and commute flows are uninformative since they conflate supply and demand. With this in mind, I identify reduced form effects and calibrate the structural model using Congestion Charge Exposure to isolate exogenous variation in the way Central London's toll affects traffic for untolled commuters.

Reduced form estimates show that exogenously reducing traffic between a residence-workplace pair causes commuters to sort towards those locations and increases their driving rates. In particular, decreasing traffic by ten percent increases the number of commuters 9.2 percent and increases commuters driving by 12.1 percent relative to other modes. The effect on mode choice is large and the significant commuting elasticity suggests that traffic affects

residence and workplace locations, motivating structural analysis to account for general equilibrium effects.

To quantify the welfare effects of London’s Congestion Charge, I specify the structure of labour demand, housing supply, and traffic externalities and run counterfactual policy simulations. Commuting decisions determine neighbourhoods’ labour supply while labour demand reflects agglomeration externalities as in Ahlfeldt et al. (2015) and differences between high- and low-skill labour as in Tsivanidis (2019) and Lee (2019). The model also assumes commuters impose traffic on the roads comprising their routes to work, generating endogenous traffic externalities along a realistic road network.

I quantify the model by estimating its central parameters using the reduced form analysis and calibrating location and road characteristics to match Greater London’s commuting and traffic patterns. The model also provides a framework for integrating census counts of high- and low-skill population and employment with commute flows that do not separate by worker skill. Given commute costs, the model implies unique workplace and residential fixed effects rationalizing each skill group’s spatial distribution of population and employment. I use this relationship to estimate skill-specific commuting and run policy counterfactuals that identify unequal welfare effects.

Counterfactual simulations show that London’s Congestion Charge increases mean utility by 0.025 percent and disproportionately benefits low-skill commuters. Aggregate gains come alongside a 0.07 percent reduction in welfare inequality as high- and low-skill workers both benefit from traffic reductions but tolled workplaces are more important for high-skill commuters. Counterfactual simulations also show that the policy shifts employment immediately outside the Congestion Charge Zone and increases commuters’ driving rates.

Simulations also suggest that London’s Congestion Charge reduced the region’s road traffic by 0.715 percent and that this effect would be larger without endogenous sorting and

substitution away from public transit. In the baseline counterfactual, I find that the policy suburbanizes population and increases the number of commuters driving by 0.581 percent. I also find that endogenous traffic reductions concentrate in the tolled area and on radial roads leading downtown but the policy made traffic grow on many untolled orbital roads. Combined, sorting and mode choice eliminate fourty percent of the potential traffic decline predicted by a model where commuters do not cause externalities.

These results suggest that commuters often choose public transit or convenient locations to avoid traffic and readily shift to longer drives when roadspace becomes available. In the model, London’s Congestion Charge removes downtown-bound drivers from the road and creates incentives to drive on overlapping untolled commutes. This is an example of the fundamental law of road congestion documented by Duranton and Turner (2011), Hsu and Zhang (2014), and Garcia-Lopez et al. (2020)—tolling an important destination rather than in-demand roads reduces downtown driving but reallocates untolled roadspace to other trips, limiting regional traffic reductions.

More broadly, this paper contributes to the literature studying the determinants of urban structure and economic activity. Despite a long-standing interest in the welfare implications of commute costs, congestion, and road pricing (Verhoef, 2005; Brinkman, 2016; Zhang and Kockelman, 2016) empirical analysis of this issue is scant (Proost and Thisse, 2019). This is likely due to data and methodological limitations. I contribute a new framework for connecting commuting and traffic data, use a unique natural experiment to estimate effects of congestion externalities and road pricing on commuting, and develop a general equilibrium model to aggregate commuters’ choices to outcomes in labour and housing markets.

I also contribute to a literature using quantitative general equilibrium models and commuting data to study transportation policies (Tsivianidis, 2019; Tyndall, 2018; Severen,

2019; Heblich et al., 2020). This work builds on the literature estimating the effects of urban transportation on economic activity (Baum-Snow et al., 2005; Baum-Snow, 2007; Duranton and Turner, 2011; Gonzalez-Navarro and Turner, 2018; Jerch et al., 2020) using discrete choice models adapted from international trade (Eaton and Kortum, 2002; Ahlfeldt et al., 2015; Monte et al., 2018). My work also relates to Fajgelbaum and Schaal (2017) and Allen and Arkolakis (2019) who consider the effects of endogenous congestion on interregional trade routes. This paper is the first empirical study of road pricing using a quantitative spatial equilibrium framework that incorporates traffic externalities along a realistic road network for counterfactual estimation and welfare analysis.

My results highlight the importance of considering sorting and the labour market when measuring road pricing's benefits; equilibrium welfare gains are an order of magnitude smaller than those implied by a traditional measure of the value of travel time saved. Related work often considers tolls to alleviate bottleneck congestion when drivers respond by rescheduling their trips (Vickrey, 1967; Arnott et al., 1993; Parry, 2009; Hall, 2020). Others analyse travel surveys (Couture et al., 2018; Akbar and Duranton, 2017), exploit temporary shocks (Gibson and Carnovale, 2015; Hanna et al., 2017), or experimentally simulate congestion pricing policies (Martin and Thornton, 2017; Kreindler, 2018). I add to this literature by identifying traffic's effect on the location of economic activity and using structural estimates of commuters' travel demand for policy analysis.

Finally, I measure the regional welfare effects of London's canonical policy experiment. Leape (2006) summarises early analysis of London's Congestion Charge and Green et al. (2016) and Green et al. (2018) show evidence that London's Congestion Charge affected traffic accidents and air pollution in the tolled zone. Tang's (2018) partial equilibrium analysis summarises amenity improvements by estimating effects on home prices in the tolled area. Despite these benefits, Zhang and Kockelman (2016) and Brinkman (2016)

argue that cordon fees can have negative welfare effects if they break productive employment clusters. My analysis of London’s Congestion Charge finds that sorting attenuates the policy’s value but agglomeration externalities are not the culprit. Cordon fees are limited because they create incentives to drive on untolled commutes that share roadspace with tolled traffic.

This paper proceeds by discussing the data in section 2. Section 3 then describes London’s economic geography and congestion charge policy. Section 4 introduces an Index of Congestion Charge Exposure and establishes that tolling downtown affected traffic throughout Greater London. Section 5 fixes ideas in a model of commuters’ travel demand and section 6 estimates road traffic’s effect on commuting. Section 7 embeds this demand system in a structural model that I calibrate in section 8 and use to run counterfactual simulations in section 9. Finally, section 10 offers concluding remarks.

2 Data

I use data from three main sources. Commute flows, employment, and population come from the United Kingdom’s 2001 and 2011 decennial censuses. Traffic data come from the UK Department for Transport’s (DfT) network of traffic count points. I spatially join commute and traffic data using routes computed by the OpenStreetMap project’s Open Source Routing Machine (OSRM). Appendix A.1.1 describes additional data which include boundary shapefiles and public transit information from Transportation for London and the Greater London Authority’s open data repositories.

I restrict my study area to the Greater London Region and develop a panel of outcomes in 2001 and 2011 for middle super output areas (MSOAs) which are census geographies designed to contain 2,000 to 6,000 resident households. Normalizing to constant geographic boundaries delivers a balanced panel of 967 MSOAs.

2.1 Commute flows

Commute flows between MSOA pairs come from the 2001 and 2011 UK census interaction data sets.² I use these data to measure the number of workers commuting between MSOA pairs by car and other means including taxis, public transit, or walking. My sample is a balanced panel of 934,122 directional MSOA pairs, excluding commuters living and working in the same MSOA.³ I observe each pair's car and non-car commuting in both 2001 and 2011 for a total of 3,736,488 commute flow counts.⁴

2.2 Employment and population

I define each area's residential population as the number of employed individuals living there who do not work from home; employment is the same universe measured at usual workplaces reported to the census. I observe employment and population of workers in three skill groups and two modes of transportation for all MSOAs in Greater London.

The modes of transportation are car and non-car and skill groups are based on the three-class UK National Statistics Socio-economic Classification (NS-SeC) which I aggregate into two skill groups that approximate college and non-college workers. High-skill occupations include managers, professionals, small business owners, intermediate occupations (e.g. nurse, salesperson), and technical occupations (e.g. plumber, chef), which all require some form of post-secondary education or training. Low-skill occupations include routine and semi-routine (e.g. receptionists, care workers) occupations.⁵

²Data are accessed via the WICID query system.

³I also exclude counts of individuals who report no fixed workplace address.

⁴The UK census changed their method for anonymizing small cells between 2001 and 2011. The data appendix describes my procedure for adjusting these data so that they are comparable over time.

⁵High skill is NS-SeC 1 through 5 and low skill is NS-SeC 6 and 7. Unfortunately, I do not separately observe intermediate and technical occupations.

2.3 Traffic

Traffic data come from the UK department for transport (DfT) who count traffic flows at roughly 10,000 count points (CPs) across England's major road network. Trained enumerators count hourly traffic from 07:00 to 19:00 on a non-winter weekday annually or on two, four, or eight year cycles.⁶ My raw data cover all years from 2000 to 2015 and are an unbalanced panel of 1,200 to 2,200 counts in Greater London depending on the year.⁷ To control for composition effects caused by high-traffic CPs being counted most frequently, I restrict my sample to those observed at least once from 2000 to 2002 (pre-toll) and again from 2003 to 2015 (post-toll). This gives a balanced panel of 1,399 count points whose spatial distribution is shown in figure 1 which confirms their comprehensive coverage of Greater London's road network.

I define a road link as a travel direction-count point pair and observe each link's hourly traffic volume each time it is counted. I also observe link length, end point locations, and enumerators' coordinates. I aggregate across hours to calculate morning and evening rush hour traffic each time a link is counted. To this end, I define morning rush-hour traffic as total traffic volume counted between 07:00 and 10:00 and evening rush-hour traffic as volume counted from 16:00 to 19:00. Finally, I average across years from 2000 to 2002 and again from 2003 to 2015 to compute pre- and post-toll traffic volumes during each rush hour.

2.4 Routes

I compute time-minimizing paths between all MSOA pairs in Greater London using the OpenStreetMap project's Open Source Routing Machine version 5.22.0 (Luxen and Vetter,

⁶Counts are restricted to weekdays between March and October excluding public and school holidays.

⁷I observe a total of unique 2,991 count points located in the London Region and omit observations with imputed and automatic traffic counts. I end my sample at 2015 to avoid measurement issues caused by the post-2015 introduction of automatic counts.

2011). I run the router 934,122 times through a fall 2019 snapshot of the London Region’s road network to compute fastest paths between centroids of all directed MSOA pairs.⁸ I observe a detailed polyline representing the route’s path through London’s roads.

I define a commute’s route in each rush hour as the set of count points within 50 metres of its path. For each directed MSOA pair, the algorithm builds a 100 metre buffer centred along its path and intersects that buffer with traffic enumerator coordinates. Road links with enumerators located within the buffer are matched with the MSOA pair. The algorithm assigns a travel direction to each link-path intersection by finding the path’s closest vertex to each matched link and noting the dominant zonal or meridional direction travelled one vertex towards the destination.⁹ A commute’s morning route R_{ij}^{morn} is the set of road links matched to the path from i to j and the evening route R_{ij}^{eve} is the set used for return travel.¹⁰

Appendix figure A.1 presents an example of the spatial join. The bottom right corner is a residence MSOA centroid and the top left corner is a workplace MSOA centroid. Black dots indicate count points and highlight those matched to one direction of travel between this MSOA pair. Matching the 934,122 routes to the balanced panel of road links preserves 922,550 MSOA pairs with traffic measured both pre- and post-toll.

I aggregate across road links to compute rush hour traffic volume for each residence-workplace pair in the census data. An MSOA pair’s morning traffic is the sum of morning rush-hour traffic volume on road links comprising the workplace-bound route and evening traffic is the sum of evening traffic volume on the residence-bound route. Formally, those living in i and working in j in period t face morning traffic $\sum_{\ell} \mathbb{1}\{\ell \in R_{ij}^{morn}\} \sum_{h=7}^9 traffic_{h\ell t}$

⁸The road data come from the Geofabrik online repository.

⁹If a CP’s closest vertex is the route’s destination, I take travel direction from the second last vertex to the destination.

¹⁰Morning and evening a commute’s workplace- and residence-bound paths can differ due to street elements such as one-way roads and turn restrictions but it is always the case that $R_{ij}^{morn} = R_{ji}^{eve}$.

and evening traffic $\sum_{\ell} \mathbf{1}\{\ell \in R_{ij}^{eve}\} \sum_{h=16}^{18} traffic_{\ell h t}^h$ where $traffic_{\ell h t}^h$ is the traffic volume counted on link ℓ in hour h , each road link ℓ is a directed count point, and period t is either pre- or post-toll. I then measure traffic faced by commuters living in i and working in j in period t as the average of their morning and evening rush-hour traffic.

Section 4 describes a related procedure for computing commuters' Congestion Charge Exposure and I control outliers by dropping MSOA pairs with traffic growth or charge exposure in the top or bottom 2.5 percent of the distribution. Appendix table A.1 summarises pairwise traffic and commuting data in the main analysis sample.

3 Background

3.1 London's economic geography

The Greater London Region extends over 25 kilometres from the city centre to cover 33 local authorities and is surrounded by a greenbelt that has restricted urban expansion since the 1960s. Central London has had an active congestion charge since 2003 which section 3.2 describes in detail. Figure 1 maps Greater London and its major arterial roads; line widths correspond to speed limits and dots are count points used to measure traffic. The purple area is Outer London and Inner London is subdivided into the yellow Congestion Charge Zone and green adjacent area.

London's road network relies on several disjoint ring and radial roads and its high capacity motorways are concentrated in Outer London. Meanwhile, appendix figure A.2 shows that the region's employment is concentrated in Central London with pockets of employment along major transportation corridors and suburban employment centres in the south. Figure A.3 confirms that London also exhibits a strong commuter gravity pattern and appendix figure A.4 shows that population centralized and employment decentralized

over the 2000s. Finally, table 1 shows that one third of London commuters drove to work in 2001 and this share fell to one quarter in 2011. Driving is more popular among those who live and work in Outer London, but even in the suburbs driving's share of commutes has fallen from about two thirds in 2001 to just under half in 2011.

3.2 London's Congestion Charge

Road pricing proposals in the UK date back to the Smeed commission's 1964 report to the Department for Transport who were studying economic and technical possibilities for future policies (Smeed, 1964). The Smeed report was ahead of its time but by the late 1990s political will and technological advances made congestion charging a serious consideration in Central London.

London's road speeds fell steadily from the 1970s to the late 1990s (Santos et al., 2008) and a 1999 survey identified public transit and congestion as locals' two most salient policy concerns (Leape, 2006). That same year, the Greater London Authority Act created London's metropolitan government and gave future London mayors power to enact road user charging (Greater London Authority, 1999). Ken Livingstone was elected London's first mayor in 2000 and his campaign promises had included downtown road pricing. Transportation for London (TfL) subsequently began studying policy options and enacted a £5 cordon fee for vehicles entering Central London in February of 2003.

London's Congestion Charge Zone (CCZ) comprises approximately 20 km² within the Inner Ring Road containing most of the city's major attractions and central business district. The Inner Ring Road remains tolled and used by through-traffic to avoid the CCZ. The CCZ was temporarily expanded westward in 2007 to include Chelsea and Kensington, but the Western Extension Zone (WEZ) charge was unpopular with residents and abandoned in 2010. Figure 1 shows the CCZ relative to the London Region (dots are traffic

count points and lines are London’s major road network) and appendix figure A.5 focuses on a subset of Inner London to detail the CCZ’s geography.

Central London’s Congestion Charge started at £5, grew over time, and is levied on motor vehicles entering the CCZ between 07:00 and 18:30 on weekdays using automatic license plate recognition. The policy exempts motorcycles, vehicles-for-hire, and alternative fuel vehicles and gives zone residents a ninety percent discount.¹¹ Figure 2 presents a timeline of relevant policy developments during my study period.

Transportation for London (2003) documented a 27 percent drop in motor vehicles entering Central London after the toll’s introduction and Leape (2006) asserts that most of this is substitution to public transit and detouring around the zone. All told, Tang (2018) estimates a causal effect on Central London traffic closer to a nine percent drop in traffic that capitalized into a three percent rise in home values. In a follow-up survey, Outer London residents reported that they reduced driving to the CCZ for a range of activities after its introduction (Transportation for London, 2005).

3.3 Regional traffic declines

Figure 3 maps Greater London’s widespread traffic decline in traffic during the 2000s. Each line segment is the roadway monitored by a single count point, links intersecting London’s Outer Ring Road are highlighted in blue, line colours denote deciles of log-traffic change, and thickness is proportional to traffic change’s distance from zero. London’s median count point saw traffic fall by approximately ten percent in the years following the toll’s introduction. Traffic declined most on central roads and suburban links on radial roads. Meanwhile, traffic grew in the eastern suburbs and on major circumferentials including London’s North Circular Road.

¹¹The charge was extended to private vehicles for hire including minicabs and Uber in late 2018.

I now provide suggestive evidence that Central London's toll shifted traffic away from CCZ-bound roads and motivate the systematic evaluation presented in section 4. To this end, I define radial road links as those not intersecting the Outer Ring Road and run the event study regression:

$$\ln(\text{traffic}_{\ell t}) = \sum_{\substack{\tau=2000, \\ \tau \neq 2002}}^{2015} D_{\tau}^t \times \beta_{\tau} \text{radial}_{\ell} + \sum_{\substack{\tau=2000, \\ \tau \neq 2002}}^{2015} D_{\tau}^t \times f_{\tau}(\text{lat}_{\ell}, \text{lon}_{\ell}) + \alpha_t + \alpha_{\ell} + e_{\ell t} \quad (1)$$

where $\text{traffic}_{\ell t}$ is rush-hour traffic volume observed on link ℓ in year t from 2000 to 2015, radial_{ℓ} indicates links that are not on the Outer Ring Road, $D_{\tau}^t = 1\{\tau = t\}$ are annual dummies excluding 2002, $f_{\tau}(\text{lat}_{\ell}, \text{lon}_{\ell})$ is a year-specific second-order polynomial in longitude and latitude, and α_t and α_{ℓ} are year and link fixed effects. The semi-elasticities β_{τ} measure traffic trends on radial roads relative to nearby links on the Outer Ring Road. Looking forward, section 4 shows that many radial roads are on important routes to the CCZ, suggesting that negative values of β_{τ} after 2002 indicate Central London's Congestion Charge reduced radial-road traffic.

Figure 4 presents least squares estimates of β_{τ} normalized to zero in 2002—the final year without Central London's toll—error bars denote link-clustered 95 percent confidence intervals. The plot shows that the Outer Ring Road was on a similar traffic trend to nearby radial roads before 2003 but relative declines in radial-road traffic emerge after Central London's toll began. This pattern persists for several years and a corresponding difference-in-differences estimate shows that radial-road traffic fell by 6.8 percent relative to ring-road traffic in the post-toll period. Increasingly negative β_{τ} estimates after 2006 may reflect the 2005 CCZ toll increase and the Western Extension Zone charge that was active from 2007 to 2011.¹² Section 4 captures this trend with a continuous measure of Congestion Charge

¹²Appendix figure A.6 shows a similar, albeit less precisely estimated, pattern using a variant of equation 1 saturated with year-specific controls for being in the CCZ, WEZ, or Inner London.

Exposure and an identification strategy that provides direct evidence that Central London's toll affected regional traffic patterns.

4 Congestion Charge Exposure

The Congestion Charge Zone ends in Central London but indirectly affects traffic on roads heading downtown. Road links that are most frequently used en-route to Central London are most exposed to the congestion charge and see the largest reductions in traffic following the toll's introduction. In this section, I develop an Index of Congestion Charge Exposure to measure the ripple effects of Central London's toll on suburban traffic and establish that Central London's toll affected regional traffic patterns. I then argue that Congestion Charge Exposure captures exogenous variation in traffic caused by the policy and provides a basis for measuring effects on regional traffic and commuting.

4.1 Measurement

Congestion Charge Exposure measures the pre-2003 traffic crossing each road link that would be tolled in the post-policy period. Congestion Charge Exposure on road link ℓ in rush hour $h \in \{morn, eve\}$ is

$$CCE_{\ell}^h = \sum_{r,s} CCZ_{rs} 1\{\ell \in R_{rs}^h\} \frac{traffic_{rs0}^h}{length_{rs}^h} \quad (2)$$

where $1\{\cdot\}$ is the indicator function, R_{ij}^h is the set of road links connecting residence i and workplace j at rush hour h , $CCZ_{ij} = CCZ_j(1 - CCZ_i)$ indicates tolled commutes, $traffic_{ij0}^h$ is pre-toll traffic associated with commuting between residence i and workplace j , and $length_{ij}^h$ is length of monitored roads on that commute.

Figure 5 demonstrates the intuition behind CCE with an illustrative example. The left

panel shows an Inner London road link crossed by tolled routes shown as blue paths and origin MSOAs are shaded in proportion to the amount of morning traffic contributed to this link's morning charge exposure.¹³ Origins are all east of the link and fan outward as the road network leads drivers to this link en-route to the CCZ. Looking forward, variation in Congestion Charge Exposure tracks how Central London's toll affects regional traffic patterns and identifies traffic's effect on commuting.

Figure 6 maps Congestion Charge Exposure's spatial distribution. Lines are road segments monitored by each count point with thickness and darkness corresponding to link-level CCE values averaged across directions and rush hours. The map shows that CCE is largest on radial roads running east-west across the city. In addition, several southern road links headed towards the central business district are very exposed to Central London's toll. Meanwhile, London's North Circular Road does not stand out as it is not exposed to substantial amounts of tolled traffic.

From a commuter's perspective, Congestion Charge Exposure aggregates across road links in the same fashion as traffic. Specifically, the Congestion Charge Exposure Index of those living in i and working in j is

$$CCE_{ij} = \frac{1}{2} \sum_h \sum_{\ell} 1\{\ell \in R_{ij}^h\} CCE_{\ell}^h \quad (3)$$

which sums exposure across road links en-route and averages across the morning trip to work and evening trip home. An MSOA pair's CCE summarises its exposure to traffic generated by drivers travelling to the CCZ.

Looking forward, differences in Congestion Charge Exposure across otherwise similar commutes establish that Central London's toll affects regional traffic patterns, causes sorting and general equilibrium effects, and calibrate commuters' roadspace demand for coun-

¹³Pre-toll morning traffic contributed is $\sum_j CCZ_{ij} 1\{\ell \in R_{ij}^{morn}\} \frac{traffic_{ij0}^{morn}}{length_{ij}}$.

terfactual policy simulations. The right hand panel of figure 5 sheds light on the source of this cross-commute variation by contrasting commutes with high and low Congestion Charge Exposure. The red commute is more exposed as it shares the road with CCZ-bound traffic on the westbound A13. Meanwhile, the yellow commute has relatively low Congestion Charge Exposure as it mostly travels parallel to the CCZ and spends less time on major radial roads.

Figure 7 plots commutes' Congestion Charge Exposure against log-changes in traffic between the pre- and post-toll periods. The darkest area is a 95 percent confidence interval around conditional means and the grey shaded areas trace conditional quantiles at fifty CCE bins. The negative correlation indicates that Congestion Charge Exposure predicts declining traffic after Central London's toll began with explanatory power throughout the distribution, a least squares fit on the underlying data has an R^2 of 0.09. Section 4.2 proceeds to show that this relationship is robust to rich geographic controls and argues that it reflects a causal effect of Central London's toll on regional traffic patterns.

4.2 Effect on regional traffic

I establish that Central London's Congestion Charge affected regional traffic patterns using the quasi-difference-in-differences regression

$$\ln(\text{traffic}_{ijt}) = \varphi \text{CCE}_{ij} \times \text{Post}_t + X'_{ij}\Gamma_t + \gamma_{b(ij)t} + \gamma_{it} + \gamma_{jt} + \gamma_{ij} + \nu_{ijt} \quad (4)$$

where traffic_{ijt} is traffic faced when choosing to commute between residence i and workplace j in year t , pre- or post-toll, and Post_t is a post-toll dummy. The vector X_{ij} includes measures of public transit availability, road characteristics, and indicators of commutes into

the CCZ or WEZ, all interacted with post-2003 dummies.¹⁴ Distance-by-year fixed effects $\gamma_{b(ij)t}$ flexibly control for differential trends in one kilometre bins in driving distance. The fixed effects α_{it} and α_{jt} capture all changes at the residence and workplace levels and α_{ij} forces identification based on changes over time. The parameter of interest φ captures the effect of Central London's toll on traffic faced by an average commuter independent of their tolled status. This parameter excludes the toll's direct effect since I explicitly control for commutes crossing either the WEZ or CCZ boundary.

Revisiting figure 5 elucidates how least squares estimates of equation 4 identify CCE's effect on traffic. The right panel shows similar length commutes to the same workplace where the red commute has higher CCE because of its reliance on the westbound A13 which is also an important route downtown. However, directly comparing these commutes' traffic trends will not identify a causal effect if the red residence's distance from the North Circular Road directly influences traffic growth. Residence-by-year fixed effects account for this by forcing comparisons relative to the average change in traffic from each residence—the comparison illustrated here identifies a negative effect of CCE on traffic only if the red commute's traffic falls or the yellow commute's traffic grows relative to other trips from their residences.

Table 2 presents least squares estimates of equation 4. CCE is scaled to have a standard deviation of one and standard errors are two-way clustered at both residence and workplace levels. All regressions include MSOA-pair fixed effects and are weighted by one plus initial

¹⁴ Public transit controls include dummies for residence and workplace centroids both 1,500 metres from a rail transit station, 1,500 metres from a Tramlink light rail station, and additional dummies for pairs on the same rail transit or Tramlink line. Tramlink is a suburban light rail system using a mixture of dedicated rights of way and at-grade track. Rail transit includes the London Underground, Overground, Docklands Light Rail, TfL rail, and Thameslink. I also include dummies for residence and workplace within 750 metres of the same express bus line, 500 metres of the same 24 hour bus line, and 500 metres from the same daytime-only bus line and control for the share of monitors observed each year, crossing the WEZ boundary, and crossing the CCZ boundary. Finally, I control for the natural logarithm of the total number of count points behind commute in each period, this is not interacted with a post dummy and captures additional measurement error.

commute flows.

Negative effects throughout table 2 indicate that Central London's Congestion Charge affected regional traffic patterns and that CCE is a reasonable measure of effects on untolled commutes. Comparing columns 1 and 2 shows that controlling for distance, workplace, and residence-by-year fixed effects attenuates CCE's effect on traffic growth. Column 3, which includes a complete vector of fixed effects and controls, associates a one standard deviation increase in Congestion Charge Exposure with a 2.6 percent decrease in traffic. Change in log-traffic has a standard deviation of 0.205 in this sample, so a one standard deviation increase in CCE moves a commute 13 percent of a standard deviation down in the distribution of traffic growth.

Appendix table A.2 shows that CCE's effect on traffic is smaller for far apart MSOA pairs, diminishes in CCE, is independent of whether the trip involves Outer London, and is larger for pairs with poor transit access. Appendix table A.3 shows that a similar relationship holds at the road link level and that CCE's effect on traffic is larger during the evening rush hour. Appendix table A.4 goes on to show that estimates are robust to dropping MSOAs in the CCZ, WEZ, or both; controlling for trends on pairs with new rail transit connections after 2001; and controlling for trends on the interaction of residence and workplace distance to London's central business district. Finally, appendix table A.5 shows that results are robust to using proxies for Congestion Charge Exposure that are based an unweighted sum of tolled routes crossing each link or that treat the WEZ as tolled. In what follows, I investigate the effects of these traffic spillovers on commuting and economic geography.

5 A theory of commuters' travel demand

I now present a discrete choice model of commuters' location and mode choices as a basis for estimating effects of London's Congestion Charge on economic geography and welfare. In the model, commuting summarises heterogeneous workers' location choices as in Ahlfeldt et al. (2015). Workers face a two-step discrete choice problem in which they first choose locations to live and work based on wages, rents, amenities, and transportation opportunities then choose a mode of transportation. These assumptions lead to a nested discrete choice demand system over locations and commute modes with a log-linear commuter-gravity equation and a logit mode choice as in traditional partial equilibrium treatments of travel demand (McFadden, 1974).

The model demonstrates the importance of estimating congestion's effect on commuting and shows how choice shares and preferences map to first-order benefits of reducing congestion that are equivalent to the value travel time saved (Small et al., 2007). The model also shows that traffic must have empirically relevant general equilibrium effects if it has a causal effect on commuting. Looking forward, section 7 embeds this demand system into a general equilibrium framework to account for inequality and endogenous effects of commuting decisions on road traffic, housing, and labour markets.

5.1 Preferences

The city is partitioned into N neighbourhoods that act as both workplaces and residences for a fixed population Ω with $\pi_g \Omega$ workers in each skill group g . Residence i provides group g an amenity B_{ig} , costs Q_i per unit floorspace in rent, and workplace j offers income w_{jg} . All workers have Cobb-Douglas preferences over residential floorspace and a consumption good comprising β of their spending and incur a disutility of $d_{ijm} \geq 1$ from commuting between i and j on mode m .

Worker ω has idiosyncratic preferences $\epsilon_{ij\omega}$ over locations and $\nu_{m\omega}$ over modes of transportation distributed Fréchet with cumulative distribution functions $G(\nu) = e^{-\nu^{-\eta}}$ and $F(\epsilon) = e^{-\epsilon^{-\theta}}$ independent of skill. Workers form these preferences sequentially, first observing $\epsilon_{ij\omega}$ and choosing a residence-workplace pair ij then observing $\nu_{m\omega}$ and choosing a commute mode.¹⁵

Ex-post, worker ω of skill g living in i , working in j , and commuting on mode m has indirect utility

$$u_{ijmg\omega} = \frac{B_{ig}}{Q_i^{1-\beta}} \frac{w_{jg}}{d_{ijm}} \nu_{m\omega} \epsilon_{ij\omega}$$

which is the result of Cobb-Douglas preferences fixing the optimal consumption bundle for each commute ij . Workers then choose i , j , and m to maximize utility given wages, rents, amenities, and commute costs.

5.2 Commute costs

Each commute ij is served by two modes of transportation: driving a car c and other modes o . The cost of commuting from i to j on mode m is

$$d_{ijm} = toll_{ij}^{1\{m=c\}} traffic_{ij}^{\kappa 1\{m=c\}} e^{\alpha_{im} + \alpha_{jm} + d_m(ij)} \quad (5)$$

where κ is driving's commute cost elasticity of road traffic and $d_m(ij)$ is an arbitrary function of distance.

Related literature typically assumes commute costs are an exponential function of travel time such as $d_{ij} = e^{\kappa_d Time_{ij}}$ where κ_d is cost's travel time elasticity (Ahlfeldt et al., 2015; Tsivanidis, 2019; Severen, 2019). Assuming travel time is a linear in log-traffic so that

¹⁵Gonzalez-Navarro and Turner (2018) provide motivating evidence of commuters' nested decision making by finding that public transit expansion increases ridership by causing residents to substitute across modes rather than cities.

$time_{ij} = \kappa_0 + \kappa_T ln(traffic_{ij})$ implies that equation 5 is isomorphic to the standard model where traffic's marginal disutility is $\kappa = \kappa_d \times \kappa_T$. Therefore, abstracting from other issues such as travel time uncertainty implies that κ combines traffic's expected time costs and travel time's disutility.¹⁶

With many modes of transportation, appendix A.2.1 shows that commuters' nested discrete choice problem summarises commute costs between locations ij with cost indices

$$\bar{d}_{ij} = \gamma \left(\sum_m d_{ijm}^{-\eta} \right)^{-\frac{1}{\eta}}. \quad (6)$$

where $\gamma = \Gamma \left(1 - \frac{1}{\eta} \right)^{-1}$.

5.3 Commuting

Since workers choose locations before mode of transportation, skill does not directly affect mode choice and appendix A.2.1 shows that the probability a worker chooses mode m to commute between residence i and workplace j is

$$\pi_{m|ij} = \sum_g \pi_{m|ig} \pi_g = \frac{d_{ijm}^{-\eta}}{\sum_m d_{ijm}^{-\eta}} \sum_g \pi_g = \frac{d_{ijm}^{-\eta}}{\sum_m d_{ijm}^{-\eta}} \quad (7)$$

and the share of type g commuters living in i and working in j is

$$\pi_{ij|g} = \frac{1}{\Phi_g} (w_{jg} B_{ig} Q_i^{\beta-1})^\theta \bar{d}_{ij}^{-\theta} \quad (8)$$

¹⁶Since the data aggregate traffic over complete rush hour periods, accommodating more rush hour traffic on a given road link requires decreasing speeds as feasible headways fall. In the case of a bottleneck, increasing total rush hour traffic increases the amount of time with an active queue and even commuters who respond by altering departure times prefer less road traffic (Arnott et al., 1993).

where $\Phi_g = {}^\theta \sum_{r'} \sum_s (w_{sg} B_{rg} Q_r^{\beta-1})^\theta \bar{d}_{rs}^{-\theta}$. Each commute then has $\Omega \pi_{ij} = \Omega \sum_g \pi_g \pi_{ij|g}$ commuters $\Omega \pi_{ijc} = \Omega \pi_{c|ij} \sum_g \pi_g \pi_{ij|g}$ of whom drive. Appendix A.2.1 also shows that mean worker welfare is the equilibrium expected utility

$$E(u) = \gamma' \gamma \sum_g \pi_g \left[\sum_i \sum_j \left(\frac{B_{ig} w_{jg}}{\bar{d}_{ij} Q_i^{1-\beta}} \right)^\theta \right]^{\frac{1}{\theta}}. \quad (9)$$

where $\gamma' = \Gamma(1 - \frac{1}{\theta})$

5.4 Empirical implications

The following theorem shows that commuting elasticities reflect traffic's effects on sorting, mode choice, and welfare:

Theorem 1 (Traffic's effect on commuting) *Traffic's first-order effect on the number commuting between residence i and workplace j is*

$$\frac{\partial \ln(\pi_{ij})}{\partial \ln(\text{traffic}_{ij})} = -\theta \kappa \pi_{c|ij} (1 - \pi_{ij}) \quad (10)$$

its first-order effect on mode choice is

$$\frac{\partial \ln \left(\frac{\pi_{c|ij}}{1 - \pi_{c|ij}} \right)}{\partial \ln(\text{traffic}_{ij})} = -\eta \kappa \quad (11)$$

and the marginal external cost of congestion on this commute is

$$MCC_{ij} \equiv -\frac{\partial \ln(E(u))}{\partial \ln(\text{traffic}_{ij})} = \kappa \pi_{ijc} \quad (12)$$

where $E(u)$ is mean worker utility, π_{ij} is the probability of living in i and working in j , $\pi_{c|ij}$ is the conditional probability of driving on commute ij , $\pi_{ijc} = \sum_g \pi_{c|ij} \pi_{ij|g} \pi_g$ is the

unconditional choice probability of driving on ij .

Proof: See appendix A.2.1. □

Equation 10 shows that commute flow elasticities reflect location preferences through θ and increase in driving's conditional choice probability $\pi_{c|ij}$ to reflect heterogeneity in road quality and transit infrastructure. The location choice probability $(1 - \pi_{ij})$ enters to reflect the weight of outside options and κ reflects traffic's marginal disutility. For a given value of κ , commuting's traffic elasticity increases in θ to reflect the variance of commuters' idiosyncratic valuations of particular locations. Empirically, if increasing traffic reduces the number commuting, it must be that they choose different residences, find jobs in different workplaces, or both—evidence of sorting.

Equation 11 shows that traffic's effect on the log-odds of driving reflects its marginal disutility through κ and the density of workers near the margin of switching modes through η . If increasing traffic decreases the log-odds of driving between given locations, it must be because public transit becomes relatively more attractive—endogenous mode choice.

In terms of welfare, equation 12 shows that before housing and labour markets adjust, the marginal cost of congestion on a commute depends on the commute cost elasticity κ and the popularity of driving. This partial equilibrium externality holds wages, rents, and households' choices constant and becomes a traditional measure of the value of travel time saved from reducing traffic if κ is the product of traffic's time cost and time's opportunity cost (Small et al., 2007).

As a practical matter, appendix A.2.2 argues that a non-zero commuting elasticity is necessary and sufficient for commute costs to have general equilibrium effects as they determine where people live and work, which then determines wages and rents. This is a particular microfoundation for road links' travel demand elasticities in a canonical model of traffic externalities on a road network (e.g. Small et al. 2007, section 3.4.4). Section

6 proceeds to provide evidence of these forces and section 7 accounts for them in a general equilibrium framework with endogenous labour demand, housing supply, and traffic externalities.

6 Traffic's effect on commuting

This section provides reduced form evidence that marginal changes in traffic affect long-run commuting patterns, discusses implications for studying London's Congestion Charge, and establishes the exclusion restrictions that calibrate commuters' travel demand system for policy analysis. Guided by theorem 1, I argue that traffic's effect on the number of commuters choosing a residence-workplace pair captures sorting across locations while mode shares capture switching between driving and public transit. This section identifies these effects using exogenous variation in traffic caused by exposure to London's Congestion Charge and finds that exogenously decreasing traffic between a residence-workplace pair causes commuters to sort towards those locations and increases the odds that they drive. It follows that by affecting regional traffic patterns, London's congestion charge induced mode choice and location sorting that can have general equilibrium effects as wages and rents adjust.

6.1 Empirical strategy

I estimate the effect of rush-hour traffic on the number commuting and log-odds of driving using regressions of the form

$$\ln(\pi_{ijt}) = \delta^n \ln(\text{traffic}_{ijt}) + X'_{ij} \beta_t + \theta_{b(ij)t} + \alpha_{it} + \alpha_{jt} + \alpha_{ij} + u_{ijt} \quad (13)$$

$$\ln(\pi_{ijtc}) - \ln(\pi_{ijto}) = \delta^c \ln(\text{traffic}_{ijt}) + X'_{ij} \beta_t^c + \theta_{b(ij)t}^c + \alpha_{it}^c + \alpha_{jt}^c + \alpha_{ij}^c + u_{ijt}^c \quad (14)$$

where π_{ijt} is the number of commuters living in i , working in j , in year t , and π_{ijtm} is the subset choosing mode m which takes values c for drivers and o for all other modes. As in section 4, X_{ij} includes direct effects of the CCZ and WEZ as well as measures of public transit availability and road characteristics interacted with a post treatment dummy.¹⁷ Distance-by-year fixed effects $\theta_{b(ij)t}$ control for trends in one kilometre bins of driving distance, α_{it} and α_{jt} capture changes at residences and workplace levels, and α_{ij} forces identification based on changes over time.

The goal is to identify traffic elasticities δ^n and δ^c which are proportional to the number of commuters near the margin of each decision and traffic's marginal cost. Intuitively, increasing a commute's rush-hour traffic increases drivers' travel times and creates uncertainty that burdens commuters targeting early arrival times to avoid being late. In the model presented in section 5 these elasticities calibrate congestion charging's welfare implications and provide evidence of general equilibrium effects as traffic affects sorting, wages, and rents.

My preferred congestion elasticity estimator is two stage least squares (TSLS) adding one to each mode's commute flows before taking logarithms. I weight all regressions by 2001 commuters plus one, base inference on covariance matrices that are two-way clustered by residence and workplace, and appendix A.1.5 presents corresponding maximum likelihood estimates.¹⁸

In practice, I prefer the linear model with one added to flows as the data's small count adjustment makes observations with one and zero commuters indistinguishable. As described in the data appendix, the UK data service protects respondent anonymity by randomly

¹⁷Footnote 14 lists controls.

¹⁸Results using the inverse hyperbolic sine transformation resemble those shown here.

adjusting small counts in commuting matrices to either zero or three. Further, poisson and negative binomial estimators drop clusters of observations in which the dependent variable is always zero, a particularly undesirable property given that I cannot distinguish true zeros from censored observations.¹⁹

6.2 Identification

Before discussing the instrumental variables that identify traffic elasticities, it is useful to note that fixed effects in equation 13 account for most salient omitted variables. Pair fixed effects force identification based on changes over time and control for constant route characteristics such as long-lived infrastructure. Residence-by-year fixed effects control for differential trends in rent or amenities such as pollution, shop openings, and transit stations. Workplace-by-year fixed effects control for trends in workplace productivity or attractiveness including changing parking availability, input access, and floorspace availability. Baseline observables account for the toll itself, traffic measurement issues, and changing effects of distance and public transit. However, least squares elasticity estimates are biased by traffic's correlation with unmeasured determinants of commuting.

Even in panel data, unmeasured changes in commute characteristics will increase both commuting and traffic. In terms of mode choice, new public transit connections can decrease both road traffic and the odds of driving on a route. There could also be workplaces that are complements with certain residences because of parking or transit availability or because both attract a certain type of commuter. These complementarities would increase both commuting and traffic between certain origin-destination pairs, positively biasing estimates of δ^n and δ^c .

¹⁹Dropping zero-flow clusters is necessary because the poisson model would perfectly fit their outcomes by setting the associated fixed effect to $-\infty$. Unfortunately, current techniques for managing sparse commuting data focus on cross-sectional data and do not address this issue (e.g. Dingel and Tintelnot (2020)).

To identify traffic's effect on commuter demand, I instrument traffic with each route's Congestion Charge Exposure Index. CCE is defined to capture latent demand for tolled trips sharing the road with each commuter and section 4 shows that variation is largely driven by geography and idiosyncrasies in London's road network. Since I control for the toll's direct effect, commutes with high CCE see traffic decline after 2003 because of reduced driving on intersecting routes to tolled destinations. Central London's toll discourages drivers from sharing roads with highly exposed commutes, making them faster and more reliable for commuters driving between these locations. In this sense, Congestion Charge Exposure predicts changes in traffic that are exogenous to commuters' location and mode choices conditional on fixed effects and controls.

I estimate parameters of interest using IV regressions with the linear first stage described by equation 4 in section 4 which establishes the relevance condition $E[CCE_{ij} \times traffic_{ij} | X_{ij}, \theta_{b(ij)t}, \alpha_{it}, \alpha_{jt}, \alpha_{ij}] \neq 0$. Specifically, table 2 shows a strong first stage with the expected negative relationship between CCE and traffic growth. To formally examine the exclusion restriction, it is useful to combine equations 2 and 3 to yield

$$E \left[\sum_{h,\ell,r,s} \Delta u_{ij} 1\{\ell \in R_{ij}^h \cap R_{rs}^h\} CCZ_{rs} \frac{traffic_{rs0}^h}{length_{rs}^h} \right] = 0$$

unobservable trends in commuting must be unrelated to the amount of traffic initially on overlapping tolled commutes.²⁰ This condition could be violated if routes with more charge exposure were also more likely to see unobserved infrastructure or transit improvements.

Appendix figure A.7 shows that CCE is related to certain measures of transit connections but not systematically predicted by better or worse public transit in 2001.²¹ To test the

²⁰Rewriting the instrument in terms of a set intersection is valid because $1\{\ell \in R_{ij}^h\} 1\{\ell \in R_{rs}^h\} = 1\{\ell \in R_{ij}^h \cap R_{rs}^h\}$.

²¹Appendix figure A.8 shows that the CCZ is not systematically related to the toll or years underlying pairwise traffic observations conditional on origin, destination, and distance fixed effects. However, there is

importance of rail transit expansion between 2001 and 2011, I identify newly connected pairs as those with both workplace and residence within 1,500 metres of a station in 2011 where at least one of these stations opened after 2001. Appendix table A.6 shows that new connections are positively correlated with CCE but this relationship evaporates when controlling for residence and workplace fixed effects. The same table defines an indicator for pairs with both workplace and residence within 750 metres of stops on the same express or 24 hour bus route to show that CCE's spatial distribution is conditionally uncorrelated with important bus routes. Nevertheless, my preferred specification includes a rich set of controls for baseline public transit connections.

6.3 Main results

Table 3 presents estimates of the number commuting's elasticity with respect to traffic. The first column presents ordinary least squares (OLS) estimates and the remaining columns present TSLS estimates instrumenting traffic with Congestion Charge Exposure, all regressions include MSOA pair fixed effects and are weighted by one plus initial total flows, standard errors and F-statistics are two-way clustered at residence and workplace levels.

Column 1 of table 3 shows that the expected attenuation bias in OLS estimates is severe. Column 2 presents TSLS estimates without controlling for distance or residence and workplace trends, finding a strong negative relationship between changes in commuter counts and traffic. Column 3 adds residence and workplace-by-year fixed effects, attenuating traffic's effect on commuting. Column 4 adds controls which increase the elasticity estimate so that a ten percent decrease in a commute's road traffic increases the number of commuters choosing that location pair by 9.18 percent.²² In terms of observed variation, this amounts

a positive conditional correlation between CCE and the number of road links monitored post-toll.

²²Unreported results suggest that traffic monitoring network controls can account for the difference between columns 3 and 4 of table 3, but independently adding transit controls also increases elasticity estimates.

to a standard deviation increase in log-traffic decreasing log-commuting by one standard deviation.

Table 4 presents similar estimates of traffic’s effect on log-drivers relative to non-drivers, the mode choice regression. Again, OLS estimates display positive bias and the preferred estimates are negative and economically substantial. Comparing columns 2 and 3 shows that controlling for workplace and residence trends turns the relationship between traffic and drivers’ relative commute share from positive to negative.²³ This suggests that routes connecting places becoming more car oriented have low Congestion Charge Exposure. Column 4 controls for route characteristics to find that a ten percent decrease in traffic increases the share of drivers relative to other modes by 12.1 percent.

In terms of observed variation, the mode choice elasticity of table 4 column 4 implies that a standard deviation decrease in log-traffic increases the log-odds of driving by 1.6 standard deviations. Evaluating for an MSOA pair where half of commuters drive associates a 3 point increase in the share of commuters driving with a ten percent decrease in traffic. In context, this is a large negative effect of increasing traffic on the share of commuters choosing to drive.

6.4 Robustness

Appendix A.1.5 presents poisson and negative binomial control function instrumental variables (CFIV) estimates of traffic’s effect on commuting. CFIV regressions drop a number of location pairs without commuting to give smaller and less powerful elasticity estimates than TSLS. Despite their drawbacks, CFIV estimates support the conclusion that traffic affects commuting—a weighted poisson estimate of δ^c implies that a ten percent decrease

²³Unreported results suggest that bias from workplace and residence trends are largely due to transit infrastructure. Adding the full vector of controls without workplace and residence-by-year effects yields $\delta^c = -0.729$ with a two-way clustered standard error of 0.535.

in a commute's traffic increases the number choosing that location pair by 3.2 percent.

Appendix table A.4 presents additional robustness tests. The first three rows show that commuting elasticities are insensitive to dropping MSOAs in the CCZ, WEZ, or both and that dropping the CCZ accentuates effects on mode choice. The final two rows show that elasticities are also unaffected by controlling for a post-toll dummy interacted with rail transit connections opening after 2001 or the interaction of residence and workplace distance from the central business district.

Finally, appendix table A.7 breaks down mode choice effects by separately estimating traffic's effect on the number of commuters driving and using other modes. Results indicate that decreasing a route's traffic significantly increases the number of commuters driving with no effect on the number taking other modes. This is consistent with assuming that road traffic has no direct effect on the cost of commuting by public transit.

6.5 Heterogeneous elasticities

To find heterogeneous effects of traffic on labour market outcomes, I begin by classifying each MSOA as a low-skill workplace if its low skill employment share is above the sample median. Similarly, low-skill residences are those with above median low-skill population share. I then define low-skill MSOA pairs as those where both the workplace and residence are classified as low-skill.

Table 5 presents variants of equations 13 and 14 with log-traffic and instruments interacted with dummies indicating low-skill pairs. All regressions are fit by weighted TSLS, include full sets of controls and a low-skill-post toll dummy, columns 3 and 4 include separate workplace and residence-by-year fixed effects for low-skill and non-low-skill pairs. The two-way clustered Sanderson and Windmeijer (2016) weak IV F-statistics in columns 1 and 2 are smaller than in the homogeneous case but the first stage remains strong enough to

provide reliable estimates.

The first two columns of table 5 suggest that traffic has a larger effect on the location choices of low-skill commuters. Mode choice elasticities are slightly larger between high-skill locations but this difference is not statistically significant. Columns 3 and 4 allow different workplace and residence trends at high- and low-skill MSOA pairs to control for differential wage and amenity growth and give a similar pattern of location effects to columns 1 and 2, but the first stage is weak and results should be interpreted cautiously.

A larger commute elasticity for low-skill workers is a surprising but common finding of commuter gravity estimates (Tsivanidis, 2019; Zárate, 2019; Lee, 2019). While my data cannot pin down the mechanism, I propose that this pattern reflects differences in sorting behaviour and labour market outcomes rather than time preferences. This is consistent with the heterogeneous location fundamentals and homogeneous commute costs in the travel demand system I use for policy analysis going forward.

Appendix tables A.8, A.9, and A.10 investigate additional sources of treatment effect heterogeneity. Table A.8 interacts log-traffic and instruments with distance tercile dummies to find that elasticities are largely independent of trip distance, consistent with the structural model's constant elasticity disutility of traffic. Table A.9 interacts traffic with a dummy indicating commutes within Inner London, finding that traffic reduces the number commuting most within Inner London and mode choice elasticities are only negative and statistically significant if the workplace or residence is in Outer London. Table A.10 follows theorem 1 to provide some evidence in support of the nested logit model by showing that traffic's effect on the number commuting is smallest on commutes with low driving rates in 2001.

7 An equilibrium commuting model

Traffic's significant effect on long-run location choices motivates using a general equilibrium model to quantitatively assess congestion charging's implications for city structure, inequality, and welfare. This section extends the travel demand model in section 5 to account for general equilibrium effects on road traffic, housing, and labour markets. The full model also leverages data describing employment and population of high- and low-skill workers to estimate congestion charging's unequal effects.

7.1 Labour supply, residential population, and wages

Commuters' travel demand, described in section 5, jointly determines labour supply and housing demand in each neighbourhood. Labour supply of skill g to workplace j is $L_{jg} = \Omega \sum_i \pi_{ijg}$ and group g population in neighbourhood i is $N_{ig} = \Omega \sum_j \pi_{ijg}$. As in Ahlfeldt et al. (2015) and Lee (2019), each group's wages satisfy

$$L_{jg} = w_{jg}^\theta \sum_i \frac{\bar{d}_{ij}^{-\theta}}{\sum_s w_{sg}^\theta \bar{d}_{is}^{-\theta}} N_{ig} \quad (15)$$

which defines a unique vector of wages rationalizing location decisions and transportation costs.

7.2 Housing and rents

Rents are determined by residential floorspace demand

$$H_i = \frac{1-\beta}{Q_i} \Omega \sum_g \sum_j \pi_{ijg} w_{jg}.$$

To close the housing market, I assume that an immutable zoning code fixes each MSOA's supply of residential floorspace at H_i . This simplification seems reasonable given London's scarcity of developable land and limited opportunities to convert between residential and commercial uses (Cheshire and Hilber, 2008; Hilber and Vermeulen, 2016).

7.3 Labour demand

Competitive firms use a Cobb-Douglas function of floorspace and a CES aggregate of both skill groups' labour to produce a traded output whose price is normalized to one. I assume commercial floorspace available in each workplace is fixed and simplify notation by normalizing floorspace input and assuming total factor productivity varies across workplaces.

Firms in location j produce $A_j L_j^\alpha$ where α is labour's share of production and total labour input is $L_j = \left(\sum_g a_{jg} L_{jg}^\rho\right)^{\frac{1}{\rho}}$. Firms substitute workers' skills with elasticity $\sigma = \frac{1}{1-\rho}$ and a_{jg} is skill g 's input intensity at workplace j defined so that $\sum_g a_{jg} = 1$. Demand for group g labour in workplace j is then

$$w_{jg} = \alpha a_{jg} A_j L_{jg}^{-\frac{1}{\sigma}} L_j^{\alpha-\rho} \quad (16)$$

and free entry ensures firms spend $F_j = \frac{1-\alpha}{\alpha} \sum_g w_{jg} L_{jg}$ on floorspace. In my empirical application with two skill groups h and l , equilibrium wages and employment reveal each workplace's skill intensity as

$$\frac{a_{jh}}{1 - a_{jh}} = \frac{w_{jh}}{w_{jl}} \left(\frac{L_{jh}}{L_{jl}} \right)^{\frac{1}{\sigma}}. \quad (17)$$

7.4 Road traffic

The route from i to j is an exogenously determined set of road links with endogenous traffic levels. Road link ℓ corresponds to a travel direction along a stretch of road and its

traffic in rush hour h is the sum of commute flows and other traffic so that $\text{traffic}_\ell^h = \Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}^h\} + \mathcal{O}_\ell^h$. The first term, $\Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}^h\}$, is the endogenous number of commuters whose routes include link ℓ and \mathcal{O}_ℓ^h is non-commuter traffic.²⁴

Each link-hour's non-commuter traffic is a reduced form function of Congestion Charge Exposure $\mathcal{O}_\ell^h = \varphi^h CCE_\ell^h + e_\ell^h$ where CCE_ℓ^h is a known function of the city's road tolls. This gives interior solutions to road link traffic

$$\text{traffic}_\ell^h = \Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}^h\} + \varphi^h CCE_\ell^h + e_\ell^h.$$

I impose non-negativity constraints $\text{traffic}_\ell^h \geq 0$ by assuming that non-commute traffic endogenously responds to low demand from commuters such that $\mathcal{O}_\ell^h = 0$ if $\Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}^h\} + \varphi^h CCE_\ell^h + e_\ell^h < 0$. Finally, commuters between i and j face the sum of all traffic on the links they cross going to work in the morning and returning in the evening averaged across these two trips.

7.5 Equilibrium with exogenous productivity

In equilibrium the mass of group g workers living in i and working in j is

$$\Omega \pi_{ijg} = \Omega \frac{\pi_g}{\Phi_g} (w_{jg} B_{ig} Q_i^{\beta-1})^\theta \bar{d}_{ij}^{-\theta} \quad (18)$$

where $\bar{d}_{ij} = \gamma^{-1} \left(\sum_m d_{ijm}^{-\eta} \right)^{-\frac{1}{\eta}}$. Labour market clearing implies that wages satisfy

$$w_{jg} = \alpha a_{jg} A_j \left(\Omega \sum_i \pi_{ijg} \right)^{-\frac{1}{\sigma}} \left(\sum_{g'} a_{jg'} \left(\Omega \sum_i \pi_{ijg'} \right)^\rho \right)^{\frac{\alpha-\rho}{\rho}} \quad (19)$$

²⁴ Appendix A.1.4 discusses an empirical counterpart to the accounting identity describing links' traffic, provides evidence of non-commuter traffic's first-order importance, shows that measurement error precludes identifying each road link's non-commuter traffic, and argues that these errors are random across road links.

floorspace market clearing implies

$$Q_i = \frac{1-\beta}{H_i} \Omega \sum_g \sum_j \pi_{ijg} w_{jg} \quad (20)$$

and

$$F_j = \frac{1-\alpha}{\alpha} \sum_g \sum_i \pi_{ijg} w_{jg}. \quad (21)$$

Given parameters $\{\theta, \sigma, \alpha, \beta\}$, a closed city labour market equilibrium is endogenous allocations $\{\pi_{ijg}\}$, prices $\{w_{jg}, F_j, Q_i\}$, and expected utility (equation 31) determined by fundamentals $\{B_{ig}, A_j, a_{jg}\}$, commute costs indices $\{\bar{d}_{ij}\}$, aggregate skill shares $\{\pi_g\}$, and population Ω so that equations 18, 19, 20, and 21 hold.

The labour market determines location choices $\pi_{ij} = \sum_g \pi_{ijg}$ and commute costs are

$$\begin{aligned} \bar{d}_{ij} &= \gamma^{-1} \left(d_{ijc}^{-\eta} + d_{ijo}^{-\eta} \right)^{-\frac{1}{\eta}} \\ \text{s.t. } d_{ijc} &= toll_{ij} \cdot traffic_{ij}^\kappa \cdot e^{\alpha_{ic} + \alpha_{jc} + d_c(ij)} \end{aligned} \quad (22)$$

which depend on endogenous allocations through road traffic

$$\begin{aligned} traffic_{ij} &= \frac{1}{2} \sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} traffic_\ell^h \\ \text{s.t. } traffic_\ell^h &= \max \left(\varphi^h CCE_\ell^h + e_\ell^h + \Omega \sum_{r,s} 1\{\ell \in R_{rs}^h\} \pi_{c|rs} \pi_{rs}, 0 \right) \end{aligned} \quad (23)$$

where $h \in \{morn, eve\}$. Combining with equation 7 gives

$$\begin{aligned} \ln \left(\frac{\pi_{c|ij}}{1 - \pi_{c|ij}} \right) = & \eta \alpha_{ic} + \eta \alpha_{jc} + \eta d_c(ij) - \eta \ln(toll_{ij}) + \eta \ln(d_{ijo}) \\ & - \eta \kappa \ln \left[\frac{1}{2} \sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} \max \left(\varphi^h CCE_\ell^h + e_\ell^h + \Omega \sum_{r,s} 1\{\ell \in R_{rs}^h\} \pi_{c|rs} \pi_{rs}, 0 \right) \right] \end{aligned} \quad (24)$$

which implicitly defines mode shares $\{\pi_{c|ij}\}$ as a function of location choices.

Given parameters $\{\theta, \eta, \kappa, \varphi^{morn}, \varphi^{eve}, \sigma, \alpha, \beta\}$, a closed city general equilibrium is a labour market equilibrium plus traffic and mode shares $\{traffic_\ell^h, \pi_{c|ij}\}$ so that equations 18, 19, 20, 21, 22, 23, and 24 hold given routes, labour market fundamentals, transit costs, mode share fixed effects, cost residuals, charge exposure, and tolls $\{R_{rs}^h, B_{ig}, A_j, a_{jg}, \alpha_{ic}, \alpha_{jc}, d_{ijo}, e_\ell^h, CCE_\ell^h, toll_{ij}\}$. Theorem 2, stated and proved in appendix A.2.3, guarantees the existence of a closed city general equilibrium with exogenous productivity.

7.6 Endogenous productivity

I introduce symmetric agglomeration externalities that decay across neighbourhoods as in Ahlfeldt et al. (2015), allowing total factor productivity to increase in nearby labour inputs such that

$$A_j = a_j \Upsilon_j^\lambda \text{ s.t. } \Upsilon_j = \sum_s e^{-\delta d(s,j)} \frac{L_s}{Area_s}$$

where $Area_s$ is the land area of neighbourhood s and $d(s,j)$ is the euclidean distance between centroids of s and j set to $\sqrt{\frac{Area_j}{3.14}}$ when $s = j$. The parameter $\delta > 0$ governs spillovers' spatial decay and λ is productivity's elasticity of nearby CES employment indices. This externality adapts Ahlfeldt et al.'s (2015) single-skill model to allow for cross-

neighbourhood agglomeration externalities from firms' labour inputs.²⁵

8 Model estimation and calibration

For counterfactual policy analysis, I require estimates of parameters $\{\theta, \eta, \kappa, \varphi^{morn}, \varphi^{eve}, \sigma, \beta\}$, commute cost indices $\{\bar{d}_{ij}\}$, amenities $\{B_{ig}\}$, wages $\{w_{jg}\}$, skill intensities $\{a_{jg}\}$, regional skill shares $\{\pi_g\}$, and traffic residuals $\{e_\ell^h\}$. Incorporating endogenous productivity also requires taking a stand on agglomeration's decay δ and elasticity λ . This subset of model parameters and fundamentals is sufficient for policy analysis since I solve the model for changes in outcomes rather than levels.

I use traffic and commuting data to sequentially estimate the model's central parameters based on reduced form estimates presented in section 6. I then calibrate standard labour and housing demand parameters to existing literature and solve for heterogeneous wages and amenities rationalizing high- and low-skill population and employment distributions across MSOAs.

This sequential estimation procedure has the benefit of identifying central travel demand parameters $\{\theta, \eta, \kappa\}$ before taking a stand on labour demand and housing supply. The orthogonality conditions established in section 6.2 identify parameters that are portable to any gravity commuting model that delivers the first-order elasticities derived in theorem 1. Additional structure on labour demand and traffic's sources delivers unique amenities, wages, skill intensities, commute costs, and traffic residuals rationalizing observed traffic, population, employment, and commuting as an equilibrium of the model given parameters and Congestion Charge Exposure.

²⁵The theory appendix discusses localized human capital spillovers as in Lee (2019).

8.1 Sequential parameter estimation

Aggregating across groups' location choices, taking logarithms, and substituting the identity $\bar{d}_{ij} = \gamma^{-1}(1 - \pi_{c|ij})^{\frac{1}{\eta}} d_{ijo}$ gives the structural gravity equation

$$\ln(\pi_{ij}) = \ln \left(\sum_g \frac{\pi_g}{\Phi_g} (w_{jg} B_{ig} Q_i^{\beta-1})^\theta \right) - \frac{\theta}{\eta} \ln(1 - \pi_{c|ij}) + \ln \left(\frac{\gamma^\theta}{d_{ijo}^\theta} \right).$$

The first term summarises amenities, rents, and wages that I capture with time-varying residence and workplace fixed effects. Adding time subscripts and decomposing $\ln \left(\frac{\gamma^\theta}{d_{ijo}^\theta} \right)$ into controls, fixed unobservables, and a residual gives the estimating equation

$$\ln(\pi_{ijt}) = \omega_{jt} + b_{it} - \frac{\theta}{\eta} \ln(1 - \pi_{c|ijt}) + X'_{ij} \delta_t + \delta_{ij} + u_{ijt}. \quad (25)$$

Assuming $E[CCE_{ij} \times Post_t \times u_{ijt}] = 0$, I use a weighted TSLS regression to identify the parameter ratio $\frac{\theta}{\eta}$.²⁶

Table 6 presents TSLS estimates of equation 25 weighted by 2001 commuting and adding $\frac{1}{N \times (N-1)}$ to choice probabilities before taking logarithms, standard errors in parenthesis are clustered by both workplace and residence. Both regressions include workplace-, residence-, pair-, and distance band-by-year fixed effects and column 2 adds the controls described in section 6.1. Results are insensitive to controls and I set $\frac{\theta}{\eta} = 2.229$ to match table 6's second column.

²⁶Linearly decomposing the structural unobservable gives $\ln \left(\sum_g \frac{\pi_g}{\Phi_g} (w_{jg} B_{ig} Q_i^{\beta-1})^\theta \right) = \omega_{jt} + b_{it} + \ln \left(\sum_g \frac{\pi_g}{\Phi_g} \frac{B_{ig}^\theta}{\sum_{g'} \pi_{g'} B_{ig'}^\theta} \frac{w_{jg}^\theta}{\sum_{g'} \pi_{g'} w_{jg'}^\theta} \right)$ where the final term is included in the residual of equation 25. This omitted variable will only bias parameter estimates if Congestion Charge Exposure is systematically correlated with residence-workplace pairs that offer high amenities and wages to one group relative to others.

I then use equation 7 to identify $\eta\kappa$ from the TSLS regression

$$\ln(\pi_{c|ijt}) - \ln(1 - \pi_{c|ijt}) = \alpha_{it}^c + \alpha_{jt}^c - \eta\kappa \ln(\text{traffic}_{ijt}) + X'_{ij}\beta_t^c + \theta_{b(ij)t}^c + \alpha_{ij}^c + u_{ijt}^c \quad (26)$$

under the assumption $E[CCE_{ij} \times Post_t \times u_{ijt}^c] = 0$. Section 6.3 presents results and I set $\eta\kappa = 1.208$ to match column 4 of table 4.

Assuming commute costs are an exponential function of travel time, that travel time's marginal disutility per minute is 0.01 (Ahlfeldt et al., 2015; Tsivanidis, 2019; Severen, 2019), and that log-traffic has a constant effect on travel time given by κ_T implies that traffic's utility elasticity is $\kappa = 0.01 \times \kappa_T$. Herzog (2020) shows that a log-point increase in traffic increases travel times by 0.16 seconds per metre and converting seconds to minutes and evaluating for a 14 kilometre trip delivers $\kappa_T = 37.33$ —a ten percent increase in traffic volume increases a representative trip's travel time by 3.7 minutes.²⁷ Multiplying by 0.01 gives traffic's commute cost elasticity $\kappa = 0.3733$ —absorbing ten percent more traffic reduces utility by 3.7 percent—which implies $\eta = 3.24$ and $\theta = 7.21$.

8.2 Wages and adjusted amenities

I calibrate each group's wages and amenities to rationalize their spatial distributions of employment and population as in Ahlfeldt et al. (2015) and Lee (2019). I first identify commute cost indices (up to scale) by adjusting 2011 commute flows for traffic using elasticity estimates from section 6.3 and regressing them on controls, distance, residence, and workplace fixed effects $\ln(\pi_{ij2011}) + 0.918\ln(\text{traffic}_{ij2011}) = \theta_{b(ij)} + \alpha_i + \alpha_j + e_{ij}$. I then recover commute cost indices as the structural residual $\bar{d}_{ij} = \exp(\ln(\pi_{ij2011}) - \alpha_i - \alpha_j)$ which includes the non-linear distance function captured by distance band fixed effects.

²⁷Herzog (2020) also demonstrates that a linear-in-logs functional form is a reasonable fit for London's link-level travel supply function.

Given θ , \bar{d}_{ij} , and data on high- and low-skill employment and population in 2011, I solve the system of equations described by equation 15 for high- and low-skill wages that rationalize locations and commute costs. Given wages, equilibrium populations reveal adjusted amenities $\tilde{B}_{ig} = \Phi_g^{-1}(B_{ig}Q_i^{\beta-1})^\theta$. I identify this cluster by aggregating across each group's gravity equation and re-arranging to compute $\tilde{B}_{ig} = \frac{\pi_{ig}}{\sum_j w_{jg}^\theta \bar{d}_{ij}^\theta}$ up to scale.²⁸

8.3 Commuting by skill

I use the model to separately estimate commute flows of high- and low-skill workers. Taking the ratio of each group's location choices (equation 8) gives

$$\frac{\pi_{ij|h}}{\pi_{ij|l}} = \frac{w_{jh}^\theta \tilde{B}_{ih}}{w_{jl}^\theta \tilde{B}_{il}}. \quad (27)$$

Given wages and adjusted amenities, I solve for skill specific commute flows that satisfy equation 27 and add up to observed 2011 commuting $\pi_{ij} = \pi_h \pi_{ij|h} + \pi_l \pi_{ij|l}$ for each MSOA pair.²⁹

8.4 Labour and housing demand parameters

I set housing and labour demand parameters according to existing estimates. I set housing's expenditure share $(1 - \beta) = .4$ (as in Severen (2019)) a relatively large value that is reasonable in a city as expensive as London. On the labour demand side, I set firms' floorspace share to $(1 - \alpha) = .2$ and the elasticity of substitution between high- and low-

²⁸I scale wages to have geometric means of £120 and £160 for low- and high-skill respectively. After identifying skill-specific commute flows, an expected wage of £170 rationalizes 2011 location choices as a labour market equilibrium.

²⁹Groups' population and employment are imbalanced by those commuting into and out of the region. In 2011 the ratio of high-skill commuter population to employment in the region is 1.03 and this ratio is 2.12 for low-skill workers. Therefore, regional skill shares π_g are the average of each group's total commuter population and employment in 2011 and normalized to sum to one. This procedure sets $\pi_h = 0.755$, London's local labour market is approximately 76 percent high-skill workers.

skill workers to $\sigma = 1.3$ following Card (2009), Tsivanidis (2019), and Lee (2019). Given wages and the elasticity of substitution, I evaluate equation 17 to identify skill intensities that rationalize high- and low-skill workers' relative employment and wages in each MSOA.

I set the agglomeration elasticity $\lambda = 0.7$ and decay $\delta = 4.2$ to match estimates from Ahlfeldt et al. (2015) given a 5 km-per-hour walking speed. These parameters are consistent with the rapidly decaying agglomeration forces found in similar contexts (Arzaghi and Henderson, 2008; Ahlfeldt et al., 2015; Lee, 2019).

8.5 Traffic residuals

I set traffic residuals e_ℓ^h that rationalize the distributions of traffic and CCE across road links in each rush hour. First, I run panel regressions of link-level morning and evening traffic on Congestion Charge Exposure to identify φ^h in each rush hour from

$$traffic_{\ell t}^h = \varphi^h CCE_{\ell t}^h \times Post_t + X_i' \Gamma_t^h + \gamma_t^h Post_t + \gamma_\ell^h + e_{\ell t}^h$$

where $h \in \{morn, eve\}$ denotes morning and evening rush hour traffic, γ_ℓ^h are road-link fixed effects, and X_i are baseline controls indicating links within the CCZ, WEZ, year of measurement, and log-distance from the boundary interacted with a CCZ dummy. I extract each hour's marginal effect φ^h and set $e_\ell^h = traffic_{\ell 2011}^h - \varphi^h CCE_{\ell 2011}^h - \Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}\}$ which is non-commuter traffic without the toll.³⁰ Appendix table A.3 presents detailed results which suggest that a standard deviation increase in link-level charge exposure decreases daily rush hour traffic volume by 66 vehicles on average and that this effect is larger in the evening rush hour.

³⁰CCE has a standard deviation of one and a minimum value of zero in each rush hour.

8.6 Model fit

The model predicts high- and low-skill commuters using cars and public transit at the same rate holding all else equal. Interestingly, the data suggest that this assumption is a reasonable approximation in Greater London. In particular, section 6.5 finds little evidence of heterogeneous mode choice elasticities of traffic and workplace data suggest roughly equal car shares of low- and high-skill commuters. In 2011, 37 percent of low-skill workers drive to work and 35 percent of high-skill workers drive.

The top left panel of figure 8 assesses the model's fit by plotting observed 2011 driver shares against model predictions at workplaces and residences. Colours delineate skill groups and residential or workplace counts, large points are binned means, and the red dashed 45 degree line indicates a perfect fit. Figure 8 shows that the model generally fits mode shares well except for moderately over-predicting low-skill driver shares.

The model's ability to predict each group's spatial distribution of population and employment provides an additional goodness of fit test. To this end, the top-right panel of figure 8 plots the logarithm of MSOA employment and population counts against their model predicted values for high- and low-skill commuters. The model has some trouble capturing residential patterns but accurately predicts both groups' employment.

The bottom panels of figure 8 assess the model's backcasting ability by calibrating to 2001 data and plotting actual against predicted changes in choice shares. Results indicate that the model fits changes in mode shares and location choices well with the exception of a few outliers in employment and population growth.

9 Counterfactuals

I now simulate removing Central London's toll to estimate counterfactual changes in commuters' locations, driving rates, road traffic, and welfare. Simulations hold exogenous variables at their 2011 values and remove an £8 toll assuming revenues were redistributed lump sum into wages. Counterfactuals shock commute costs by removing the toll's monetary cost from commutes into the CCZ and setting each road link's Congestion Charge Exposure to its normalized minimum of zero. This counterfactual accounts for revenue collected from commuters with homogeneous rebates but omits non-commuter revenues; implicitly assuming that only commuters can benefit from the policy and that uncounted revenue offsets adjustment costs of tolled non-commute drivers. The effect of removing the congestion charge follows CCE's spatial distribution, shown in figure 6, to create an initial shock to non-commuter road traffic.³¹

Given the initial shock to commute costs from reducing tolls and increasing non-commuter traffic, I solve for relative prices and allocations using an iterative hat algebra procedure following Dekle et al. (2007). Since I fix Greater London's population, exogenously reducing road traffic gives commuters welfare gains to the extent that real wages rise and commute costs fall. Percent increase in welfare caused by Central London's Congestion Charge is the change in mean expected utility

$$100 \times \left(\frac{E(u) - E(u)'}{E(u)'} \right) = 100 \times \left(\left[\frac{\sum_g \pi_g E(u|g) \widehat{E}(u|g)}{\sum_g \pi_g E(u|g)} \right]^{-1} - 1 \right)$$

where $\widehat{E}(u|g) = \frac{E(u|g)'}{E(u|g)}$ is group g 's counterfactual mean utility $E(u|g)'$ relative to its

³¹The initial traffic shock sets each link-hour's traffic equal to its residual plus commuter traffic $e_\ell^h + \Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}^h\}$ and I account for cross-commute spillovers as the computational algorithm converges to a new equilibrium. Each commuter also loses the transfer of £0.14 from their daily wage as part of the initial shock. Appendix A.2.5 continues discussing this counterfactual's quantitative assumptions.

observed value with the CCZ in place. Appendix A.2.4 defines a hat algebra equilibrium and describes a computational procedure for arriving at one.

9.1 Welfare, mode choice, and road traffic

Table 7 presents percent change in equilibrium utility, inequality, segregation, traffic, and driving's share of commutes caused by London's Congestion Charge. Inequality is the ratio of high- to low-skill expected utility and I measure skill groups' segregation using dissimilarity indices, the sum of absolute cross-group differences in MSOA employment or population shares. Panel A assumes an £8 toll with lump-sum revenue redistribution and panel B isolates the effects of Congestion Charge Exposure by assuming commuters do not pay tolls in either equilibrium.³² Columns sequentially add model mechanisms that section 9.3 describes in detail and the final column presents results using the full model with agglomeration externalities.

The final column of panel A of table 7 shows that London's Congestion Charge increased expected utility by 0.025 percent. This effect is small in absolute terms but of the same order of magnitude as recent estimates of gross welfare gains from public transit expansion (Severen, 2019).³³ Multiplying the mean welfare gain by 2.9 million commuters in the 2011 flow data, a £170 mean daily wage, and 260 work days per year gives an annual benefit of approximately £32 million.

In terms of inequality, the second and third rows of table 7's final column show that the congestion charge favours low-skill commuters and decreases utility inequality by 0.07 percent. The fourth and fifth rows show that reduced inequality comes alongside convergence

³²Simulations with toll exemptions are equivalent to a counterfactual where tolls persist and traffic exogenously falls according to road links' Congestion Charge Exposure.

³³Severen (2019) estimates that Los Angeles Metro Rail increases welfare between 0.051 to 0.10 percent (depending on the time horizon and time savings for drivers) and that this benefit fails to exceed annual operating subsidies. Estimated benefits of public transit improvement in developing world cities typically exceed those from richer cities (Tsivanidis, 2019; Zárate, 2019).

in high- and low-skill location choices.

Additional simulations suggest that progressiveness follows from low-skill workers' relatively loose ties to tolled workplaces. Specifically, panel B of table 7 runs counterfactual simulations that isolate the policy's indirect effects by exempting commuters from tolls and shocking commute costs with non-commuter traffic reductions. The first column shows that traffic reductions give similar benefits to high- and low-skill commuters before they sort. Moving to the final column of panel B shows that sorting disproportionately benefits high-skill commuters in a simulation with traffic reductions but no tolls, suggesting that the policy is progressive because high-skill employment relies on the tolled zone's exogenous labour productivity.

The second last row of panel A of table 7 shows that London's Congestion Charge increases the driving rate of the region's commuters by 0.581 percent. Comparing columns 2 and 3 shows that endogenous traffic attenuates this mode shift as commuters endogenously sort and crowd each-other onto public transit.

The final row of table 7 panel A shows that London's Congestion Charge reduces total traffic by 0.715 percent. Comparing to the initial traffic shock (columns 1 and 2) shows that traffic would have fallen by an additional 0.475 percentage points if commuters did not endogenously sort and substitute from public transit to driving. Further, comparing with panel B shows that 43 percent of the remaining endogenous traffic effect is caused by the cordon fee itself.³⁴

Figure 10 maps simulated percent change in road link daily traffic caused by London's Congestion Charge, red hues indicate declines, green indicates growth, and line width is proportional to difference from zero. The policy reduces traffic within the CCZ and on

³⁴Comparing traffic reductions in the final rows of Panels A and B of table 7 suggests that the monetary incentive to avoid tolled commutes increases the equilibrium traffic reduction by $0.715 - 0.411 = 0.304$ percentage points, 43 percent of the equilibrium traffic decline.

radial roads headed downtown but increases traffic on many orbital roads. Interestingly, the policy also reduces traffic on London’s tolled Inner Ring Road despite its use by vehicles bypassing the CCZ—the Inner Ring Road’s traffic falls because it is an important part of many tolled trips.

9.2 Economic geography

Figure 9 maps percent change in each skill group’s MSOA employment and population caused by London’s Congestion Charge, red hues indicate declines and green indicates growth. The left panel shows that the policy decreases both groups’ employment in the CCZ and increases employment in Inner London’s CCZ-adjacent MSOAs. Interestingly, low-skill employment also falls throughout the suburbs and concentrates in a ring around Central London. Effects on high-skill suburban employment are less prominent, suggesting their workplaces are less geographically flexible than low-skill workplaces.

Figure 9’s right panel shows that London’s Congestion Charge causes population growth throughout the suburbs. In fact, it appears the policy reallocates high- and low-skill population to different suburban neighbourhoods—low-skill population grows more in northern and eastern suburbs while high-skill population growth favours the west and eastern neighbourhoods along the Thames. However, it is important to note that table 7 indicates that the policy decreased residential dissimilarity in total.

Finally, appendix figure A.9 maps simulated percent change in residential and commercial floorspace rents. Commercial floorspace rents follow employment growth’s spatial distribution and residential rent growth’s spatial distribution is correlated with changing high-skill population. Simulated residential rents also grow in the CCZ and Canary Wharf relative to nearby MSOAs. This suggests that the changing distribution of high-skill labour market access is one potential mechanism for the CCZ’s effect on home values found by

Tang (2018).

9.3 Mechanisms

In the model, three mechanisms mediate transportation policy's welfare effects: commuters' location and mode choices, endogenous wages and rents, and endogenous road traffic caused by commuters who drive. Table 7 unpacks these mechanisms by presenting the congestion charge's effect on aggregate outcomes in four simplified models.

Column 1 of table 7 eliminates all three margins of adjustment and aggregates equation 12 of theorem 1 across commuters to compute the initial policy shock's aggregate effect.³⁵ This partial equilibrium effect is equivalent to the traditional value of time savings approach used by Hall (2020), Yang et al. (2020), Herzog (2020), and others. Column 2 allows locations, wages, and rents to adjust, shutting down all externalities and exogenously fixing traffic after the initial shock. Column 3 endogenizes road traffic to commuters' decisions but omits cross-neighbourhood productivity externalities. Panel B repeats simulations without commuters' monetary incentives to avoid driving into the CCZ.

Column 1 of table 7 shows that a partial equilibrium value of time savings calculation substantially overstates the congestion charge's aggregate welfare effect. Repeating the calculation from section 9.1 using this partial equilibrium measure would associate a £200 million annual welfare gain with the policy, far larger than the £32 million equilibrium benefit. Column 2 of panel A shows that allowing commuters to sort across locations and modes of transportation attenuates their gains relative to the partial equilibrium measure. Moving to column 3 shows that adding endogenous road traffic further reduces gains and attenuates the policy's effect on equilibrium driving rates.

³⁵Column 1 of table 7 presents the first-order effects $-dlnE(u|g) = -\sum_{i,j} \pi_{ij|g} (dln(w_{jg}) - dln(toll_{ij}) - \kappa dln(traffic_{ij}))$ where $dln(w_{jg}) = -\frac{8}{w_{jg}} \sum_{ij} \pi_{ijcg} CCZ_{ij}$ accounts for recycled toll revenue.

Interestingly, comparing columns 3 and 4 shows that agglomeration externalities increase the CCZ's benefits to commuters. This occurs despite endogenous employment decentralization because London's Congestion Charge also decreases suburban employment and causes jobs to coalesce in Inner London's CCZ adjacent area.

Finally, I find that location and mode choices combine to attenuate the congestion charge's equilibrium traffic reduction by 0.475 percentage points relative to a model with no externalities. Table 8 separates these mechanisms by comparing initial and equilibrium traffic declines (columns 1 and 4) to two intermediate cases. The second column isolates the effect of location choices using equilibrium locations and initial mode shares to compute a 0.842 percent traffic decline. The third column isolates mode choices using initial locations and equilibrium mode choice to compute a 1.012 percent traffic decline. These calculations show that induced roadspace demand reflects a combination of shifts towards higher driving rates on each commute and increases popularity of long commutes.

For an unambiguous accounting, it is useful to define endogenous traffic caused by the policy to allow the decomposition

$$\begin{aligned} \text{Endog. Traffic} &= \sum_{i,j,\ell,h} (\pi_{c|ij}\pi_{ij} - \pi'_{c|ij}\pi'_{ij})1\{\ell \in R_{ij}^h\} \\ &= \underbrace{\sum_{i,j,\ell,h} (\pi_{c|ij} - \pi'_{c|ij})\pi_{ij}1\{\ell \in R_{ij}^h\}}_{\text{Mode shares} = 38\%} + \underbrace{\sum_{i,j,\ell,h} (\pi_{ij} - \pi'_{ij})\pi'_{c|ij}1\{\ell \in R_{ij}^h\}}_{\text{Location choices} = 62\%} \end{aligned}$$

where $\pi'_{(.)}$ is a counterfactual choice share without the congestion charge and $\pi_{(.)}$ is its counterpart observed with the policy in place. This calculation reveals that 38 percent of roadspace demand induced by London's Congestion Charge comes from mode choices and the other 62 percent comes from location sorting—location choices contribute slightly more than mode choices to endogenous traffic.

10 Conclusion

This paper studies the effects of Central London’s Congestion Charge on the region’s large and diverse labour market. To this end, I develop a method to measure effects of a centralized cordon fee on traffic patterns throughout the city and use a new equilibrium framework for policy analysis. In the process, I present reduced-form empirics establishing that Central London’s Congestion Charge reduced traffic on roads leading downtown and use this variation to estimate elasticities of commuters’ travel demand. These elasticities then calibrate a general equilibrium model with heterogeneous worker skills, productivity externalities, mode choice, and endogenous traffic externalities to estimate the aggregate effects of London’s Congestion Charge.

My results show that tolling drivers to enter a regional employment centre generates widespread traffic reductions that benefit commuters in the long-run. Further, regional traffic reductions disproportionately benefit low-skill commuters who are less tied to tolled workplaces. I also find that exogenously reducing traffic between neighbourhoods attracts additional commuters, increases their odds of driving over taking transit, and that ignoring this endogenous sorting would dramatically overstate both traffic declines and welfare gains caused by London’s Congestion Charge.

All things considered, I find that London’s Congestion Charge gives commuters £32 million in annual benefits. This is approximately ten percent of 2011 toll revenue (Transportation for London, 2012) and just seven percent of the remaining £477 million dead-weight loss of congestion (Herzog, 2020) reflecting the Congestion Charge Zone’s limited extent. Despite the policy’s limitations, these are progressive gains that required only modest capital investment, bring government revenue, and improve downtown amenities (Tang, 2018). With this in mind, similar, or even more ambitious, cordon fees are an attractive option for other centralized cities looking to alleviate congestion’s costs.

These conclusions leave several important avenues for future work. In particular, this paper estimates commuters' welfare gains without evaluating the net benefits for non-commute travel, which are an important piece of a comprehensive cost-benefit analysis. I am also unable to evaluate trip timing decisions that may be particularly important for understanding non-commuters' behaviour. Finally, my empirical setting precludes analysis of more complex policies with tolls that vary smoothly over time and space and may not generalize to developing countries or cities with mediocre public transit. Continued research in these directions can improve our ability to guide policy makers and refine our understanding congestion charging's costs and benefits.

References

- Adam, Z., Walasek, L., and Meyer, C. (2018). Workforce commuting and subjective well-being. *Travel behaviour and society*, 13:183–196.
- Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the Berlin wall. *Econometrica*, 83(6):2127–2189.
- Akbar, P. and Duranton, G. (2017). Measuring the cost of congestion in highly congested city: Bogotá.
- Allen, T. and Arkolakis, C. (2019). The welfare effects of transportation infrastructure improvements. Technical report, National Bureau of Economic Research.
- Allen, T., Arkolakis, C., and Li, X. (2015). On the existence and uniqueness of trade equilibria. *Manuscript, Yale Univ.*
- Arnott, R., De Palma, A., and Lindsey, R. (1993). A structural model of peak-period

congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, pages 161–179.

Arzaghi, M. and Henderson, J. V. (2008). Networking off madison avenue. *The Review of Economic Studies*, 75(4):1011–1038.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Baum-Snow, N., Kahn, M. E., and Voith, R. (2005). Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000 [with comment]. *Brookings-Wharton papers on urban affairs*, pages 147–206.

Berge, L. et al. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. Technical report, Center for Research in Economic Analysis, University of Luxembourg.

Brinkman, J. C. (2016). Congestion, agglomeration, and the structure of cities. *Journal of Urban Economics*, 94:13–31.

Card, D. (2009). Immigration and inequality. *American Economic Review*, 99(2):1–21.

Cheshire, P. C. and Hilber, C. A. (2008). Office space supply restrictions in Britain: the political economy of market revenge. *The Economic Journal*, 118(529):F185–F221.

Christian, T. J. (2012). Trade-offs between commuting time and health-related activities. *Journal of urban health*, 89(5):746–757.

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

- Dekle, R., Eaton, J., and Kortum, S. (2007). Unbalanced trade. *American Economic Review*, 97(2):351–355.
- Dingel, J. I. and Tintelnot, F. (2020). Spatial economics for granular settings.
- Duranton, G. and Turner, M. A. (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, 101(6):2616–52.
- Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.
- Fajgelbaum, P. D. and Schaal, E. (2017). Optimal transport networks in spatial equilibrium. Technical report, National Bureau of Economic Research.
- 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.
- Gibson, M. and Carnovale, M. (2015). The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics*, 89:62–73.
- Gonzalez-Navarro, M. and Turner, M. A. (2018). Subways and urban growth: Evidence from Earth. *Journal of Urban Economics*, 108:85–106.
- Google (2020). Google API Terms of Service | Google Developers. <https://developers.google.com/terms>.
- Greater London Authority (1999). Greater London authority act 1999.
- Green, C., Heywood, J. S., and Navarro Paniagua, M. (2018). Did the London congestion charge reduce pollution?
- Green, C. P., Heywood, J. S., and Navarro, M. (2016). Traffic accidents and the London congestion charge. *Journal of Public Economics*, 133:11–22.

- Hall, J. D. (2020). Can tolling help everyone? estimating the aggregate and distributional consequences of congestion pricing. *Journal of the European Economic Association*.
- Hanna, R., Kreindler, G., and Olken, B. A. (2017). Citywide effects of high-occupancy vehicle restrictions: Evidence from “three-in-one” in Jakarta. *Science*, 357(6346):89–93.
- Heblich, S., Redding, S. J., and Sturm, D. M. (2020). The making of the modern metropolis: evidence from London.
- Herzog, I. (2020). The marginal external cost of traffic in Greater London.
- Hilber, C. A. and Vermeulen, W. (2016). The impact of supply constraints on house prices in England. *The Economic Journal*, 126(591):358–405.
- Hilbrecht, M., Smale, B., and Mock, S. E. (2014). Highway to health? commute time and well-being among canadian adults. *World Leisure Journal*, 56(2):151–163.
- Hsu, W.-T. and Zhang, H. (2014). The fundamental law of highway congestion revisited: Evidence from national expressways in japan. *Journal of Urban Economics*, 81:65–76.
- Jerch, R., Barwick, P., Li, S., Wu, J., et al. (2020). Road rationing policies and housing markets. Technical report, Department of Economics, Temple University.
- Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic perspectives*, 20(1):3–24.
- Kreindler, G. E. (2018). The welfare effect of road congestion pricing: Experimental evidence and equilibrium implications. *Unpublished paper*.
- Kunn-Nelen, A. (2016). Does commuting affect health? *Health economics*, 25(8):984–1004.

- Leape, J. (2006). The London congestion charge. *Journal of Economic Perspectives*, 20(4):157–176.
- Lee, K. (2019). Benefits of working near college workers: Human capital spillovers across neighborhoods.
- 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*, GIS '11, pages 513–516, New York, NY, USA. ACM.
- Martin, L. A. and Thornton, S. (2017). Can road charges alleviate congestion? *Available at SSRN 3055428*.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328.
- Molnar, A. and Mangrum, D. (2018). The marginal congestion of a taxi in new york city. Technical report, Working paper.
- Monte, F., Redding, S. J., and Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, 108(12):3855–90.
- Office for National Statistics (2012). Annual survey of hours and earnings: 2011. Technical report.
- Office of National Statistics (2014). Commuting and personal well-being, 2014. Technical report.
- Parry, I. W. (2009). Pricing urban congestion. *Annu. Rev. Resour. Econ.*, 1(1):461–484.
- Proost, S. and Thisse, J.-F. (2019). What can be learned from spatial economics? *Journal of Economic Literature*, 57(3):575–643.

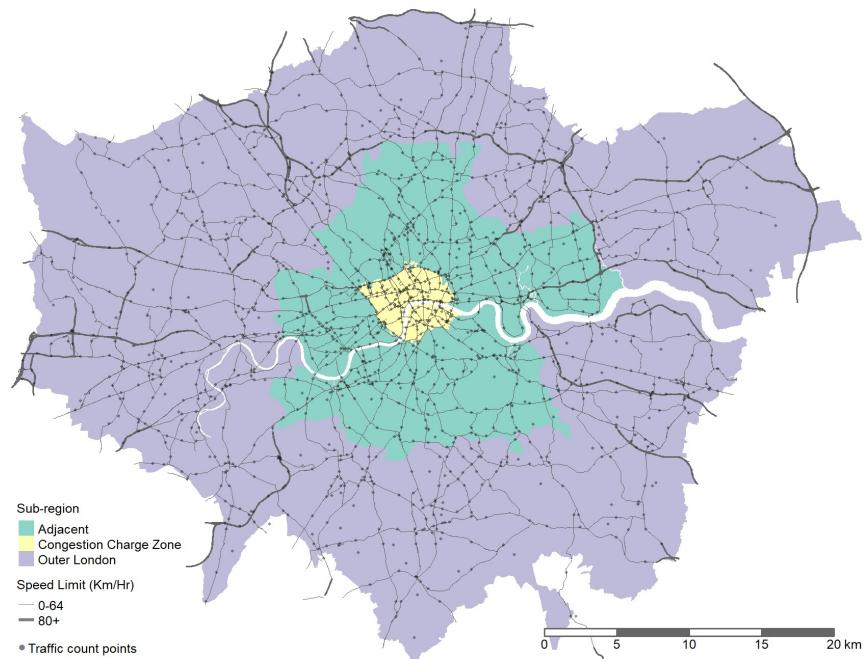
- Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics*, 39(3):347–366.
- Russo, A., Adler, M., Liberini, F., and van Ommeren, J. N. (2019). Welfare losses of road congestion.
- Sanderson, E. and Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.
- Santos, G., Button, K., and Noll, R. G. (2008). London congestion charging. *Brookings-Wharton Papers on Urban Affairs*, pages 177–234.
- Severen, C. (2019). Commuting, labor, and housing market effects of mass transportation: Welfare and identification.
- Simeonova, E., Currie, J., Nilsson, P., and Walker, R. (2018). Congestion pricing, air pollution and children’s health. Technical report, National Bureau of Economic Research.
- Small, K. A., Verhoef, E. T., and Lindsey, R. (2007). *The economics of urban transportation*. Routledge.
- Smeed, R. (1964). Road pricing: the economic and technical possibilities. *The Smeed Report, Ministry of Transport. London*.
- Stillwell, J. and Duke-Williams, O. (2007). Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991. *Journal of the Royal Statistical Society: Series A (statistics in society)*, 170(2):425–445.
- Tang, C. K. T. (2018). The cost of traffic: Evidence from the London congestion charge. Technical report.

- Transportation for London (2003). Congestion charging: Six months on. Technical report.
- Transportation for London (2005). Central London congestion charging scheme impacts monitoring. Technical report.
- Transportation for London (2012). Annual report and statement of accounts 2011/12. Technical report.
- Tsivanidis, N. (2019). The aggregate and distributional effects of urban transit infrastructure: Evidence from Bogotá's transmilenio. *Unpublished manuscript*.
- Tyndall, J. (2018). The local labour market effects of light rail transit.
- Verhoef, E. T. (2005). Second-best congestion pricing schemes in the monocentric city. *Journal of Urban Economics*, 58(3):367–388.
- Vickrey, W. (1967). Optimization of traffic and facilities. *Journal of Transport Economics and Policy*, pages 123–136.
- Wardrop, J. G. (1952). Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- 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.
- Zárate, R. D. (2019). Factor allocation, informality and transit improvements: Evidence from Mexico City.
- Zhang, K. and Batterman, S. (2013). Air pollution and health risks due to vehicle traffic. *Science of the total Environment*, 450:307–316.

Zhang, W. and Kockelman, K. M. (2016). Optimal policies in cities with congestion and agglomeration externalities: Congestion tolls, labor subsidies, and place-based strategies. *Journal of Urban Economics*, 95:64–86.

11 Figures

Figure 1: The London Region



Points are DfT traffic count points in the balanced panel and lines are a simplified representation of region's major roads taken from the Geofabrik OpenStreetMaps repository (side roads not shown).

Figure 2: Policy timeline

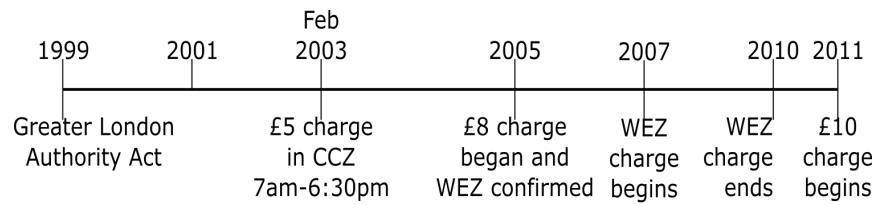
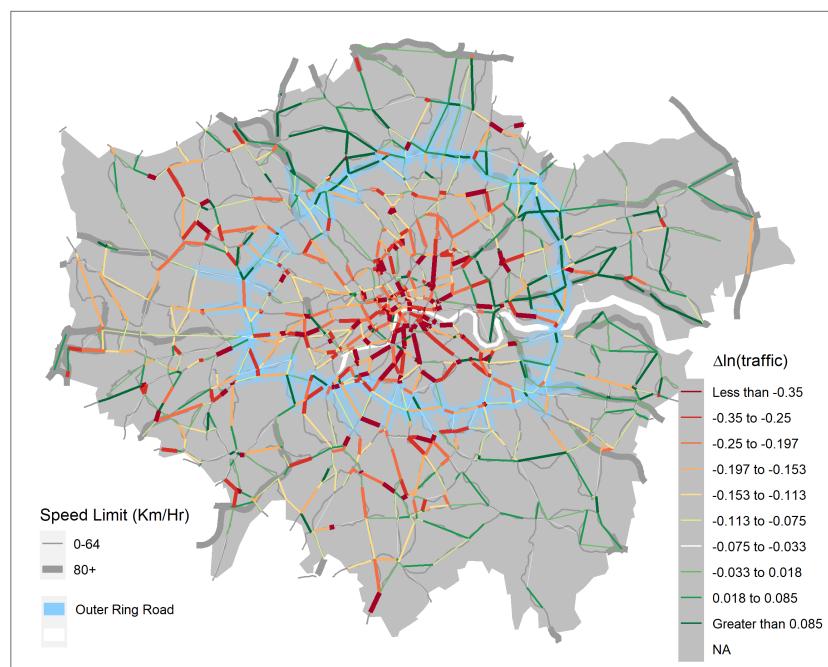
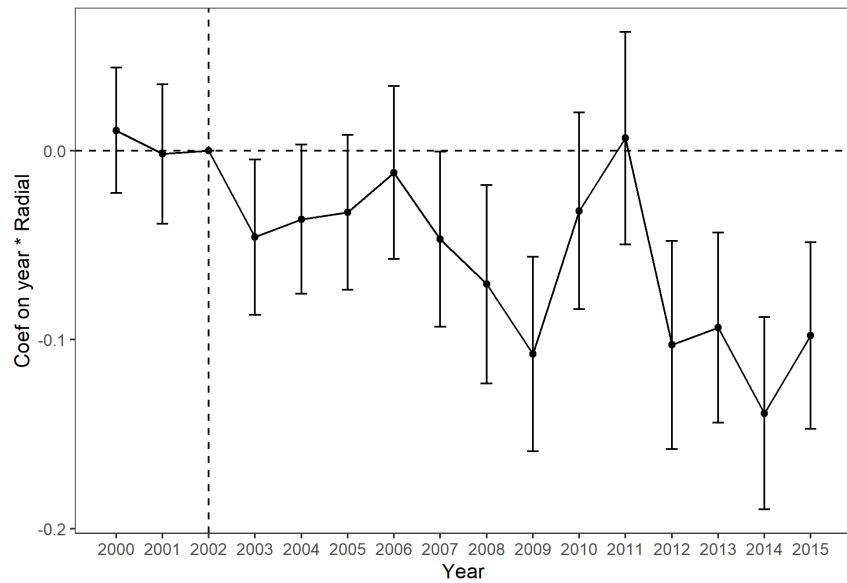


Figure 3: Log-change in count point traffic



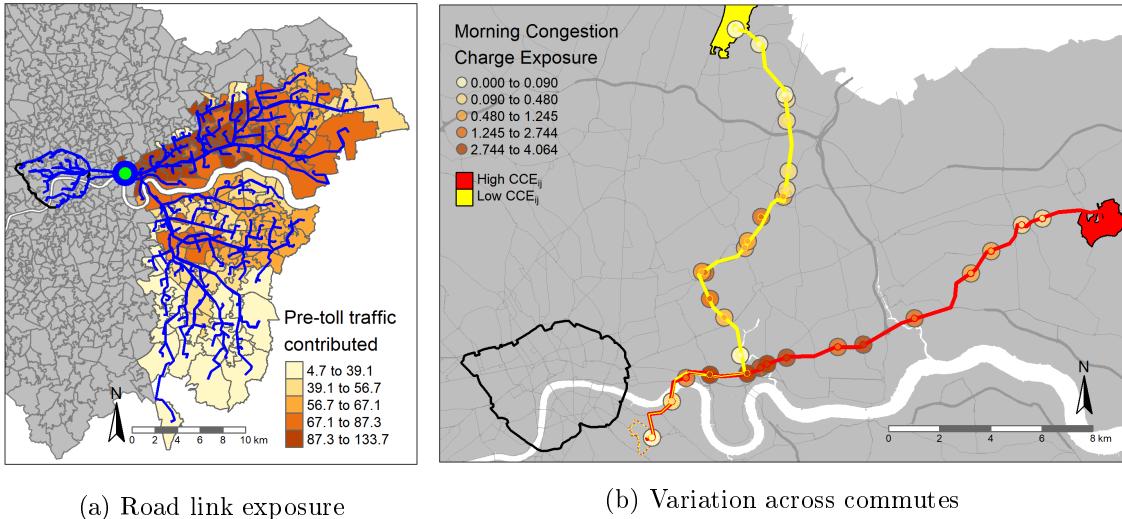
Colours and thickness are proportional difference in logarithms of post- and pre-toll traffic at each count point (relative to zero). I aggregate traffic across travel directions and rush hour to compute count point traffic in each period. Count points appear as linear representations of the road links they monitor.

Figure 4: London's congestion charge and radial road traffic



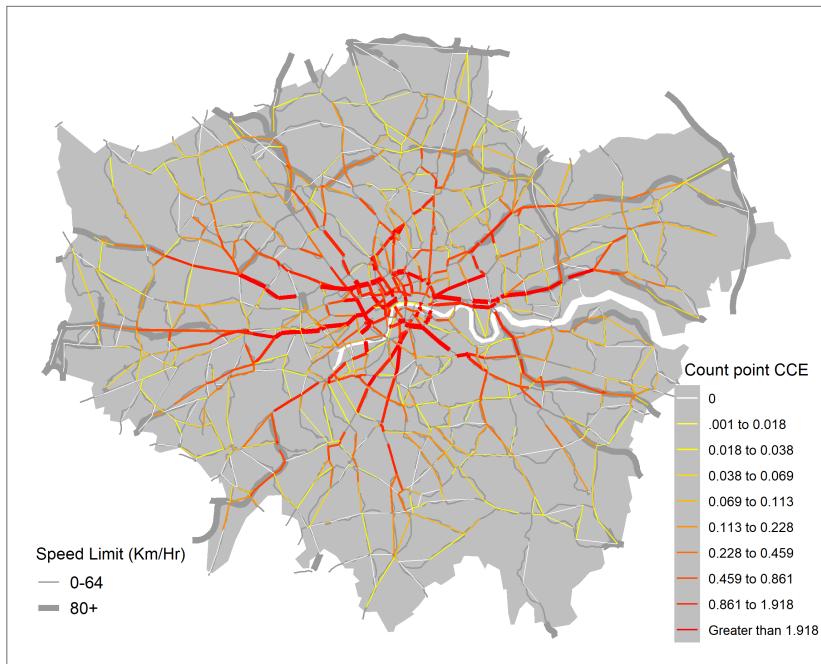
Points are OLS estimates of β_τ the link-year regression $\ln(\text{traffic}_{\ell t}) = \sum_{\substack{\tau=2000, \\ \tau \neq 2002}}^{2015} (D_\tau^t \times \beta_\tau \text{radial}_\ell + D_\tau^t \times f_\tau(\text{lat}_\ell, \text{lon}_\ell)) + \alpha_t + \alpha_\ell + e_{\ell t}$, normalized to zero in 2002, $\text{Traffic}_{\ell t}$ is the mean of morning and evening rush-hour traffic volumes in that link-year, and error bars denote link-clustered 95 percent confidence intervals.

Figure 5: Computing Congestion Charge Exposure



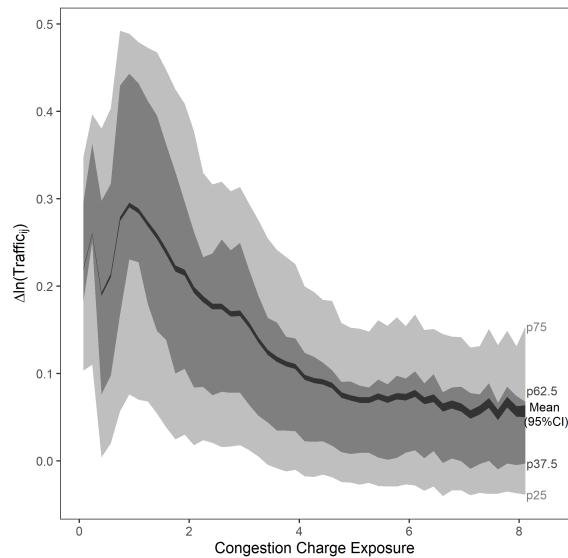
The left panel demonstrates calculating Congestion Charge Exposure for the green highlighted road link where blue paths are OSRM routes contributing to Congestion Charge Exposure and MSOA colours map to quintiles of the amount of morning traffic crossing this link attributed to each origin. The right panel demonstrates the source of cross-commute variation in Congestion Charge Exposure with links' representative points shaded in proportion to their CCE and major roads shown as grey lines with thickness proportional to their speed limits.

Figure 6: The Congestion Charge Exposure Index



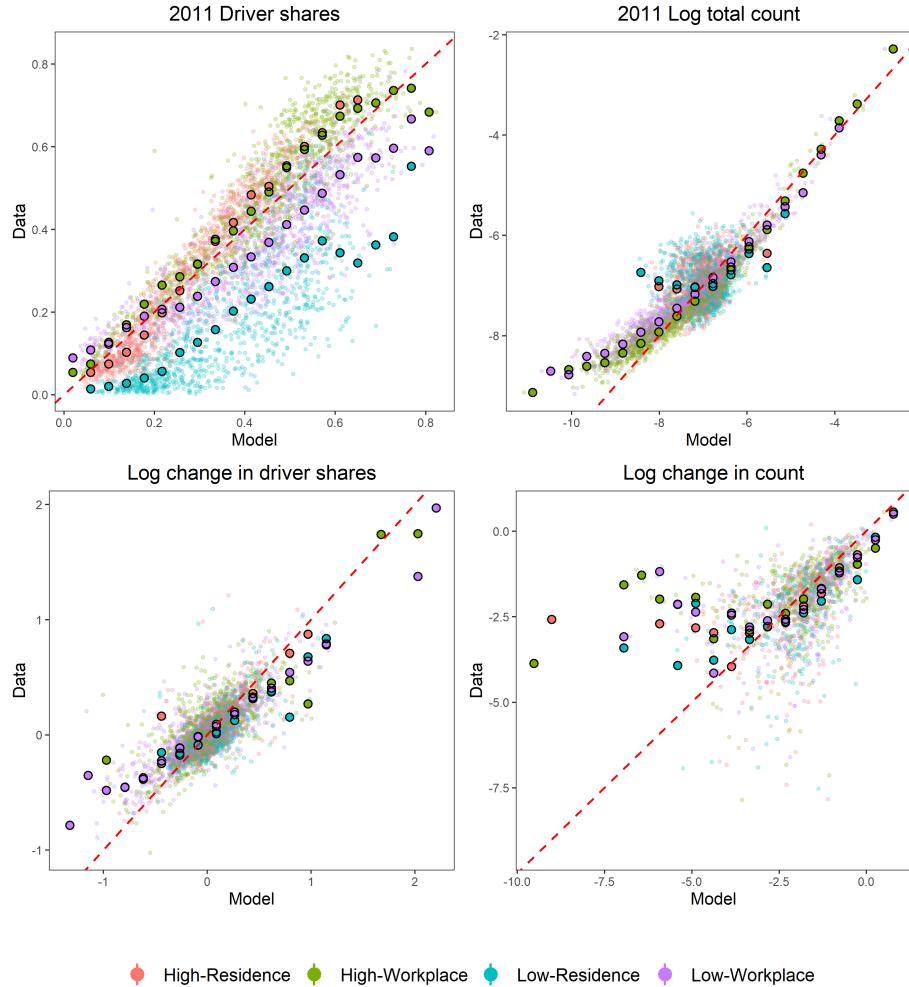
Line color and thickness are both mapped to CCE averaged across travel directions and rush hours at each count point. Count points appear as linear representations of the road links they monitor and grey lines are a subset of the major road network.

Figure 7: Congestion Charge Exposure and traffic declines



The vertical axis is the natural logarithm of differences in MSOA pairs' post- and pre-toll traffic and the horizontal is their Congestion Charge Exposure. The dark area is bounded by 95 percent confidence intervals around conditional means of log-change in traffic computed in 50 evenly spaced bins of CCE, the dark shaded area is bounded by conditional 37.5 and 62.5 percentiles, and the light shaded area is bounded by conditional 25 and 75 percentiles.

Figure 8: Model predicted and actual MSOA outcomes.



Small points plot observed and simulated driver shares and natural logarithm of total counts and large points are binned means. Panel A presents predictions and data for 2011 and changes in panel B are normalized so that there is no total population growth over time.

Figure 9: Percent change employment and population caused by the congestion charge

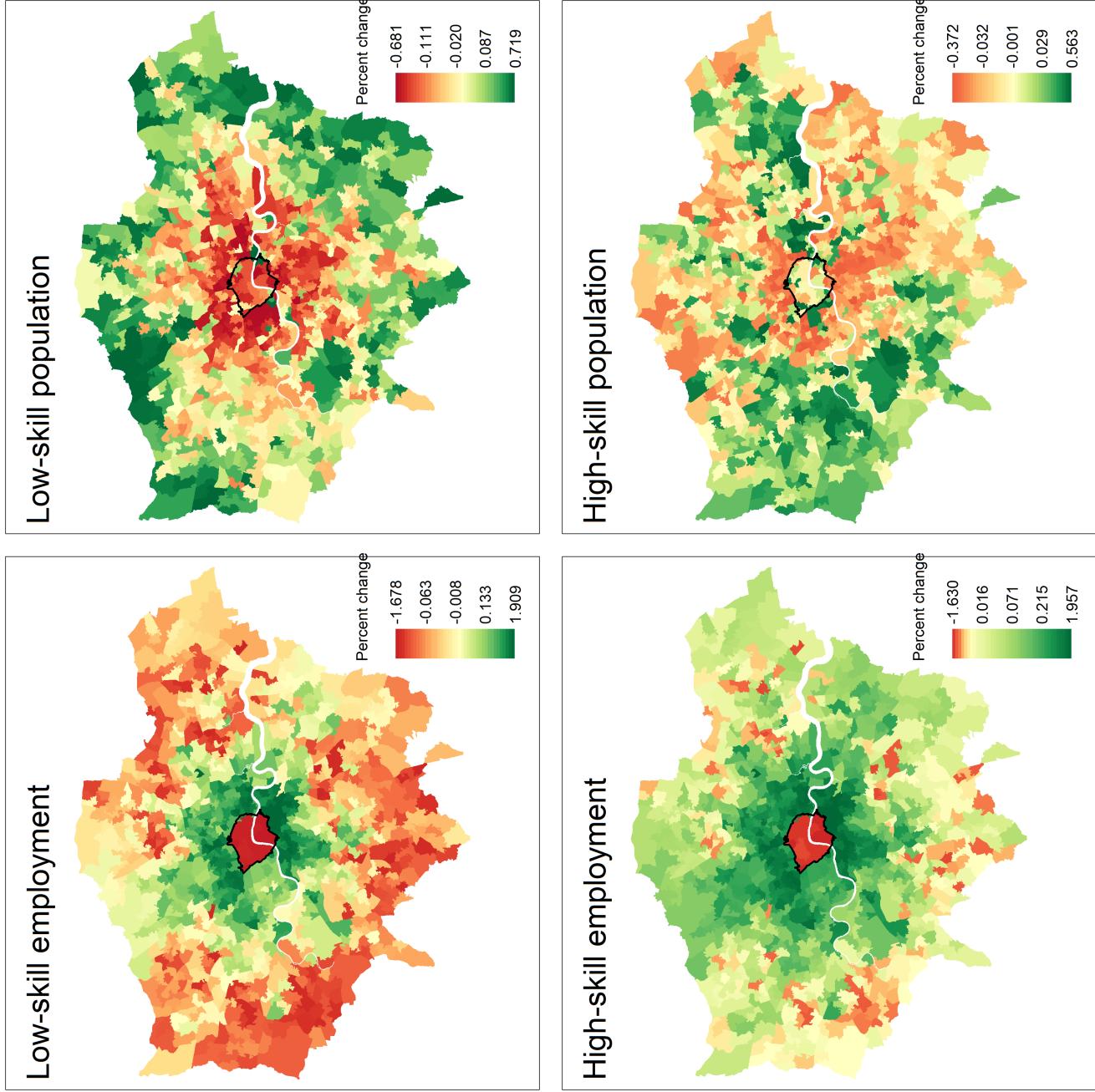
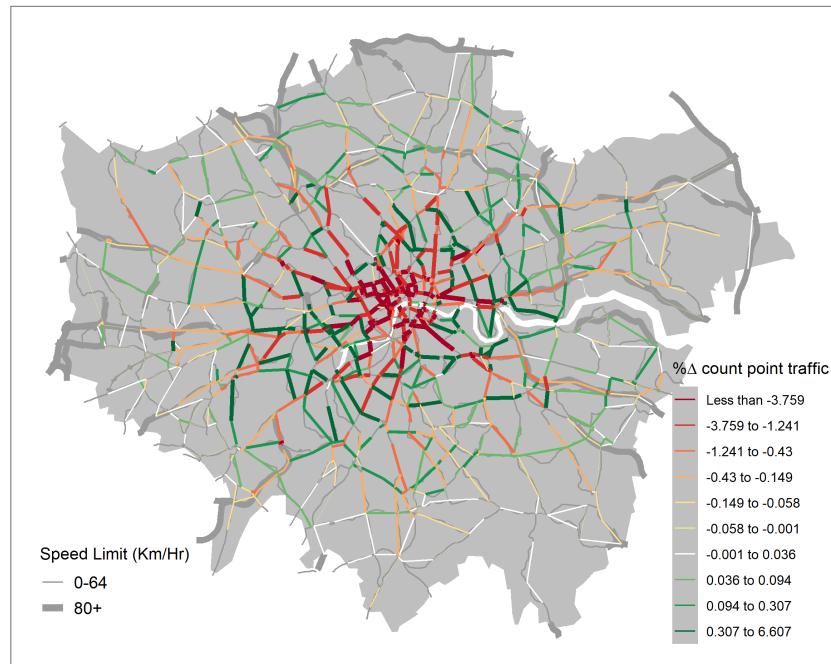


Figure 10: Percent change in traffic caused by the congestion charge



Line color and thickness are both mapped to percent change in traffic (relative to zero) aggregated across travel directions and rush hours at each count point.

12 Tables

Table 1: London commute flows by mode of transportation

	2001	2011	log-change
Entire London Region			
Driver share	0.362	0.254	-0.354
Other share	0.638	0.746	0.156
Total flows	2,447,652	2,936,727	0.182
Within Outer London			
Driver share	0.627	0.445	-0.343
Other share	0.373	0.555	0.397
Total flows	792,876	1,205,940	0.419
Outer London flow share	0.324	0.411	0.237

Top panel summarises commute flows between all MSOA pairs in the London Region and the bottom panel summarises flows for pairs with both MSOAs in Outer London boroughs listed in appendix A.1.1.

Table 2: Congestion Charge Exposure and traffic

	<i>Dependent variable:</i>		
	ln(Traffic)		
	(1)	(2)	(3)
CCE × Post	-0.042*** (0.004)	-0.040*** (0.003)	-0.026*** (0.002)
Dist × Post FE	No	Yes	Yes
Pow/Por × Post FE	None	Both	Both
Controls	No	No	Yes
Observations	1,709,220	1,709,220	1,709,220

Note:

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

Each column presents results of an OLS regression using 2001 total flows plus one as weights and all regressions include pair fixed effects. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14. CCE is scaled to have a standard deviation of one.

Table 3: Traffic's effect on the number commuting

	<i>Dependent variable:</i>			
	ln(Number of commuters)			
	(1)	(2)	(3)	(4)
ln(Traffic)	−0.096*** (0.026)	−0.966*** (0.260)	−0.619*** (0.121)	−0.918*** (0.188)
Dist × Post FE	Yes	No	Yes	Yes
Pow/Por × Post FE	Both	None	Both	Both
Controls	Yes	No	No	Yes
Excl. IV	None	CCE × Post	CCE × Post	CCE × Post
First Stage F-stat	None	106.09	224.54	274.51
Observations	1,709,220	1,709,220	1,709,220	1,709,220

Note:

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

Each column presents results of an OLS or TSLS regression using 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. The dependent variable is the natural logarithm of total flows plus one. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14. First stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table 4: Traffic's effect on mode shares

	<i>Dependent variable:</i>			
	ln(Drivers) - ln(Non-drivers)			
	(1)	(2)	(3)	(4)
ln(Traffic)	0.031 (0.054)	0.894** (0.449)	-1.019*** (0.219)	-1.208*** (0.352)
Dist × Post FE	Yes	No	Yes	Yes
Pow/Por × Post FE	Both	None	Both	Both
Controls	Yes	No	No	Yes
Excl. IV	None	CCE × Post	CCE × Post	CCE × Post
First Stage F-stat	None	106.09	224.54	274.51
Observations	1,709,220	1,709,220	1,709,220	1,709,220

Note:

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

Each column presents results of an OLS or TSLS regression using 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. The dependent variable is the natural logarithm of drivers plus one minus the natural logarithm of commuters by other modes plus one. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14. First stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table 5: Commuting elasticities by skill

	<i>Dependent variable:</i>			
	ln(Commuters)	ln($\frac{\text{Drivers}}{\text{Non-drivers}}$)	ln(Commuters)	$\frac{\text{Drivers}}{\text{Non-drivers}}$
	(1)	(2)	(3)	(4)
ln(Traffic) × low-skill	-1.686*** (0.281)	-0.948** (0.371)	-3.595*** (1.238)	-0.182 (1.004)
ln(Traffic) × (1 - low-skill)	-0.726*** (0.201)	-1.352*** (0.387)	-0.505 (0.320)	-0.011 (0.579)
Pow/Por FE	by year	by year	by skill-year	by skill-year
First Stage F-stat (low-skill)	24.78	24.78	6.49	6.49
First Stage F-stat (1 - low-skill)	40.53	40.53	84.39	84.39
Controls	Yes	Yes	Yes	Yes
Observations	1,709,220	1,709,220	1,709,220	1,709,220

Note:

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

Each column presents results of a TSLS regression using 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. Dependent variables add one to flow variables before taking logarithms and standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14 and add skill-group-by-year controls. Excluded instruments are CCE-skill interactions and partial first stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table 6: Structural gravity

	<i>Dependent variable:</i>	
	$\ln(\pi_{ij})$	
	(1)	(2)
$\ln(1-\pi_{c ij})$	-2.074*** (0.370)	-2.229*** (0.392)
Controls	No	Yes
Dist \times Post FE	Yes	Yes
Pow/Por \times Post FE	Both	Both
Excl. IV	CCE \times Post	CCE \times Post
First Stage F-stat	126.59	106.31
Observations	1,709,220	1,709,220

Note:

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

Each column presents results of a TSLS regression using 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. The dependent variable is the natural logarithm of the MSOA pair's share of annual commuters plus $\frac{1}{N \times (1-N)}$ and independent variable is the natural logarithm of the pair's non-car commute share plus one. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14. First stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table 7: Aggregate effects of the congestion charge.

		Endogenous adjustment mechanisms			
		(1) None	(2) Sorting, wages, and rents	(3) Full model exog. TFP	(4) Full model endog. TFP
Utility	Mean	0.156	0.090	0.021	0.025
	Low ; High	0.21 ; 0.143	0.163 ; 0.073	0.067 ; 0.01	0.082 ; 0.012
Dissimilarity	Inequality	-0.067	-0.090	-0.057	-0.070
	Population	–	-0.099	-0.071	-0.087
	Employment	–	-0.228	-0.160	-0.139
	Car share	–	2.052	0.576	0.581
Traffic		-1.190	-1.190	-0.718	-0.715
Panel B: Commuters exempt from tolls					
Utility	Mean	0.131	0.127	0.029	0.029
	Low ; High	0.130 ; 0.131	0.132 ; 0.125	0.018 ; 0.032	0.015 ; 0.032
Dissimilarity	Inequality	0.000	-0.007	0.005	0.017
	Population	–	-0.003	-0.003	0.018
	Employment	–	-0.006	-0.006	-0.008
	Car share	–	2.790	0.659	0.656
Traffic		-1.190	-1.190	-0.411	-0.411

Each column presents percent difference between actual and counterfactual expected utility, aggregate rent, and drivers' share of all commuters where counterfactuals remove the CCZ with an increasingly rich model. Panel A assumes an £8 toll with lump sum revenue redistribution and panel B assumes commuters are exempt from tolls. Column 1 fixes wages, rents, and commuter decisions at baseline; column 2 endogenizes labour market outcomes; column 3 endogenizes road traffic; and column 4 adds cross-neighbourhood agglomeration externalities.

Table 8: Accounting for endogenous traffic

	No externalities	No mode choice	No location choice	Equilibrium
% Δ Traffic Percentage points lost	-1.190 0	-0.8419 0.348	-1.0115 0.179	-0.715 0.475

The first row presents percent change in traffic caused by London's Congestion Charge and the second presents differences with the first column, which restates the initial shock to traffic. The final column restates the equilibrium traffic decline, which has the endogenous component

$$\Omega \sum_{i,j,\ell} \pi'_{c|ij} \pi'_{ij} 1\{\ell \in R_{ij}\}.$$

A Appendices

A.1 Data and econometric appendix

This section describes details of the data construction and additional econometric results. First, appendix table A.1 presents summary statistics of MSOA pair commute and traffic data for the main analysis sample. The table summarises raw counts of drivers and non-drivers and transforms their by first adding one to all counts as in linear regressions of sections 6 and 8. An MSOA pair's $\ln(\text{traffic})$ is the natural logarithm of the mean of its morning and evening traffic.

A.1.1 Additional data sources

The study area is the London census region which is comprised of 33 boroughs. Following the Office for National Statistics, Inner London is MSOAs in the City of London, Camden, Hackney, Hammersmith and Fulham, Haringey, Kensington and Chelsea, Islington, Lambeth, Lewisham, Newham, Southwark, Tower Hamlets, Wandsworth, and Westminster; Outer London is the remaining 19 boroughs. MSOA shapefiles and cross-walks to boroughs, to OAs, and across time come from the Office for National Statistics Open Ge-

ography Portal. I define constant boundary MSOAs by aggregating splitters and mergers to the union of their 2001 and 2011 geographies.

Shapefiles describing Congestion Charge and Western Extension Zone boundaries come from Transportation for London. I define an MSOA as inside the CCZ or WEZ if it has a non-empty intersection with the relevant charge zone with the exception of ONS code E02000967 (contains Regent's park), E02000890, E02000878, and E02000371 which I exclude from the CCZ. Rare cases of constant boundary MSOAs intersecting both the CCZ and the WEZ default to inclusion in the CCZ and not the WEZ; the only exception is ONS code E02000978 (contains Hyde Park and defined in the WEZ not the CEZ).

Bus routes come from Transportation for London's Open Data Portal as of September 2019. The data contain the locations of all stops in the London Buses network and defines routes as sequences of stops. I identify express bus routes as those with names containing a capital "X" and identify night bus routes as those with names containing a capital "N."

My rail transit database builds on stop locations released by Transportation for London in response to public data request FOI-1451-1819. I augment these data by manually characterizing opening dates, stop locations, and routes of the London Underground, Overground, Thameslink, TfL Rail, Docklands Light Rail (DLR), and Tramlink systems. I define heavy rail transit as all rail systems operating entirely in exclusive rights of way to separate differentiate and Tramlink, a suburban light rail system using a mixture of dedicated rights of way and at-grade track.

Baseline controls exclude rail stations built after 2001, but several stations and lines opened during my study period. This includes Wood Lane tube station and substantial expansions to Overground and Dockland's Light Railway systems. Fortunately, reduced form results are robust to controlling for post-2001 Overground, DLR, Underground extensions.

A.1.2 Commute flows

For census each year (2001 and 2011) I create two 967 by 967 commute flow matrices for constant boundary MSOAs in Greater London.

The raw data are measured for census blocks known as output areas (OAs) and workplace zones (WZs). In each year, I obtain three commute flow matrices that each describe one of three mode choices: drives car or van, other (includes transit, car passenger, walk, etc), and works from home. The universe of interest is Greater London residents aged 16 and over in employment the week before the census in 2011. The 2001 census set an upper age bound of 74, but this discrepancy over time should be unimportant in practice. In 2001, the data are square matrices of output area to output area flows and in 2011 the workplaces are given by workplace zones instead of output areas.³⁶ I aggregate flows to constant boundary MSOAs before merging the data over time.³⁷

The 2001 census randomly adjusts observations of 1 and 2 commuters to 0 and 3 at a highly granular level before aggregating to three modes of transportation and releasing the data. The 2011 data are not adjusted and are instead available for less granular mode categories. Therefore, 2001 data available for research contain synthetic counts of zero and three so that naively aggregating to MSOAs and merging over time creates synthetic multiples of three that are inconsistent with 2011 data.

To make the data comparable over time, I adjust 2011 counts to match the 2001 census small cell adjustment method and bin both years to multiples of three before aggregating to MSOA pairs. The adjustment procedure is as follows:

1. Randomly adjust observations of 1 and 2 to 0 and 3 in 2011 data using the adjustment probability matrix proposed by Stillwell and Duke-Williams (2007):

³⁶Output areas are defined to contain 40 to 125 households and the newer workplace zone geography is designed to contain a roughly constant number of workers.

³⁷2011 data are from census table WU03BUK and 2001 are from the level 3 special workplace statistics.

	0	3
1	2/3	1/3
2	1/3	2/3

where cell ij is the probability the value in row i is adjusted to the value in column j . For example, a cell with one commuter has a $2/3$ probability of being adjusted to zero.

2. Bin both years' data around multiples of three using the function $bin(x) = 3 \times round(x/3)$ where $round$ rounds to the nearest integer.
3. Aggregate both years to constant boundary MSOAs.
4. Create internally consistent commute flows by aggregating across modes of transportation for each MSOA pair.

Throughout, I omit commuters who report working from home or within their residence MSOA. The result is two comparable flow matrices per year with zeros along the diagonals.

Table A.11 presents the share of London Region employment captured by my adjusted MSOA pair commuting data in each year to assess two sources of measurement error. First, commute flows mechanically undercount employment by excluding workers who live outside of the London Region. Second, small cell adjustments produce random measurement error which should cancel out in aggregate. Commute flows account for 72 percent of regional employment in 2001 and 75 percent in 2011. Undercounting is more prominent for drivers and roughly constant across years, suggesting that it is driven by unobserved inflows to the region rather than unexpected small cell adjustment errors.

A.1.3 Validating routing assumptions

I compute paths between MSOA pairs using OSRM version 5.22.0, which minimizes driving times along an uncongested road network accounting for turn restrictions, one-way roads, speed limits etc. However, these paths may deviate from drivers' preferred routes in rush-hour traffic. Routes extracted from Google Maps assuming typical rush-hour traffic would be ideal but their terms of service explicitly forbid scraping Google's routing Application Programming Interface (API) (Google, 2020). This section uses a limited sample of fastest-in-traffic paths to validate the assumptions implicit in OSRM routing.

I benchmark OSRM paths against a non-random sample of fastest-in-traffic paths between MSOA pair centroids obtained from the API of an authoritative mapping and routing service during January 2020. These are travel-time minimizing paths given traffic predicted on January 2020 for a 09:00 arrival time on an upcoming October Wednesday. I use these paths to define fastest-in-traffic routes, pairwise traffic, monitored road length, and I trim outliers as in the main text for morning rush hour in the pre- and post-toll periods.

Figure A.10 plots pairwise traffic computed with OSRM routes against fastest-in-traffic counterparts in levels and log-changes for the 36,565 MSOA pairs covered by the latter data. Small points are MSOA pairs, large points are binned means, and the red 45 degree line indicates a perfect fit. Prominent diagonals in both plots suggest a good fit but OSRM slightly under-estimates traffic at low volumes and generates low outliers in log-changes.

Table A.12 shows that corresponding ordinary and weighted least squares regressions of fastest-in-traffic routed against OSRM traffic have slopes near one and an R^2 of 0.690 in levels. The fit is slightly worse in changes which give a slope of 0.76 and R^2 of 0.566. Adding origin, destination, and distance band fixed effects attenuates this slope to 0.57 but increases the R^2 and the final column weights by one plus baseline total commuting to find a similar slope and an R^2 of 0.697. Finally, table A.13 shows that similar patterns emerge

in regressions using monitored road length in place of traffic.

All told, these patterns suggest that OSRM routes give a realistic view of the traffic commuters face. In hindsight, this seems reasonable given that *Wardrop's First Principle* predicts that identical drivers will arbitrage away differences in travel costs across routes serving an origin-destination pair (Wardrop, 1952; Small et al., 2007).

A.1.4 Accounting for traffic

It is tempting to decompose road link traffic between commuters and other traffic, but my data are unsuited for this purpose. In particular, the accounting identity that theoretically separates endogenous and exogenous traffic (see section 7.4) does not hold for all road links in my data. This appendix compares the distributions of traffic and commuting and argues that incompatibilities stem from random measurement errors that do not bias aggregate statistics or this paper's central results.

Without measurement error, the number of car commuters crossing road link ℓ in hour h is $n_\ell^h = \sum_{i,j} 1\{\ell \in R_{ij}^h\} n_{ijc}$ and daily mean traffic is $\text{traffic}_\ell = \frac{1}{H} \sum_h (n_\ell^h + \mathcal{O}_\ell^h)$ with H hours each day and \mathcal{O}_ℓ^h non-commuters each hour. This accounting identity is broken by measurement error in traffic_ℓ , routes R_{ij}^h , and commute flows n_{ijc} . First, I only observe traffic from 7 am to 7 pm and cannot guarantee that commuters drive during my specified rush hour periods. Second, I cannot guarantee that all commuters use the OSRM's optimal routes. Third, enumeration error and daily idiosyncrasies in commuters' decisions affect traffic counts. Fourth, there is rounding in commute flow data.³⁸ Fortunately, all of these errors are plausibly random across road links.

Figure A.11 plots road link traffic_ℓ against estimated car commuters $\frac{1}{H} \sum_h \sum_{i,j} 1\{\ell \in R_{ij}^h\} n_{ijc}$, large points are binned means by year and red dashes denote the 45 degree line.

³⁸Section A.1.2 discusses rounding in commute flows.

While there is considerable variation around the conditional expectation function, conditional mean traffic tends to exceed car commuting with the exception of a small number of high outliers in estimated commute flows. Ignoring outliers also reveals that traffic and commuting are positively correlated throughout the distribution, providing support for the proposition that measurement error is random across road links.

Figure A.12 presents pre- and post-toll empirical cumulative density functions of road link traffic and commuter estimates. In total, car commuters account for 79 percent of 2001 traffic and 66 percent of 2011 traffic. Encouragingly, the distributions across road links show that total traffic stochastically dominates estimated commuters, leaving a role for non-commuter traffic at all points in the distribution.

A.1.5 Maximum likelihood commuting elasticities

This appendix describes estimates of traffic's effect on commuting using the high dimensional poisson maximum likelihood estimator (MLE) implemented by Berge et al. (2018) in concert with the two-step control function instrumental variables (CFIV) estimator proposed by Rivers and Vuong (1988).

The estimating equation for location choice is

$$\pi_{ijt} = \exp \left(\delta^n \ln(\text{traffic}_{ijt}) + X'_{ij} \beta_t + \theta_{b(ij)t} + \alpha_{it} + \alpha_{jt} + \alpha_{ij} + \psi \hat{\nu}_{ijt} + u_{ijt} \right)$$

where $\hat{\nu}_{ijt}$ is the first stage residual fit by applying 2001 flow weighted OLS to equation 4. Mode choice estimates recast equation 14 into the equivalent poisson regression to avoid running a fractional logit regression. Specifically, I lengthen the data so that observations

are workplace-residence-year-mode quartets and estimate:

$$\pi_{ijtm} = \exp \left(\delta^c \ln(\text{traffic}_{ijt}) D_m + X'_{ij} \beta_{tm} + \theta_{b(ij)tm} + \alpha_{itm} + \alpha_{jtm} + \alpha_{ijt} + \alpha_{ijm} + \psi \hat{\nu}_{ijtm} + u_{ijtm} \right) \quad (28)$$

Table A.14 presents non-linear CFIV estimates of commuting's elasticity of traffic. All regressions are fit by maximum likelihood, the first row presents coefficients on log-traffic and the second row presents coefficients on the first stage residuals. The first three columns are weighted poisson estimates using 2001 flows as weights and the final column replicates column 3 using a negative binomial model. Standard errors in round parenthesis are clustered by both residence and workplace and square brackets are clustered by pair. Two-way clustered standard errors are robust to arbitrary residual correlation across routes that share an origin or destination to give a high estimate on confidence interval width while pair clustered standard errors give a low estimate.³⁹

Poisson estimates support the hypothesis that traffic effects commuting the pattern across specifications mirrors TSLS estimates. However, MLE estimates are smaller and less powerful because the estimator drops observations in clusters that always exhibit zero flows. The weighted poisson estimate in column 3 associates a 3.2 percent decrease in commuting with a ten percent increase in traffic and this effect is statistically significant using pair clustered inference. The negative binomial estimator produces a smaller but more precisely estimated negative effect.

Table A.15 presents similar non-linear CFIV estimates of driver share's elasticity of

³⁹CFIV standard errors presented here are not asymptotically valid since computational burden precludes computing bootstrap standard errors recommended in Wooldridge (2010). I present QMLE-style standard errors to give a sense of uncertainty in point estimates and compute p-values to generate intuition rather than formally test the frequentist hypothesis that δ^n or δ^c are zero. See Berge et al. (2018) for computational details of the high-dimensional fixed effects maximum likelihood estimators.

traffic. Each column is a separate regression based on equation 28 where an observation is a unique combination of an MSOA pair, year, and mode of transportation for a total of 3,435,604 data points. Standard errors in round parenthesis are two-way clustered by residence-mode and workplace-mode and square brackets present pair-mode clustered standard errors.

Poisson estimates of δ^c omit a substantial number of observations whose cluster never has any drivers. This is particularly problematic because many fixed effects are needed to force identification based on relative mode shares. As a result, MLE mode choice regressions with the full set of fixed effects fail to pick up the variation in the data that delivers a negative elasticity in the linear case.⁴⁰

A.2 Theory and quantitative appendix

A.2.1 Deriving choice probabilities and proving theorem 1

This appendix provides detailed derivations of commuters choice probabilities under the assumptions described in section 5 and concludes by proving theorem 1.

Mode choice: First, define the ex-ante expected mean utility of commuting from i to j for a worker with skill g as $\mu_{ijg} = \frac{w_{jg}B_{ig}}{Q_i^{1-\beta}\bar{d}_{ij}}$ where \bar{d}_{ij} is a commute cost index (often called the inclusive value in nested logit applications) equal to each commute's expected value of $\max_m \frac{\nu_{m\omega}}{d_{ijm}}$. With many workers, mode m 's share of skill g choices is the conditional probability of a skill g worker taking mode m after choosing to commute from i to j and

⁴⁰The fourth column of table A.15 shows that the negative binomial overdispersion parameter reached its predefined maximum value at optimal parameter values. This suggests no overdispersion in the mode choice regression.

realizing the preference $\nu_{m\omega}$:

$$\begin{aligned}
\pi_{m|ijg} &= \Pr(u_{ijgm} \geq u_{ijgm'} \forall m' \neq m) \\
&= \Pr\left(\frac{\mu_{ijg}}{d_{ijm}} \nu_{m\omega} \epsilon_{ij\omega} \geq \frac{\mu_{ijg}}{d_{ijm'}} \nu_{m'\omega} \epsilon_{ij\omega} \forall m' \neq m\right) \\
&= \Pr\left(\frac{\nu_{m\omega}}{d_{ijm}} \geq \frac{\nu_{m'\omega}}{d_{ijm'}} \forall m' \neq m\right) \\
&= \int_0^\infty \prod_{m' \neq m} \Pr(u \geq \frac{\nu_{m'\omega}}{d_{ijm'}}) \Pr(u = \frac{\nu_{m\omega}}{d_{ijm}}) du \\
&= \int_0^\infty \exp\left(-u^{-\eta} \sum_{m' \neq m} d_{ijm'}^{-\eta}\right) \exp(-(ud_{ijm})^{-\eta}) d_{ijm}^{-\eta} \eta u^{-\eta-1} du \\
&= \int_0^\infty \exp\left(-u^{-\eta} \sum_{m'} d_{ijm'}^{-\eta}\right) d_{ijm}^{-\eta} \eta u^{-\eta-1} du.
\end{aligned}$$

To evaluate this integral, define $y = -u^{-\eta} \sum_{m'} d_{ijm'}^{-\eta}$, which implies $dy = \eta u^{-\eta-1} \sum_{m'} d_{ijm'}^{-\eta} du$ or $\eta u^{-\eta-1} du = \frac{dy}{\sum_{m'} d_{ijm'}^{-\eta}}$. This, paired with the monotonic negative relationship of y and u , allows the change of variables:

$$\begin{aligned}
\pi_{m|ijg} &= \int_0^\infty \exp\left(-u^{-\eta} \sum_{m'} d_{ijm'}^{-\eta}\right) d_{ijm}^{-\eta} \eta u^{-\eta-1} du. \\
&= \int_{-\infty}^0 e^y \frac{d_{ijm}^{-\eta}}{\sum_{m'} d_{ijm'}^{-\eta}} dy \\
&= \frac{d_{ijm}^{-\eta}}{\sum_{m'} d_{ijm'}^{-\eta}}.
\end{aligned} \tag{29}$$

Since mode choice is independent of skill, the conditional probability of any worker choosing mode m given commute ij is $\pi_{m|ij} = \frac{d_{ijm}^{-\eta}}{\sum_{m'} d_{ijm'}^{-\eta}}$.

Commute cost indices are the expected costs of each commute: A Fréchet random variable x with CDF $F(x) = e^{-(\frac{x}{s})^{-\eta}}$ has expectation $E(x) = s\Gamma\left(1 - \frac{1}{\eta}\right)$. Given route choice, workers know $\mu_{ijg}\epsilon_{ij\omega}$ and d_{ijm} , but must form expectations about transport costs

since they do not know yet what mode they will choose after learning their $\nu_{m\omega}$. At this stage, inverse shock-inclusive costs of commute ij have the CDF:

$$\begin{aligned}
 G_{ij}(d^{-1}) &= Pr\left(\frac{1}{d} \geq \frac{\nu_{mw}}{d_{ijm}} \forall m\right) \\
 &= \prod_m Pr\left(\frac{d_{ijm}}{d} \geq \nu_{mw}\right) = \exp\left(-d^\eta \sum_m d_{ijm}^{-\eta}\right) \\
 &= \exp\left(-\left[\frac{d^{-1}}{\left(\sum_m d_{ijm}^{-\eta}\right)^{\frac{1}{\eta}}}\right]^{-\eta}\right) \\
 \implies E(\nu_{mw}/d_{ijm}|ij) &= \left[\sum_m d_{ijm}^{-\eta}\right]^{\frac{1}{\eta}} \Gamma\left(1 - \frac{1}{\eta}\right)
 \end{aligned}$$

and expected commute costs are

$$\begin{aligned}
 \bar{d}_{ij} &= E(\nu_{mw}/d_{ijm}|ij)^{-1} \\
 &= \left(\Gamma\left(1 - \frac{1}{\eta}\right)\right)^{-1} \left[\sum_m d_{ijm}^{-\eta}\right]^{-\frac{1}{\eta}}
 \end{aligned}$$

which are independent of skill.

Location choice: Before choosing a mode of transportation, workers jointly choose locations based on location pair shocks ϵ_{ijw} and ex-ante expected mean utility μ_{ijg} . The share

of skill g workers commuting from i to j is

$$\begin{aligned}
\pi_{ij|g} &= \Pr(\mu_{ijg}\epsilon_{ijw} \geq \mu_{rsg}\epsilon_{rs w} \forall rs \neq ij) \\
&= \int_0^\infty \prod_{rs \neq ij} \Pr(u \geq \mu_{rsg}\epsilon_{rs w}) \Pr(u = \mu_{ijg}\epsilon_{ijw}) du \\
&= \int_0^\infty \prod_{rs \neq ij} F\left(\frac{u}{\mu_{rsg}}\right) \frac{d}{du} F\left(\frac{u}{\mu_{ijg}}\right) du \\
&= \int_0^\infty \exp\left(-u^{-\theta} \sum_{rs \neq ij} \mu_{rsg}^\theta\right) \exp\left(-u^{-\theta} \mu_{ijg}^\theta\right) \mu_{ijg}^\theta \theta u^{-\theta-1} du \\
&= \frac{\mu_{ijg}}{\sum_{rs} \mu_{rsg}} = \frac{(w_{jg} B_{ig} Q_i^{\beta-1})^\theta \bar{d}_{ij}^{-\theta}}{\sum_r \sum_s (w_{sg} B_{rg} Q_r^{\beta-1})^\theta \bar{d}_{rs}^{-\theta}}
\end{aligned} \tag{30}$$

where the final line follows from evaluating the integral as in the mode choice nest.

Expected utility: The properties of the Fréchet distribution imply that mean utility of skill g workers is

$$E(u|g) = \gamma' \gamma \left[\sum_i \sum_j \left(\frac{B_{ig} w_{jg}}{\bar{d}_{ij} Q_i^{1-\beta}} \right)^\theta \right]^{\frac{1}{\theta}} \tag{31}$$

and the law of iterated expectations shows that equation 9 is equilibrium expected utility. This lays the necessary foundations for deriving empirical implications.

Proof of theorem 1: First note that commute flows from i to j are $\pi_{ij} = \sum_g \pi_g \pi_{ij|g}$. Combining with equations 5, 6, 29, 30, and partially differentiating gives elasticities in equations 10 and 11. Substituting into equation 9 and partially differentiating gives equation 12. \square

A.2.2 Negative commuting elasticity implies general equilibrium effects

In this section, I show that a negative causal effect of commute costs on commuting is necessary and sufficient for an exogenous shock to commute costs to have spatially dependent general equilibrium effects through wages and rents. It naturally follows that traffic, a determinant of commute costs, has general equilibrium effects if and only if it affects commute flows. This is true in a wide class of models generating a commuter gravity equation.

Proposition 1 *If commute flows between each location pair i, j in a finite set of locations $\mathcal{I} = \{1, 2, \dots, N\}$ s.t. $N > 3$ are determined by the following system of equations*

$$\begin{aligned}\pi_{ij} &= w_j q_i d_{ij}^\delta \text{ (gravity)} \\ \pi_i &= \sum_j \pi_{ij} \text{ (housing demand)} \\ \pi_j &= \sum_i \pi_{ij} \text{ (labour supply)} \\ w_j &= A_j \pi_j^\alpha \text{ (labour demand)} \\ q_i &= B_i \pi_i^\beta \text{ (housing supply)}\end{aligned}$$

where $\{d_{ij}\}$ are potentially endogenous commute costs, $\{\pi_{ij}, \pi_i, \pi_j\}$ are endogenous allocations, $\{w_j, q_i\}$ are endogenous prices, $\{\delta, \alpha, \beta\}$ are exogenous parameters satisfying $\delta < 0$, $\alpha \notin \{0, 1\}$, $\beta \notin \{0, 1\}$, and fundamentals $\{A_j, B_i\}$ satisfy $A_j > 0$, and $B_i > 0$ for at least four locations $i, j, r, s \in \mathcal{I}$, then

$$\delta \neq 0 \iff \exists i, j, r, s \in \mathcal{I} \text{ s.t. } \frac{dq_i}{dd_{is}} \neq 0 \text{ and } \frac{dw_j}{dd_{rj}} \neq 0.$$

Proof: I prove both necessity and sufficiency by renaming location indices and reducing

the system of equations to

$$w_j = A_j^{\frac{1}{1-\alpha}} \left(\sum_r q_r d_{rj}^\delta \right)^{\frac{\alpha}{1-\alpha}}$$

$$q_i = B_i^{\frac{1}{1-\beta}} \left(\sum_s w_s d_{is}^\delta \right)^{\frac{\beta}{1-\beta}}.$$

For necessity, setting $\delta = 0$ and combining shows that all w_j and q_i are uniquely determined by the set of fundamentals $\{A_j, B_i\}$ if commute costs have no causal effect on total flows. For sufficiency, note that with $\delta \neq 0$ perturbing any commute cost to j , such as d_{rj} , will have first-order effects on w_j since $\frac{\alpha}{1-\alpha} \neq 0$.⁴¹ A similar argument applies to the relationship between q_i and any commute cost d_{is} . \square

Remark: Necessity of $\delta \neq 0$ only applies to the case of a shock to pairwise commute costs such as a congestion charge reducing non-commuter traffic. If there were an initial shock to some fundamental A_j or B_i , such as reduced noise and air pollution in some location, there could be general equilibrium effects even if $\delta = 0$. However, these effects would be spatially independent in that they affect each neighbourhood identically.

A.2.3 Equilibrium properties

With exogenous productivity, arguments developed by Tsivanidis (2019) guarantee the existence of at least one labour market equilibrium.⁴² The following lemma appropriates this fact for my context.

Lemma 1 (Existence of a labour market equilibrium with exogenous productivity) *Given commute costs $\{\bar{d}_{ij}\}$, parameters $\{\theta, \sigma, \alpha, \beta\}$, and fundamentals $\{B_{ig}, A_j, a_{jg}, \pi_g\}$ there*

⁴¹This perturbation has higher order effects via q_r , but this is not necessary to prove the current proposition.

⁴²Labour market equilibrium's existence follows the proof in an early appendix to Tsivanidis (2019) accessible from the University of Chicago's dissertation archive as DOI: 10.6082/dk7g-zf31.

exist allocations $\{\pi_{ijg}\}$ and prices $\{w_{jg}, F_j, Q_i\}$ satisfying the conditions of a closed city labour market equilibrium.

Proof: Equations 18, 19, 20, and 21 can be combined into a continuous and homogeneous of degree one map from choice probabilities onto themselves. Applying Brouwer's fixed point theorem shows that given parameters $\{\theta, \eta, \sigma, \alpha, \beta\}$, fundamentals $\{B_{ig}, A_j, a_{jg}\}$, population $\{\Omega, \pi_g\}$, and commute costs $\{\bar{d}_{ij}\}$, there exist equilibrium allocations $\{\pi_{ijg}\}$ and prices $\{w_{jg}, F_j, Q_i\}$ satisfying equations 18, 19, 20, and 21. \square

Lemma 1 establishes that a labour market equilibrium is determined by location choices, which are a continuous function of commute costs $G(\{\bar{d}_{ij}\})$. I now establish the existence of equilibrium mode shares that pin down traffic and commute costs following from any given labour market equilibrium.

Lemma 2 (Existence of mode shares and commute costs) *Given location choices $\{\pi_{ijg}\}$, parameters $\{\eta, \kappa, \varphi^{morn}, \varphi^{eve}\}$, fundamentals $\{\pi_g, \Omega, \alpha_{ic}, \alpha_{jc}, d_{ijo}, e_\ell^h, CCE_\ell^h, toll_{ij}, R_{ij}^h\}$ and a normalization ensuring $\varphi^h CCE_\ell^h + e_\ell^h \geq 0 \forall \ell, h$, there exist traffic, mode shares, and commute costs $\{traffic_\ell^h, \pi_{c|ij}, \bar{d}_{ij}\}$ satisfying the conditions of a closed city general equilibrium.*

Proof: The normalization $\varphi^h CCE_\ell^h + e_\ell^h \geq 0$ allows re-writing equation 24 as

$$\begin{aligned} \ln \left(\frac{\pi_{c|ij}}{1 - \pi_{c|ij}} \right) = & \eta \alpha_{ic} + \eta \alpha_{jc} + \eta d_c(ij) - \eta \ln(toll_{ij}) + \eta \ln(d_{ijo}) \\ & - \eta \kappa \ln \left[\frac{1}{2} \sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} \left(\varphi^h CCE_\ell^h + e_\ell^h + \Omega \sum_{r,s} 1\{\ell \in R_{rs}^h\} \pi_{c|rs} \pi_{rs} \right) \right]. \end{aligned} \quad (32)$$

Conditional on $\{\pi_{ijg}\}$ and fundamentals, equation 32 is a continuous function mapping N^2 mode shares onto their log-odds. Exponentiating and rearranging to isolate $\pi_{c|ij}$ on the

left hand side produces a continuous mapping from mode shares onto themselves. Since $\pi_{c|ij} \in [0, 1] \forall i, j$, Brouwer's fixed point theorem implies that there exists a vector of mode shares $\{\pi_{c|ij}\}$ satisfying equation 24. Combining with location choices π_{ijg} and substituting into equation 23 defines $traffic_\ell^h$ and equation 22 determines \bar{d}_{ij} . \square

Remark: The normalization $\varphi^h CCE_\ell^h + e_\ell^h \geq 0$ eliminates corner solutions following from the non-negativity constraint $traffic_\ell^h \geq 0$. However, this imposes a potentially undesirable scale on e_ℓ^h that my quantitative analysis avoids by explicitly imposing non-negativity constraints in equation 39 to solve the model given arbitrary $e_\ell^h \in \mathbb{R}$.

Lemma 2 guarantees the existence of mode shares and traffic externalities consistent with any labour equilibrium and shows that equilibrium commute costs can be written as function of location choices $D(\{\pi_{ijg}\})$ because $\pi_{ij} = \sum_g \pi_{ijg}$. Further inspection of equation 32 shows that $D(\{\pi_{ijg}\})$ is continuous and combining with lemma 1 gives the equilibrium operator $\ddot{D}(\{\bar{d}_{ij}\}) = \{\bar{d}_{ij}\} - D(G(\{\bar{d}_{ij}\}))$ which is an $N \times N$ matrix of zeros at equilibrium commute costs $\{\bar{d}_{ij}\}$. The following theorem guarantees that these commute costs exist and characterizes a closed city general equilibrium.

Theorem 2 (Existence of a general equilibrium with exogenous productivity) *Given $\{\theta, \sigma, \alpha, \beta, \varphi^{morn}, \varphi^{eve}, B_{ig}, A_j, a_{jg}, \pi_g, \alpha_{ic}, \alpha_{jc}, d_{ijo}, e_\ell, CCE_\ell, toll_{ij}, R_{ij}^h\}$ satisfying $\varphi^h CCE_\ell^h + e_\ell^h \geq 0 \forall \ell$ and h , there exist allocations $\{\pi_{ijg}, \pi_{c|ij}, traffic_\ell^h\}$ and prices $\{w_{jg}, F_j, Q_i\}$ satisfying the conditions of a closed city general equilibrium.*

Proof: Lemmas 1 and 2 show that equilibrium commute costs are a fixed point on the continuous operator $D(G(\{\bar{d}_{ij}\}))$ and that they pin down endogenous variables $\{\pi_{ijg}, \pi_{c|ij}, traffic_{ij}, w_{jg}, F_j, Q_i\}$. It remains to show that $D(G(\{\bar{d}_{ij}\}))$ satisfies the conditions of Brouwer's fixed point theorem, which is true if the $\{\bar{d}_{ij}\} = D(\cdot)$ given by lemma 2 are confined to a compact and convex set.

Potential commute costs' compactness and convexity turns out to hold without normalization since a fixed population Ω naturally bounds traffic. To see this, first note that $\bar{d}_{ij} = \gamma^{-1} \left(d_{ijc}^{-\eta} + d_{ijo}^{-\eta} \right)^{-\frac{1}{\eta}}$ is a continuous and increasing function of $traffic_{ij} = \frac{1}{2} \sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} traffic_\ell^h$. Since $\sum_h \sum_\ell traffic_\ell^h \geq \sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} traffic_\ell^h \forall i, j$, the existence of a finite upper bound on total traffic guarantees a finite upper bound on each element of the commute cost vector $\{\bar{d}_{ij}\}$.

Total traffic is bounded because it is the sum over a finite number of road links ℓ , each with a finite maximum possible $traffic_\ell^h$ in each rush hour. Formally, $traffic_\ell^h = \Omega \sum_{i,j} \pi_{ijc} 1\{\ell \in R_{ij}^h\} + \varphi^h CCE_\ell^h + e_\ell^h$ has the finite upper bound $\Omega \sum_{i,j} 1\{\ell \in R_{ij}^h\} + \varphi^h CCE_\ell^h + e_\ell^h$. Since zero trivially bounds $traffic_{ij}$ from below, it follows that the vector $\{\bar{d}_{ij}\}$ is confined to a compact subset of \mathbb{R}^{N^2} and Brouwer's fixed point theorem guarantees the existence of a fixed point on the equilibrium commute cost operator. \square

Multiplicity: As is common in settings with externalities, I cannot guarantee uniqueness of an equilibrium with exogenous productivity. As a heuristic, it is useful to focus on the mode choice problem that determines equilibrium driver shares and road traffic for a given labour market equilibrium. Abstracting from hours of the day costs no generality and gives

$$\begin{aligned} \ln \left(\frac{\pi_{c|ij}}{1 - \pi_{c|ij}} \right) &= \eta \alpha_{ic} + \eta \alpha_{jc} + \eta d_c(ij) - \eta \ln(toll_{ij}) + \ln(\frac{\gamma^\theta}{\Phi}) + \eta \ln(d_{ijo}) \\ &\quad - \eta \kappa \ln \left(\sum_\ell 1\{\ell \in R_{ij}\} \left[\varphi CCE_\ell + e_\ell + \sum_{r,s,g} 1\{\ell \in R_{rs}\} \pi_g \Omega (w_{sg} B_{rg} Q_r^{\beta-1})^\theta \pi_{c|rs} (1 - \pi_{c|rs})^{-\frac{\theta}{\eta}} d_{rs}^{-\theta} \right] \right) \end{aligned} \quad (33)$$

assuming interior solutions on traffic. Equation 33 defines location pair ij car share demand where commute costs are prices and wages and rents are demand shifters. I now show that standard homogeneity and gross substitution conditions cannot guarantee a unique vector of mode shares satisfying equation 32, leaving the possibility that multiple vectors of mode shares and traffic externalities can follow from a labour market equilibrium.

To see the problem at hand, it is useful to re-write equation 32 as an equilibrium system $f : \mathbb{R}_{++}^{N^2} \rightarrow \mathbb{R}^{N^2}$ with elements $f_{ij}(\{\pi_{c|rs}\}) = \ln\left(\frac{\pi_{c|ij}}{1-\pi_{c|ij}}\right) - \tilde{\Lambda}_{ij}(\{\pi_{c|rs}\})$ where $\tilde{\Lambda}_{ij}(\{\pi_{c|rs}\})$ is the right hand side of equation 32 and an equilibrium requires $f(\{\pi_{c|rs}\}) = 0$. Theorem 2 of Allen et al. (2015) establishes uniqueness of $\{\pi_{c|rs}\}$ under a homogeneity condition and the gross substitutes condition $\frac{\partial f_{ij}}{\partial \pi_{c|rs}} > 0 \forall r, s \neq i, j$.

I cannot rule out multiple equilibria since $\ln\left(\frac{\pi_{c|ij}}{1-\pi_{c|ij}}\right)$ is not a homogeneous function of $\pi_{c|ij}$. However, it is interesting to note that the equilibrium system exhibits (weak) gross substitution. Differentiating equation 32 with respect to drivers' share on an arbitrary commute gives

$$\begin{aligned} \frac{\partial \ddot{\Lambda}_{ij}}{\partial \pi_{c|rs}} &= -\frac{\kappa}{\ddot{\Phi}_{ij}} \frac{1\{\ell \in R_{ij}\}1\{\ell \in R_{rs}\}(w_{sg}B_{rg}Q_r^{\beta-1})^\theta}{d_{rso}^\theta(1-\pi_{c|rs})^{\frac{\theta+n}{\eta}}} [\theta\pi_{c|rs} + \eta(1-\pi_{c|rs})] \\ \text{s.t. } \ddot{\Phi}_{ij} &= \sum_\ell 1\{\ell \in R_{ij}\} \left[\varphi CCE_\ell + e_\ell + \sum_{r,s,g} 1\{\ell \in R_{rs}\} \Omega \pi_g (w_{sg}B_{rg}Q_r^{\beta-1})^\theta \pi_{c|rs} (1-\pi_{c|rs})^{\frac{-\theta}{\eta}} d_{rso}^{-\theta} \right] \end{aligned}$$

which is always negative when routes R_{ij} and R_{rs} intersect.⁴³ It follows that $\frac{\partial f_{ij}}{\partial \pi_{c|rs}} = -\frac{\partial \ddot{\Lambda}_{ij}}{\partial \pi_{c|rs}} \geq 0 \forall r, s \neq i, j$, which holds strictly when $R_{ij} \cap R_{rs} \neq \emptyset$.

In conclusion, there can be more than one vector of mode shares consistent with a given set of wages and rents. It follows that these wages and rents can generate more than one vector of endogenous commute costs.

A.2.4 Counterfactual Procedure

I use hat algebra to compute responses to a counterfactual road tolling policy defined by tolls $toll'_{ij}$ and charge exposure CCE_ℓ^{th} (Dekle et al., 2007). The model admits multiple equilibria and this procedure gives counterfactual outcomes that are natural progressions from the observed equilibrium.

⁴³ $\frac{\partial \ddot{\Lambda}_{ij}}{\partial \pi_{c|rs}}$ must be weakly negative because $\ddot{\Phi}_{ij} > 0$, $\theta > 0$, $\eta > 0$, and $\pi_{c|rs} \in (0, 1)$ by definition.

For any endogenous variable x with counterfactual value x' , I solve for relative changes $\hat{x} = x'/x$. In a closed city, counterfactual wages, rents, mode shares, and transportation costs are determined by following system of equations.

$$\hat{A}_j = \begin{cases} 1 & \text{with exogenous productivity} \\ \left(\frac{\sum_s e^{-\delta d(s,j)} \frac{1}{Area_s} (\sum_g a_{jg} (\sum_i \pi_{ijg} \hat{\pi}_{ijg})^\rho)^{\frac{1}{\rho}}}{\sum_s e^{-\delta d(s,j)} \frac{1}{Area_s} (\sum_g a_{jg} (\sum_i \pi_{ijg})^\rho)^{\frac{1}{\rho}}} \right)^\lambda & \text{with agglomeration} \end{cases} \quad (34)$$

$$\hat{w}_{jg} = \hat{A}_j \left(\frac{\sum_i \pi_{ijg} \hat{\pi}_{ijg}}{\sum_i \pi_{ijg}} \right)^{-\frac{1}{\sigma}} \left(\frac{\sum_{g'} a_{jg'} (\sum_i \pi_{ijg'} \hat{\pi}_{ijg'})^\rho}{\sum_{g'} a_{jg'} (\sum_i \pi_{ijg'})^\rho} \right)^{\frac{\alpha-\rho}{\rho}} \quad (35)$$

$$\hat{Q}_i = \frac{\sum_j \sum_g w_{jg} \hat{w}_{jg} \pi_{ijg} \hat{\pi}_{ijg}}{\sum_j \sum_g w_{jg} \pi_{ijg}} \quad (36)$$

$$\hat{\pi}_{ijg} = \pi_g \frac{\hat{w}_{jg}^\theta \hat{Q}_i^{\theta(\beta-1)} \hat{d}_{ij}^{-\theta}}{\sum_{r,s} \pi_{rs|g} \hat{w}_{jg}^\theta \hat{Q}_r^{\theta(\beta-1)} \hat{d}_{rs}^{-\theta}} \quad (37)$$

$$\hat{\pi}_{ij} = \frac{\sum_g \pi_{ijg} \hat{\pi}_{ijg}}{\sum_g \pi_{ijg}} \quad (38)$$

$$\widehat{traffic}_{ij} = \frac{\sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} \max(\varphi^h CCE_\ell'^h + e_\ell^h + \Omega \sum_{r,s} 1\{\ell \in R_{rs}^h\} \pi_{c|rs} \pi_{rs} \hat{\pi}_{c|rs} \hat{\pi}_{rs}, 0)}{\sum_h \sum_\ell 1\{\ell \in R_{ij}^h\} traffic_\ell^h} \quad (39)$$

$$\hat{d}_{ijc} = \widehat{toll}_{ij} \left(\widehat{traffic}_{ij} \right)^\kappa \quad (40)$$

$$\hat{d}_{ijc}^{-\eta} = \frac{\hat{\pi}_{c|ij}(1 - \pi_{c|ij})}{1 - \pi_{c|ij}\hat{\pi}_{c|ij}} \implies \hat{\pi}_{c|ij} = \frac{\hat{d}_{ijc}^{-\eta}}{1 + \pi_{c|ij}(1 - \hat{d}_{ijc}^{-\eta})} \quad (41)$$

$$\hat{d}_{ij} = \frac{\gamma^{-1}}{\bar{d}_{ij}} \left[1 + \left(\frac{\pi_{c|ij}}{1 - \pi_{c|ij}} \right) \hat{d}_{ijc}^{-\eta} \right]^{-\frac{1}{\eta}} d_{ijo} = \left[1 + \pi_{c|ij}(\hat{d}_{ijc}^{-\eta} - 1) \right]^{-\frac{1}{\eta}} \quad (42)$$

Relative commuter welfare, aggregate floorspace rent, and traffic are

$$\widehat{E(u)} = \frac{\sum_g \pi_g \Phi_g^{\frac{1}{\theta}} \left[\sum_i \sum_j \pi_{ij|g} \hat{w}_{jg}^\theta \hat{Q}_i^{\theta(\beta-1)} \hat{d}_{ij}^{-\theta} \right]^{\frac{1}{\theta}}}{\sum_{g'} \pi_{g'} \Phi_{g'}^{\frac{1}{\theta}}} \quad (43)$$

$$\hat{Q} = \frac{\sum_j F_j \hat{F}_j + \sum_i H_i Q_i \hat{Q}_i}{\sum_j F_j + \sum_i H_i Q_i} = \frac{\sum_i \sum_j \sum_g w_{jg} \hat{w}_{jg} \pi_{ijg} \hat{\pi}_{ijg}}{\sum_i \sum_j \sum_g w_{jg} \pi_{ijg}} \quad (44)$$

$$\widehat{traffic} = \frac{\sum_h \sum_\ell \left[\varphi^h CCE_\ell^h + e_\ell^h + \Omega \sum_{r,s} 1\{\ell \in R_{rs}^h\} \pi_{c|rs} \pi_{rs} \hat{\pi}_{c|rs} \hat{\pi}_{rs} \right]}{\sum_h \sum_\ell traffic_\ell^h} \quad (45)$$

Given the initial shock to commute costs, \hat{d}_{ij}^0 , the following iterative procedure identifies the equilibrium prices and allocations that are closest to the initial equilibrium:

Guess starting values of \hat{w}_{jg}^0 , \hat{Q}_i^0 , $\hat{\pi}_{c|ij}^0$, and initiate commute flows $\hat{\pi}_{ijg}^0$ by evaluating equation 37 given observed flows π_{ijg} and initial shocks \hat{d}_{ij}^0 .

Main loop: for iteration $t > 0$:

1. **Labour market loop:** for iteration $k > 0$:

- (a) Evaluate the relevant case of equation 34 at $\hat{\pi}_{ijg}^{k-1}$ to define $\hat{A}_j(\hat{\pi}_{ijg}^{k-1})$.
 - (b) Evaluate equation 35 at $\hat{\pi}_{ijg}^{k-1}$ and \hat{w}_{jg}^{k-1} to define $\hat{w}_{jg}(\hat{\pi}_{ijg}^{k-1})$.
 - (c) Evaluate equation 36 at $\hat{\pi}_{ijg}^{k-1}$ and $\hat{w}_{jg}(\hat{\pi}_{ijg}^{k-1})$ to define $\hat{Q}_i(\hat{\pi}_{ijg}^{k-1})$.
 - (d) Evaluate equation 37 at \hat{d}_{ij}^{t-1} , $\hat{Q}_i(\hat{\pi}_{ijg}^{k-1})$, and $\hat{w}_{jg}(\hat{\pi}_{ijg}^{k-1})$ to define $\hat{\pi}_{ijg}(\hat{d}_{ij}^{t-1})$.
 - (e) Update endogenous variables $\hat{x}^k = \zeta \hat{x}^{k-1} + (1-\zeta) \hat{x}(\hat{\pi}_{ijg}^{k-1})$ for each $x \in \{\hat{w}_{jg}, \hat{Q}_i, \hat{\pi}_{ijg}\}$ using a weight $\zeta \in [0, 1]$.
 - (f) If $|\hat{x}^k - \hat{x}^{k-1}|$ is sufficiently small, stop and define \hat{w}_{jg}^t , \hat{Q}_j^t , and temporary commute flows $\hat{\pi}_{ijg}(\hat{d}_{ij}^{t-1})$ as each variable's k th values.
2. Evaluate equation 38 for commute flows $\hat{\pi}_{ij}(\hat{d}_{ij}^{t-1})$.
3. **Road traffic loop:** for iteration $k > 0$:
- (a) Evaluate equations 39 and 40 at $\hat{\pi}_{ij}(\hat{d}_{ij}^{t-1})$ and $\hat{\pi}_{c|ij}^{k-1}$ to update driving costs \hat{d}_{ijc}^k .
 - (b) Evaluate equation 41 at \hat{d}_{ijc}^k to define $\hat{\pi}_{c|ij}(\hat{d}_{ijc}^k)$, impose the constraint $\pi_{c|ij} \hat{\pi}_{c|ij}(\hat{d}_{ijc}^k) \in [0, 1]$, and update mode shares $\hat{\pi}_{c|ij}^k = \zeta \hat{\pi}_{c|ij}^{k-1} + (1 - \zeta) \hat{\pi}_{c|ij}(\hat{d}_{ijc}^k)$.
 - (c) If $|\hat{\pi}_{c|ij}^k - \hat{\pi}_{c|ij}^{k-1}|$ is sufficiently small, stop.
4. Evaluate equation 42 at \hat{d}_{ijc}^k to update commute cost indices \hat{d}_{ij}^t .
5. Evaluate equation 37 at \hat{d}_{ij}^t , \hat{Q}_j^t , and \hat{w}_{jg}^t to define $\hat{\pi}_{ijg}^{temp}$ and update $\hat{\pi}_{ijg}^t = \zeta \hat{\pi}_{ijg}^{temp} + (1 - \zeta) \hat{\pi}_{ijg}^{t-1}$.
6. If $|\hat{\pi}_{ijg}^t - \hat{\pi}_{ijg}^{t-1}|$ is sufficiently small, stop. Otherwise, update $t = t + 1$ and repeat main loop.

A.2.5 Quantifying congestion charge removal

To simulate removing the CCZ, I increase traffic according to routes' Congestion Charge Exposure, remove all road tolls, and confiscate recycled toll revenue. The initial shock increases each road link's traffic to $\text{traffic}'^h_\ell = \text{traffic}^h_\ell - \varphi^h CCE^h_\ell$ in both morning and evening rush hours. Exogenous routes for each rush hour aggregate shocks to MSOA pairs so that $\text{traffic}'^h_{ij} = \sum_\ell 1\{\ell \in R_{ij}\} \text{traffic}'^h_\ell$. I then average across morning and evening rush hours to compute counterfactual traffic $\text{traffic}'_{ij} = \frac{1}{2}(\text{traffic}'^{morn}_{ij} + \text{traffic}'^{eve}_{ij})$.

I scale daily wages w_{jg} to have cross-MSOA geometric means of £120 for low-skill and £160 for high-skill workers and assume they include initial per-commuter toll revenue $r = \frac{8}{w} \sum_{ij} \pi_{c|ij} CCZ_{ij} \sum_g \pi_{ijg} w_{jg}$ collected from commutes into the CCZ where $\bar{w} = \sum_j \sum_g \pi_g w_{jg}$ is the mean calibrated wage.⁴⁴ I set $\Omega = 1,943,475$ to match the sum of commute flows in the 2011 analysis sample and parametrize $toll_{ij} = \frac{1}{1 - \frac{8CCZ_{ij}}{\bar{w} + r}}$ so that $\frac{w_{jg} + r}{toll_{ij}} \approx w_{jg} + r - 8CCZ_{ij}$ which holds exactly if w_{jg} is the average wage.⁴⁵

Removing the CCZ initiates commute cost shocks $\hat{d}_{ij}^0 = \frac{1}{toll_{ij}} \left(\frac{\text{traffic}'_{ij}}{\text{traffic}_{ij}} \right)^\kappa$ and wage shocks $\hat{w}_{jg}^0 = \frac{w_{jg} - r}{w_{jg}}$. With exogenous productivity, rents and mode shares begin at their observed 2011 values and commute flows begin at values consistent with commute cost shocks, wages, and rents. With endogenous productivity, all prices and allocations begin at values constituting an equilibrium with exogenous productivity.

A.2.6 Elastic housing supply

The baseline model fixes the amount of floorspace available in each MSOA. Alternatively, housing supply can endogenously increase in rents and developable land such that $H_i^s =$

⁴⁴Wages and tolls reflect their values in early 2011 (Office for National Statistics, 2012).

⁴⁵The analytical sample contains fewer commuters than raw data because it excludes origin destination pairs that are unmatched to the DfT traffic monitoring network or are outliers in pairwise traffic or CCE as well as dropping counts of individuals working from home or within their residence MSOA.

$S_i Q_i^\phi$ where $\phi > 0$ is the housing supply elasticity.⁴⁶ Then, equilibrium housing rent is $Q_i^* = \left(\gamma^\theta \Omega \frac{B_i^\theta}{S_i} (1 - \beta) \sum_g \pi_g \sum_j w_{jg}^{1+\theta} \bar{d}_{ij}^{-1-\theta} \right)^{\frac{1}{1+\theta(1-\beta)+\phi}}$. To a first order, endogenous housing supply will decrease the rate at which commute costs and wages capitalize into housing costs.

A.2.7 Human capital spillovers

The baseline model assumes that all agglomeration externalities occur through a constant elasticity of substitution aggregate of all labour types. Alternatively, I could follow Lee (2019) and introduce localized human capital spillovers so that total productivity in work-place j is

$$A_j = a_j \Upsilon_{jh}^{\lambda_h} \Upsilon_{jl}^{\lambda_l} \text{ s.t. } \Upsilon_{jg} = \sum_s e^{-\delta_g d(s,j)} \frac{L_{sg}}{Area_s} \text{ for } g \in \{h, l\}$$

where λ_g are human capital externality elasticities. Using American data, Lee (2019) finds evidence of rapidly decaying externalities from both college and non-college employment (large δ_g) and that non-college workers create negative externalities ($\lambda_h > 0$ and $\lambda_l < 0$).

I consider human capital spillovers by repeating the counterfactual policy experiment in the main text with $\delta_h = \delta_l = 4.2$ and a range of elasticities $(\lambda_h, \lambda_l) \in [-0.1, 0.1] \times [-0.1, 0.1]$. Table A.16 presents the range of mean utility effects, which are similar to the main text and increase in the strength of agglomeration forces. Taking $(\lambda_h, \lambda_l) = (0.7, -0.7)$ following Lee (2019) implies London's Congestion Charge increases mean equilibrium utility by 0.023 percent.

⁴⁶Severen (2019) derives this housing supply function assuming land is supplied iso-elasticly to competitive housing developers.

A.2.8 Shipped inputs

Firms may use roads for deliveries, creating an unobserved mechanism through which traffic affects labour demand. I consider this possibility by assuming productivity decreases in travel costs to an exogenously located supplier so that $A_j = a_j \overline{\text{traffic}}_j^\iota$ where $\overline{\text{traffic}}_j$ is the amount of traffic incurred between workplace j and its supplier and $\iota < 0$ is a simplification of firms' elasticity of intermediate input costs with respect to traffic. This is isomorphic to traffic impeding firms' access to customers visiting from a fixed origin.

In this case, labour market clearing requires

$$w_{jg} = \alpha a_{jg} a_j \overline{\text{traffic}}_j^\iota \left(\Omega \sum_i \pi_{ijg} \right)^{-\frac{1}{\sigma}} \left(\sum_{g'} a_{jg'} \left(\Omega \sum_i \pi_{ijg'} \right)^\rho \right)^{\frac{\alpha-\rho}{\rho}}.$$

Recall that π_{ijg} decreases in traffic ceteris paribus, so the standard labour market channel increases wages to compensate commuters for driving in traffic. It follows that counterfactuals without intermediate input access overstate traffic's positive effect on wages to the extent that $\overline{\text{traffic}}_j$ is correlated with firms' labour market access $\sum_i (B_{ig} Q_i^{\beta-1})^\theta \bar{d}_{ij}^{-\theta}$ for each group.

A.3 Appendix figures

Figure A.1: Matching pairwise commutes to traffic counts

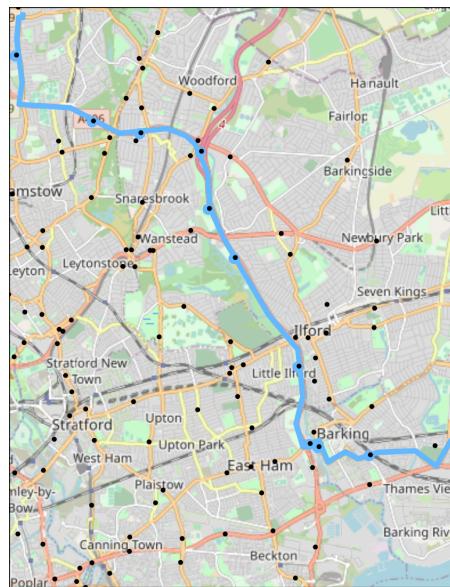
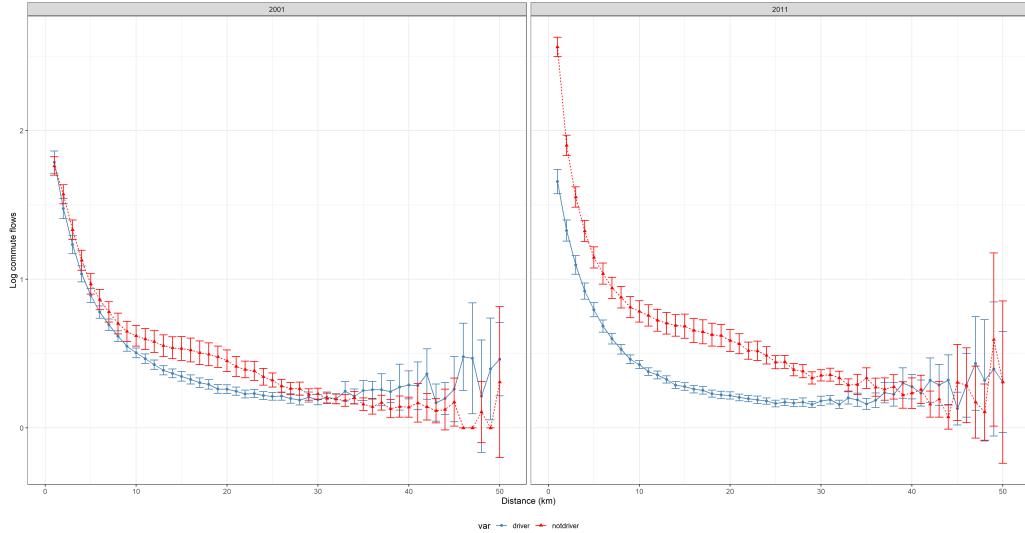


Figure A.2: MSOA employment per hectare in 2001

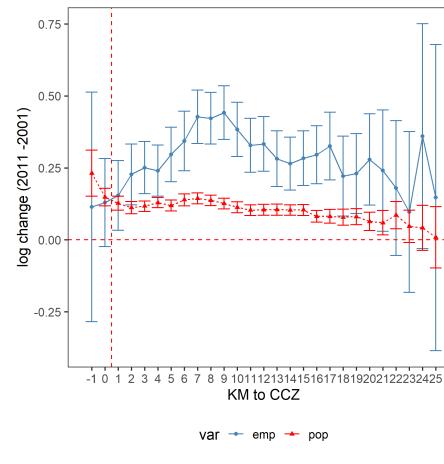


Figure A.3: Commute flows by trip distance



Points are mean MSOA pair commuter counts with 1 km bins of euclidean distance between residence and workplace. Error bars denote heteroskedasticity robust 95 percent confidence intervals.

Figure A.4: Population and employment growth gradients



Points denote mean log-change in OA population/employment growth and error bars denote heteroskedasticity robust 95 percent confidence intervals by OA centroid kilometres from the CCZ boundary.

Figure A.5: Central London's Congestion Charge Zone and Western Extension

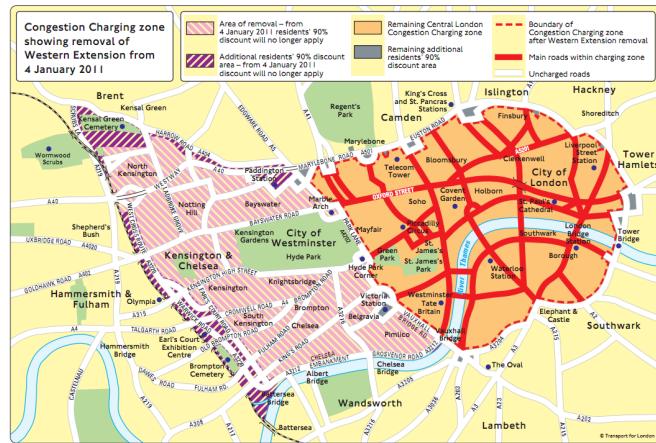
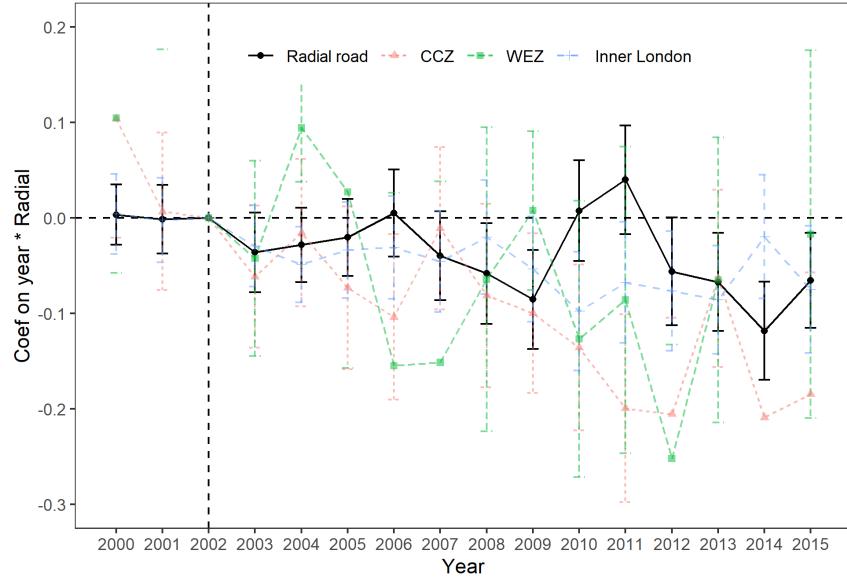


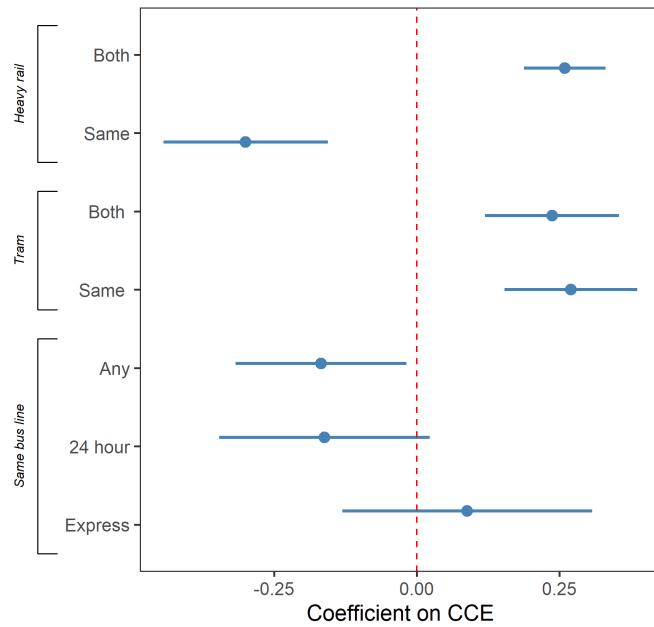
Figure produced by Transport for London and adapted from Tang (2018).

Figure A.6: London's congestion charge and radial road traffic



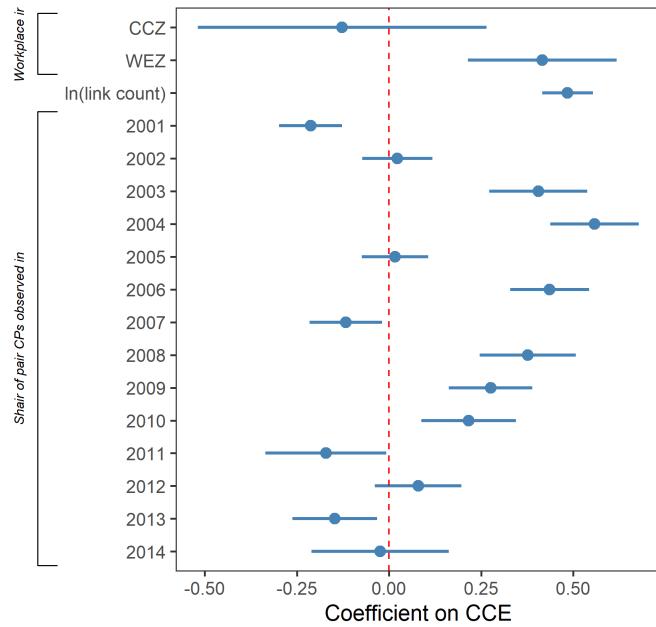
Points are OLS estimates of β_{τ}^{zone} the link-year regression $\ln(Traffic_{\ell t}) = \sum_{\tau=2000, \tau \neq 2002}^{2015} \left(D_{\tau}^t \beta_{\tau}^r Radiale + D_{\tau}^t \beta_{\tau}^c CCZ_{\ell t} + D_{\tau}^t \beta_{\tau}^w WEZ_{\ell t} + D_{\tau}^t \beta_{\tau}^l InnerLondon_{\ell t} + D_{\tau}^t f_{\tau}(late_{\ell}, lone_{\ell}) \right) + \alpha_t + \alpha_{\ell} + e_{\ell t}$, normalized to zero in 2002, $Traffic_{\ell t}$ is the mean of morning and evening rush-hour traffic volumes in that link-year, and error bars denote link-clustered 95 percent confidence intervals in zone-years where they are contained between $[-.3, .2]$.

Figure A.7: Conditional correlations between CCE and transit characteristics.



Points are estimates of β from independent weighted OLS regressions of the form $CCE_{ij} = \beta_i + \beta_j + \beta_{d(ij)} + \beta x_{ij} + u_{ij}$ where x_{ij} is a public transit characteristic for commute ij . “Both” indicates pairs with origin and destination both within 1500 metres of a station and “same” indicates pairs with origin and destination within 1500 metres of stations on the same transit line. Bars are 95 percent confidence intervals 2 way clustered by residence and workplace.

Figure A.8: Conditional correlations between CCE and route characteristics (cont.).



Points are estimates of β from independent weighted OLS regressions of the form $CCE_{ij} = \beta_i + \beta_j + \beta_{d(ij)} + \beta x_{ij} + u_{ij}$. Bars are 95 percent confidence intervals 2 way clustered by residence and workplace.

Figure A.9: Percent change in floorspace rent caused by the congestion charge.

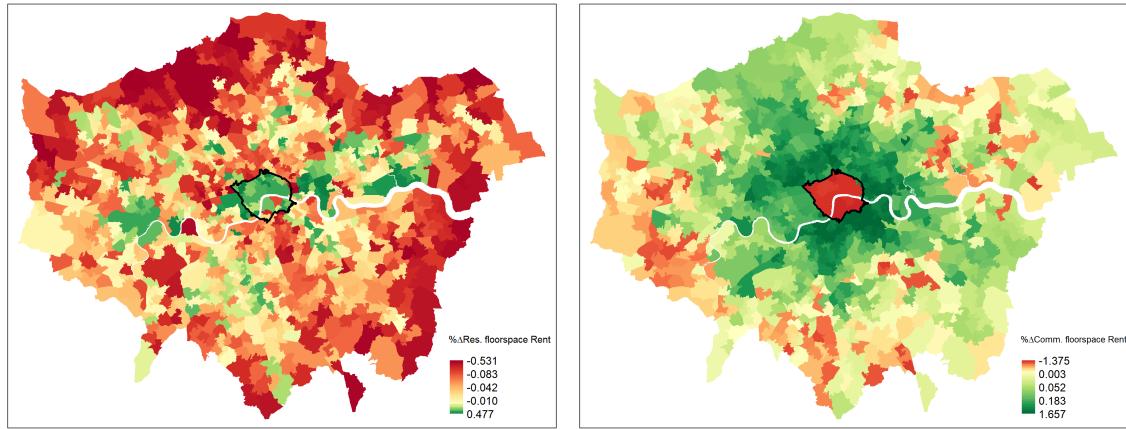


Figure A.10: Validating routing assumptions

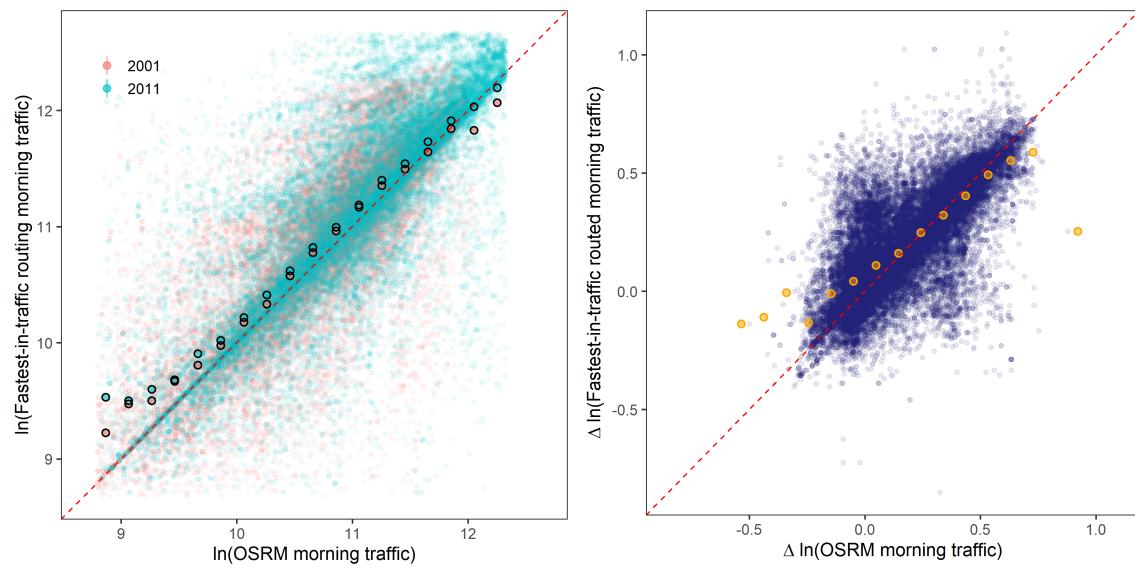
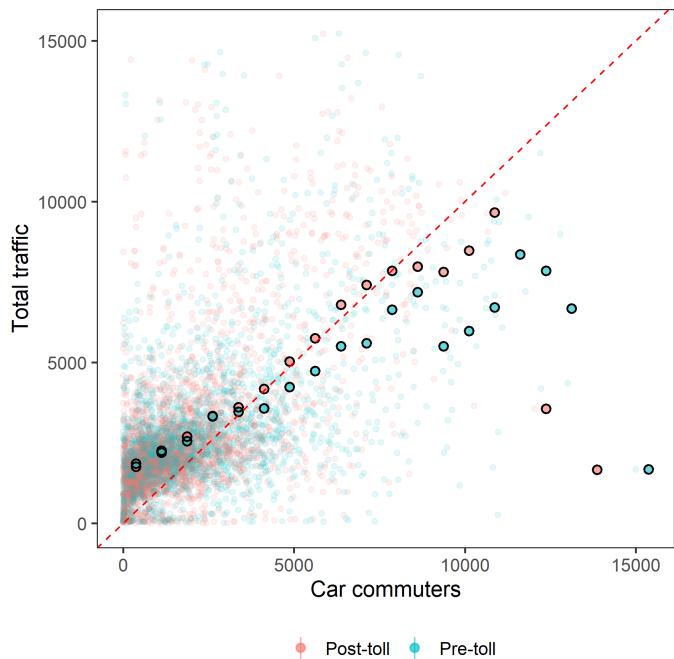
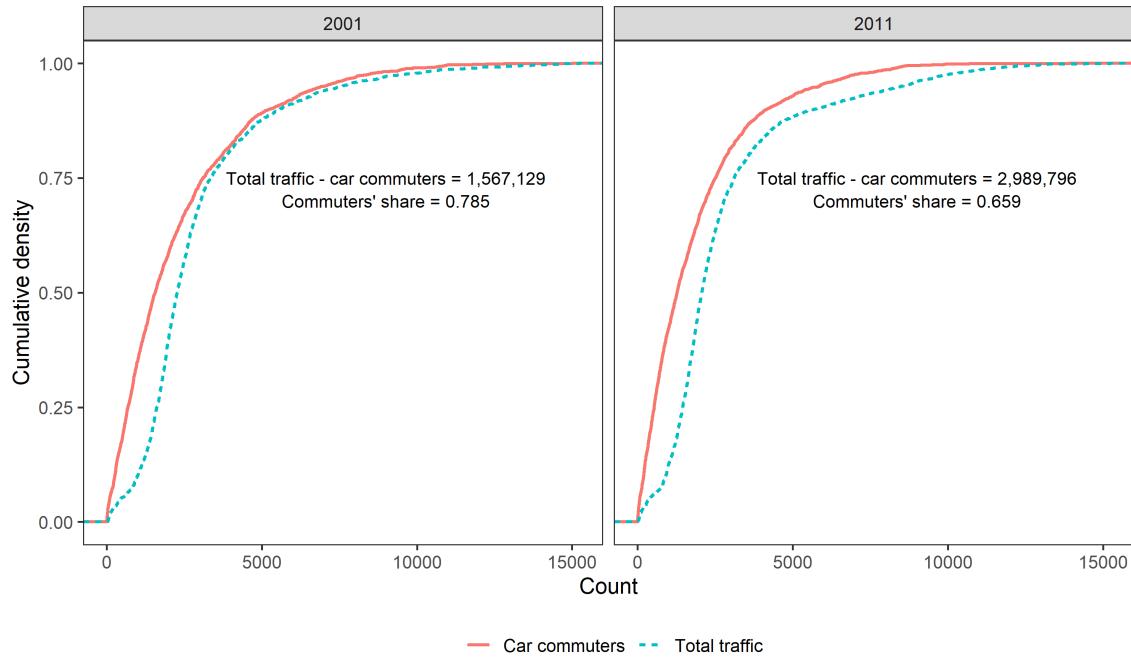


Figure A.11: Road link traffic and car commuters



Small points are road-link daily mean rush hour traffic and estimated counts of daily car commuters, large points are binned means, and red dashes denote the 45 degree line. Small point transparency is such that ten overlapping points appear opaque.

Figure A.12: Road link traffic and car commuters (cont.)



Solid lines are empirical density functions of estimated daily car commuters and dashed lines are those of road-link daily main rush hour traffic.

A.4 Appendix tables

Table A.1: Additional summary statistics for MSOA pairs in the main analysis sample

	2001		2011		Full Sample	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
Commute flows:						
Drivers	0.761	3.826	0.637	3.19	0.699	3.523
Other modes	1.227	8.448	1.587	10.174	1.407	9.353
ln(Commuters)	0.379	0.844	0.416	0.88	0.397	0.862
ln(Drivers)	-0.035	0.662	-0.125	0.723	-0.08	0.695
-ln(Other modes)						
Pairwise traffic:						
Morning	81,148.32	56,384.97	106,171.30	82,321.41	93,659.82	71,655.95
Evening	86,735.20	59,597.29	112,385.90	87,644.30	99,560.56	76,033.98
ln(Traffic)	11.045	0.876	11.245	0.958	11.145	0.924

Table A.2: Heterogeneous effects of Congestion Charge Exposure on traffic

	<i>Interaction variable:</i>			
	Distance	CCE	Both Inner London	Both heavy rail < 1,500m
CCE × Post × Q1 or 1{ $d = 1$ }	-0.027*** (0.002)	-0.186*** (0.055)	-0.024*** (0.002)	-0.025*** (0.002)
CCE × Post × Q2 or 1{ $d = 0$ }	-0.027*** (0.003)	-0.026*** (0.008)	-0.028*** (0.002)	-0.032*** (0.002)
CCE × Post × Q3	-0.019*** (0.003)	-0.027*** (0.002)		
Observations	1,709,220	1,709,220	1,709,220	1,709,220

Each column presents results of an OLS regression using 2001 total flows plus one as weights and interacting CCE with indicators describing a fixed observable. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Column 1 divides pairs into distance terciles using DfT monitored road length. Column 2 divides pairs into CCE terciles. Column 3 interacts CCE with an indicator for pairs within Inner London and column 4 interacts CCE with an indicator for both origin and destination centroids within 1,500m of a heavy rail station. For terciles, Q1 is the lowest third and Q3 is the highest, for indicators, 1{ $d = 1$ } indicates when the condition is true. All regressions include full sets of fixed effects and controls described in section 4 footnote 14 and column three adds a both Inner London by year interaction. CCE is scaled to have a standard deviation of one.

Table A.3: Congestion Charge Exposure's effect on road link traffic

	<i>Dependent variable:</i>			
	<i>Traffic</i>	<i>Traffic^{morn}</i>	<i>Traffic^{eve}</i>	
	(1)	(2)	(3)	(4)
CCE × Post	−109.755*** (23.255)	−65.527** (30.266)		
CCE ^{morn} × Post			−65.319** (32.242)	
CCE ^{eve} × Post				−102.659*** (31.623)
Controls	No	Yes	Yes	Yes
Observations	5,698	5,698	5,698	5,698
Dep. var. mean	2819	2819	2739	2898
Dep. var. std. dev.	2289	2289	2282	2414

Note:

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

Standard errors clustered by road in parenthesis. Controls are dummies indicating whether a road link was observed in each year from 2000 to 2015, is at least partially within the WEZ, within the CCZ, and interactions of CCZ inclusion with log-distance to CCZ boundary, all interacted with a post period dummy.

Table A.4: Robustness of reduced form evidence

Dep. Var	ln(Commuters)	ln($\frac{\text{Drivers}}{\text{Non-drivers}}$)	ln(Traffic)
Indep. Var	ln(Traffic)	ln(Traffic)	CCE \times Post
Drop CCZ	-0.977*** (0.221)	-1.950*** (0.246)	-0.026*** (0.002)
Drop WEZ	-0.991*** (0.193)	-1.339*** (0.381)	-0.026*** (0.002)
Drop WEZ and CCZ	-1.007*** (0.228)	-2.177*** (0.261)	-0.026*** (0.002)
New rail control	-0.922*** (0.188)	-1.202*** (0.351)	-0.027*** (0.002)
Dist to CBD interaction	-0.726*** (0.190)	-1.439*** (0.332)	-0.026*** (0.002)
Excl. IV	CCE \times Post	CCE \times Post	None

Note: *p<0.1; **p<0.05; ***p<0.01

Each cell presents results of a separate TSLS regression using 2001 total flows plus one as weights and all regressions include full sets of fixed effects and controls. Dependent variables add one to commute counts before taking logarithms and the final column presents first stage estimates. The first three rows iteratively exclude MSOAs in the CCZ, WEZ, or both, the fourth row controls for rail transit built after 2001, and the the final row controls for a post-toll dummy interacted with the interaction of origin and destination distances to Charing Cross. Standard errors in parenthesis are two-way clustered at the residence and workplace levels.

Table A.5: Alternative CCE definitions

Dep. Var	ln(Commuters)	$\ln\left(\frac{\text{Drivers}}{\text{Non-drivers}}\right)$	ln(Traffic)
Indep. Var	ln(Traffic)	ln(Traffic)	Standardized CCE Alternative \times Post
Unweighted tolls	-0.330 (0.280)	-3.996*** (0.572)	-0.018*** (0.003)
Include WEZ	-0.801*** (0.226)	-1.664*** (0.398)	-0.021*** (0.002)
Per road link	-0.946*** (0.337)	-2.345*** (0.535)	-0.015*** (0.002)
Excl. IV	CCE \times Post	CCE \times Post	None

Note: *p<0.1; **p<0.05; ***p<0.01

Each cell presents results of a separate TSLS regression using 2001 total flows plus one as weights and all regressions include full sets of fixed effects and controls. Dependent variables add one to commute counts before taking logarithms, the final column presents first stage estimates, the final row computes both traffic and CCE on a per-road link basis, and CCE is normalized to have a cross-sectional standard deviation of one throughout. Standard errors in parenthesis are two-way clustered at the residence and workplace levels.

Table A.6: Congestion Charge Exposure and transit expansion.

<i>Dependent variable:</i>			
Congestion Charge Exposure			
	(1)	(2)	(3)
New rail transit connection	0.021** (0.010)	−0.005* (0.003)	−0.003 (0.003)
Dist FE	No	Yes	Yes
Pow/Por FE	None	Both	Both
Controls	No	No	Yes
Observations	854,610	854,610	854,610
R ²	0.004	0.673	0.710

Note: *p<0.1; **p<0.05; ***p<0.01

Each column presents results of an OLS regression of the form $CCE_{ij} = \alpha NewConnection_{ij} + \alpha_i + \alpha_j + \alpha_{d(ij)} + X'_{ij}\beta + e_{ij}$ using 2001 total flows plus one as weights. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14. CCE is scaled to have a standard deviation of one.

Table A.7: Traffic's separate effects on commuting by mode

<i>Dependent variable:</i>			
ln(Drivers) ln(Non-drivers)			
	(1)	(2)	
ln(Traffic)	−1.338*** (0.292)	−0.130 (0.161)	
Observations	1,709,220	1,709,220	

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.8: Traffic's effect on commuting by trip distance

	<i>Dependent variable:</i>	
	ln(Commuters)	ln($\frac{\text{Drivers}}{\text{Non-drivers}}$)
	(1)	(2)
ln(Traffic) × (dist < 15.1 km)	-0.749*** (0.165)	-1.171*** (0.307)
ln(Traffic) × (15.1 km ≤ dist < 26.4 km)	-0.748*** (0.163)	-1.126*** (0.306)
ln(Traffic) × (26.4 km ≥ dist)	-0.752*** (0.162)	-1.106*** (0.307)
Dist × Post FE	Yes	Yes
Pow/Por × Post FE	Both	Both
First Stage F-stat (dist < 15.1 km)	190.41	190.41
First Stage F-stat (15.1 km ≤ dist < 26.4 km)	312.85	312.85
First Stage F-stat (26.4 km < dist)	363.24	363.24
Observations	1,709,220	1,709,220

Note:

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

Each column presents results of a TSLS regression using 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. Dependent variables add one to flow variables before taking logarithms and interaction terms are dummies separating the sample into equal sized groups by euclidean distance between workplace and residence. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14. Excluded instruments are CCE-distance tercile interactions and partial first stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table A.9: Traffic's effect on commuting by Inner London

	<i>Dependent variable:</i>	
	ln(Commuters)	ln($\frac{\text{Drivers}}{\text{Non-drivers}}$)
	(1)	(2)
ln(Traffic) × (Both Inner London)	-2.702*** (0.437)	0.282 (0.736)
ln(Traffic) × (1 - Both Inner London)	-0.761*** (0.196)	-1.569*** (0.334)
Dist × Post FE	Yes	Yes
Pow/Por × Post FE	Both	Both
First Stage F-stat (Both Inner London)	23.75	23.75
First Stage F-stat (Not both Inner London)	61.83	61.83
Observations	1,709,220	1,709,220

Note:

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

Each column presents results of a TSLS regression using 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. Dependent variables add one to flow variables before taking logarithms and the interaction term specifies MSOA pairs where both workplace and residence MSOAs are in the Inner London boroughs defined in appendix A.1.1. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14 and add Inner London-post-toll interactions. Excluded instruments are CCE-Inner London interactions and partial first stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table A.10: Traffic's effect on commuting by driver's share

<i>Interaction variable:</i>			
ln(Traffic) ×	$\pi_{c ij} \leq \frac{1}{3}$	$\frac{1}{3} < \pi_{c ij} \leq \frac{2}{3}$	$\frac{2}{3} < \pi_{c ij}$
ln(Number commuting)	-1.630*** (0.400)	-6.428** (2.743)	-3.322*** (0.848)
First Stage F-stat	148.83	2.94	19.43

Note:

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

Columns are coefficient estimates from a single TSLS regression using 2001 total flows plus one as weights with a full set of controls and fixed effects. The interaction variables are bins of 2001 driver's share and the regression adds one to total flows before taking logarithms. Standard errors in parenthesis are two-way clustered at the residence and workplace levels. Controls are described in section 4 footnote 14 and dummies for the three relevant $\pi_{c|ij}$ bands. Excluded instruments are CCE- $\pi_{c|ij}$ band interactions and partial first stage F-statistics are computed as in Sanderson and Windmeijer (2016) using two-way residence and workplace clustered variance covariance matrices.

Table A.11: Share of London Region employment captured by commute flows

	Drivers	Other modes	Total
2001	0.65	0.77	0.72
2011	0.60	0.81	0.75

Table A.12: Validating routes using pairwise traffic

	<i>Fastest-in-traffic routed:</i>					
	ln(Morning traffic)		$\Delta \ln(\text{Morning traffic})$			
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Morning traffic) <i>OSRM routed</i>	0.920*** (0.011)					
Post-toll		0.207*** (0.050)				
ln(Morning traffic) \times Pre-toll, <i>OSRM routed</i>		0.926*** (0.012)				
ln(Morning traffic) \times Post-toll, <i>OSRM routed</i>		0.910*** (0.012)				
$\Delta \ln(\text{Morning traffic})$ <i>OSRM routed</i>			0.763*** (0.014)	0.763*** (0.014)	0.570*** (0.017)	0.529*** (0.027)
Constant	0.972*** (0.129)	0.898*** (0.128)	0.070*** (0.005)	0.070*** (0.005)		
Observations	73,130	73,130	73,130	73,130	73,130	73,130
Por/Pow/Dist FEs	No	No	No	No	Yes	Yes
Weights	None	None	None	None	None	$\pi_{ij2001} + 1$
Projected R^2	0.689	0.690	0.566	0.566	0.690	0.697

Note:

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

Table A.13: Validating routes using monitored road length (cont)

	<i>Fastest-in-traffic routed:</i>					
	ln(Monitorred metres)		$\Delta \ln(\text{Monitorred metres})$			
	(1)	(2)	(3)	(4)	(5)	(6)
ln(monitored metres) <i>OSRM routed</i>	0.922*** (0.012)					
Post-toll		-0.038 (0.053)				
ln(monitored metres) \times Pre-toll, <i>OSRM routed</i>		0.911*** (0.012)				
ln(monitored metres) \times Post-toll, <i>OSRM routed</i>		0.918*** (0.013)				
$\Delta \ln(\text{Monitorred metres})$ <i>OSRM routed</i>			0.686*** (0.018)	0.686*** (0.018)	0.572*** (0.017)	0.595*** (0.018)
Constant	0.751*** (0.117)	0.843*** (0.121)	0.094*** (0.006)	0.094*** (0.006)		
Observations	73,130	73,130	73,130	73,130	73,130	73,130
Por/Pow/Dist FEs	No	No	No	No	Yes	Yes
Weights	None	None	None	None	None	$\pi_{ij2001} + 1$
Projected R^2	0.690	0.691	0.441	0.441	0.618	0.654

Note:

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

Table A.14: Maximum Likelihood Estimates of traffic's effect on the number commuting

	<i>Dependent variable:</i>			
	Number of commuters			
	(1)	(2)	(3)	(4)
ln(Traffic)	-0.9760 (0.4320)** [0.1609]***	-0.1024 (0.1589) [0.1178]	-0.3236 (0.2375) [0.1852]*	-0.2258 (0.2062) [0.1419]
First stage residual	0.8764 (0.8764)* [0.1783]***	0.0922 (0.1544) [0.1178]	0.2218 (0.2405) [0.1886]	0.1935 (0.2086) [0.1439]
Dist × Post FE	No	Yes	Yes	Yes
Pow/Por × Post FE	None	Both	Both	Both
Controls	No	No	Yes	Yes
Model	Poisson	Poisson	Poisson	Negative binomial
Overdispersion parameter				9.227
Observations	493,976	493,976	493,976	493,976

Note:

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

Poisson regressions use 2001 total flows plus one as weights and all regressions include MSOA pair fixed effects. Standard errors in round parenthesis are two-way clustered at the residence and workplace levels and square parenthesis are clustered by pair. Controls are described in section 4 footnote 14.

Table A.15: Maximum Likelihood Estimates of traffic's effect on mode shares

	<i>Dependent variable:</i>			
	Number of commuters (by mode)			
	(1)	(2)	(3)	(4)
ln(Traffic)	-6.9119	0.4292	0.6405	0.3222
× Driver	(0.9407)***	(0.9511)	(1.319)	(1.2147)
	[0.2517]***	[0.4359]	[0.6512]	[0.4969]
First stage residual	6.4022	-0.5145	-0.5712	-0.2815
	(0.982585)	(.99713)	(1.2892)	(1.1839)
	[0.2566]***	[0.4380]	[0.6560]	[0.4951]
Dist × Post × Driver FE	No	Yes	Yes	Yes
Pow/Por × Post × Driver FE	None	Both	Both	Both
Controls	No	No	Yes	Yes
Model	Poisson	Poisson	Poisson	Negative binomial
Overdispersion parameter	—	—	—	9,999
Observations	588,910	588,910	588,910	588,910

Note:

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

Poisson regressions use 2001 total flows plus one as weights and all regressions include MSOA pair-year and pair-mode fixed effects. Standard errors in round parenthesis are two-way clustered at the residence and workplace levels and square parenthesis are clustered by pair. Controls are described in section 4 footnote 14 and are additionally interacted with a driver dummy variable.

Table A.16: The congestion charge's effect on mean utility with asymmetric agglomeration

$\lambda_l^{\lambda_h}$	-0.1	-0.07	-0.035	0	0.035	0.07	0.1
-0.1	0.015	0.016	0.017	0.019	0.021	0.022	0.024
-0.07	0.015	0.016	0.018	0.019	0.021	0.023	0.025
-0.035	0.016	0.017	0.018	0.020	0.022	0.024	0.027
0	0.016	0.017	0.019	0.021	0.023	0.026	0.029
0.035	0.016	0.017	0.019	0.021	0.024	0.028	0.031
0.07	0.016	0.018	0.020	0.023	0.026	0.030	0.035
0.1	0.017	0.018	0.021	0.024	0.028	0.033	0.039