Economic growth and congestion – a new methodology for agglomeration elasticities

Csaba G. Pogonyi

July 15, 2017

Railway and Transport Strategy Centre
Department of Civil Engineering
Imperial College London

Extended abstract submitted for HeART 2017

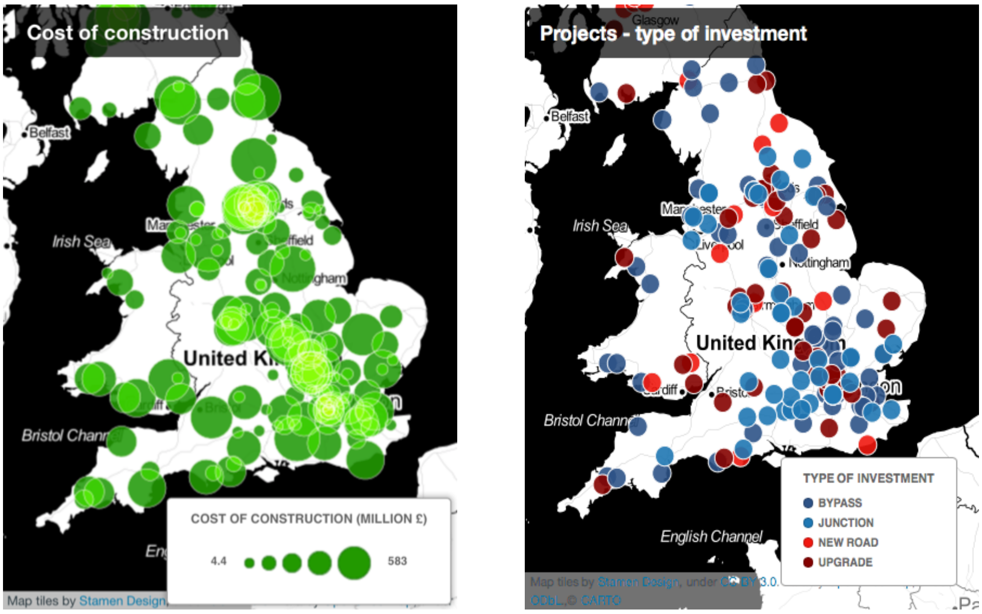
1 Introduction

This research proposes a new methodology for the estimation of agglomeration elasticities: how increased accessibility leads to more productive firms. This new methodology incorporates negative externalities arising from congestion directly into effective density measures.

Accessibility measures widely used in the literature (Melo, Graham, and Noland 2009) – like effective density measured by Euclidean distance or by yearly generalised cost – do not capture actual transport possibilities and thus lead to confounded elasticity estimations.

Causal analysis so far has implicitly assumed that completing a transport investment necessarily leads to better accessibility and this better accessibility leads to higher agglomeration, which in turn makes firms more productive. However, in transport literature it has been evident that an investment does not necessarily lead to better accessibility (due to induced congestion for example Downs 1982).

Using openly available traffic flow and journey time datasets, current effective density measures can be improved. The methodology presented in this paper makes it possible to evaluate if and how a transport investment leads to actual change in traffic flows and journey times (thus change in accessibility). We create accessibility measures, which use actual observed journey times, for every ward in England between 2009 and 2015. Especially peak time accessibility captures negative agglomeration externalities arising from traffic congestion. We find that peak time effective density provides a more realistic measure for agglomeration forces.



(a) Road investments by cost

(b) Road investments by type

Figure 1: Road investments in England 2009-2014

2 Data

We reconstructed Highway England's road network as a routable graph network for every year between 2009 and 2015, and linked them with journey time and traffic flow observations (Highway Agency 2017 and Department for Transport 2017.

We scraped road schemes from the website of the CBRD (CBRD n.d.) and created a database. This contains all road schemes - all completed, under construction or even canceled projects - in the United Kingdom between 2000 and 2016.

The Annual Respondents Database X (ARDx, UK Data Service 2016) will be used for firm productivity estimation. This data provides production characteristics for the period 2009-2015. Employment variables will be derived from the Annual Survey of Hours and Earnings (ASHE, UK Data Service 2017) database for the same period.

3 Methods

The first part evaluates the effect of road schemes on traffic conditions and calculates peak time on-road observed accessibility measures for every ward in England between 2009 and 2015. The second part estimates total factor productivity (TFP) for medium sized firms present in the ARDx dataset for the same period. The third part estimates agglomeration elasticities.

3.1 Evaluating road schemes and calculating mean effective density measures

The routable road network makes it possible to assess the impact of a road scheme on traffic conditions. We are able to observe not just changes in journey time and traffic count, but also the change in the distribution of traffic count (see 6 in Appendix). Using the Highway England graph network and the road construction database, we have created a simple differences-in-differences initial evaluation. On 2 there is no significant difference between treated (within 5 km of the scheme) and untreated (between 5 and 30 km of the scheme) road links. The assumption to be tested is that if there is no change in traffic conditions, the road scheme does not have an impact on the economy.

This research builds on the new methodology suggested by Graham and Gibbons 2017. Their accessibility measure - Mean Effective Density (MED) is

$$\rho_i = \frac{1}{n}\sum_{j=1}^{n} m_j f(d_{ij}), \tag{1}$$

where i is a geographical area (in our case a ward in England), n is the count of these areas, m is some measure of economic mass (in our case employment) for area j, and $f(d_{ij})$ is the so-called impedance function: a decreasing function of the cost of moving from i to j.

They show that by applying the law of large numbers, (1) can be decomposed to

$$\rho_i^D \xrightarrow{p} \mathrm{E}(M_j D_{ij}^{-\alpha}) = \mathrm{E}(M_j)\mathrm{E}(D_{ij}^{-\alpha}) + Cov(M_j, D_{ij}^{-\alpha}), \tag{2}$$

The product of the two expected values on the right hand side is "scaled centrality" (SC). The size of SC depends on the geographical location of the zone ("centrality") and on the "zone size distribution" (a zone with equal mass, but smaller size will have higher MED).

The covariance term on the right hand side of (2) is called "mass-impedance covariance" (MIC). In real world, mass and distance are positively correlated most of time as accessibility has a positive effect on agglomeration.

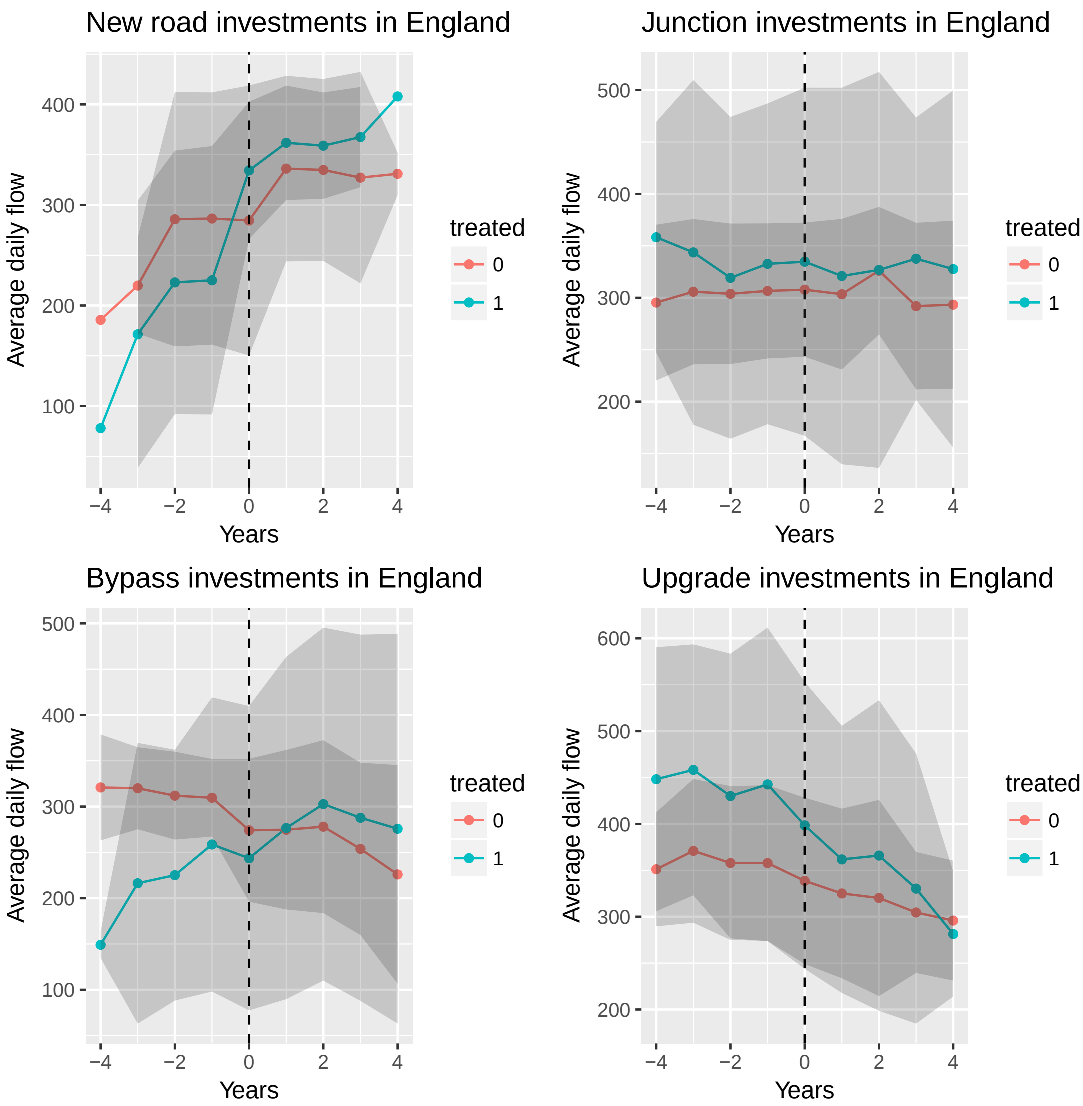


Figure 2: The effect of road projects on traffic flow in England 2009-2015 (SE around data points)

This research builds on their model and introduces an impedance function, which is as close as possible to real world conditions. $f(d_{ij})$ in (1) in our model is

$$d_{ij} = c_{FC} + \kappa_{ij}^{O} c_{AC} + \tau_{ij}^{O} w_i, \tag{3}$$

where c_{FC} is the fix cost of using a vehicle, κ_{ij}^{O} is the on-road distance between i and j, c_{AC} is the average per-km cost of using a car, τ_{ij}^{O} is the travel time and w_i is the value of travel time (VTT). O denotes either off-peak or peak travel times.

Shortest path and travel time are calculated not just for a theoretical yearly average, but separately for both peak and off-peak times.

Results after calculating MED(1) with standard Euclidean distance, distance on the road network and peak travel times can be seen on Figure 3a. The distribution of peak time MED is more spread than the other ones. This confirms our assumption that road congestion decreases MED differences among zones, thus using peak time MED values in firm productivity research may yield more precise estimates.

3.2 Estimating total factor productivity

Our basic firm production as a regression equation is

$$y_{it} = f(L, K, M) = \beta_0 + \beta_L l_{it} + \beta_K k_{it} + \beta_M m_{it} + \mu_z + \omega_{it} + \epsilon_{it}, \tag{4}$$

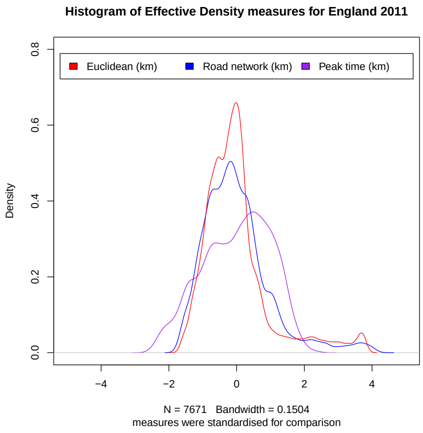
where all lower-cases are natural logarithms, i denotes firm and t denotes year. L denotes labour, K capital and M materials. μ_z is a two-digit industry specific fixed effect. ω_{it} is the observed time and firm specific deviation from the mean efficiency level and ϵ_{it} represents unexpected deviation due to measurement error or other random circumstance. We estimate this production function for every main ("one-digit") industry separately and on three geographical levels: low density areas, cities and conurbations. In addition to the linear model, we estimate TFP - Ω_{it} - using three of the most widely used semi-parametric methodologies as well: Olley and Pakes 1996, Levinsohn and Petrin 2003 and Ackerberg, Caves, and Frazer 2006.

3.3 Estimating agglomeration elasticities

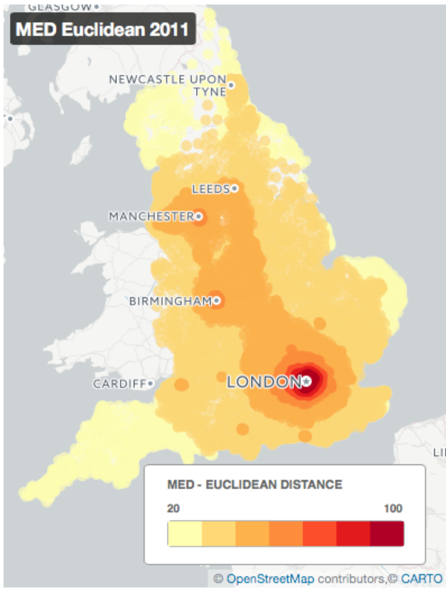
We are able to test our model on micro level panel data. This makes it possible to use sophisticated econometric methods like semiparametric regressions, GMM or Instrumental Variables.

The basic panel Fixed Effect model is

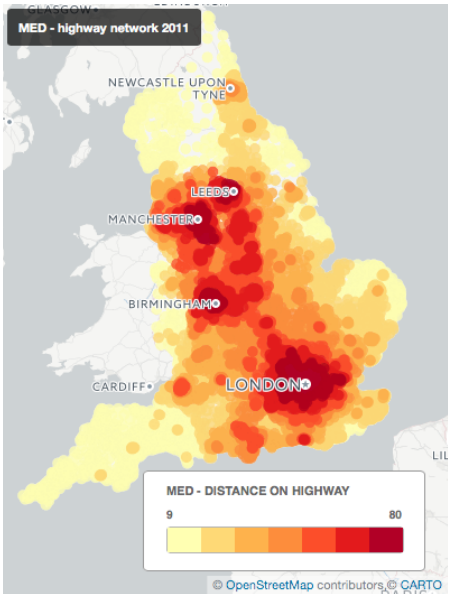
$$\omega_{it} = \beta_0 + \beta_\rho \rho_{it}^{O} + \mu_i + t_t + \epsilon_{it}, \tag{5}$$



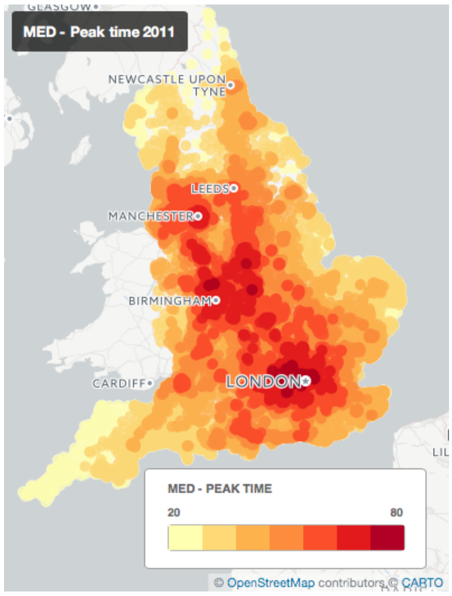
(a) Histograms of all MEDs



(b) MED – Euclidean 2011



(c) MED – Highway network 2011



(d) MED – peak time 2011

Figure 3: Mean Economic Density (MED) for England 2011

where ω_{it} is the natural logarithm of the estimated total factor productivity, ρ_{it}^{O} is the natural logarithm of one of the calculated MEDs (Euclidean, distance on the road network, off-peak or peak time), μ_i is a two-digit industry fixed effect and t_t is a time trend that allows for unobserved shocks. All plants i of firms are assigned to ward i. We assume ϵ_{it} to be a serially uncorrelated white noise error. β_ρ is the agglomeration elasticity, which shows how increased agglomeration leads to more productive firms.

Reverse causality and unobserved confounding are major problems with this setting; therefore, we use panel GMM estimation. Moreover, we test two Instrumental Variables as well: canceled road schemes and road maintenance as IVs for MED.

In order to test for nonlinearities between agglomeration and firm productivity, we employ a semi-parametric linear additive mixed model (SPLAM).

Next steps include the incorporation of public transportation and rail accessibility into the model. Another important step is to add minor road traffic flow and speed to the highway network in order to make shortest path calculations more precise.

References

Ackerberg, Daniel, Kevin Caves, and Garth Frazer. "Structural identification of production functions". In: (2006).

CBRD. *Road Schemes*. URL: http://www.cbrd.co.uk/road-schemes.

Department for Transport. *Road Safety Data*. 2017. URL: https://data.gov.uk/dataset/road-accidents-safety-data.

Downs, Arthur. "The Law of Peak-Hour Expressway Congestion". In: *Traffic Quarterly* 16.3 (July 1982). URL: https://trid.trb.org/view.aspx?id=694596.

Graham, Daniel J. and Stephen Gibbons. "Updating Agglomeration Elasticities: Phase 1a Technical Report". 2017.

Highway Agency. *Network journey time and traffic flow data*. 2017. URL: https://data.gov.uk/dataset/dft-eng-srn-routes-journey-times.

Levinsohn, James and Amil Petrin. "Estimating production functions using inputs to control for unobservables". In: *The Review of Economic Studies* 70.2 (2003), pp. 317–341.

Melo, Patricia C., Daniel J. Graham, and Robert B Noland. "A meta-analysis of estimates of urban agglomeration economies". In: *Regional Science and Urban Economics* 39.3 (2009), pp. 332–342. ISSN: 01660462. DOI: 10.1016/j.regsciurbeco.2008.12.002.

Olley, S and Ariel Pakes. "Dynamic behavioral responses in longitudinal data sets: Productivity in telecommunications equipment industry". In: *University of Pennsylvania, Philadelphia, PA. http://econweb. umd. edu/˜ haltiwan/olley_pakes. pdf* (1996).

UK Data Service. *Annual Survey of Hours and Earnings, 1997-2015*. 2017. URL: https://discover.ukdataservice.ac.uk/catalogue?sn=6689#close.

— *UK Data Service and Administrative Data Service welcome new Digital Economy Bill*. 2016. URL: https://www.ukdataservice.ac.uk/news-and-events/newsitem/?id=4643.

4 Appendix

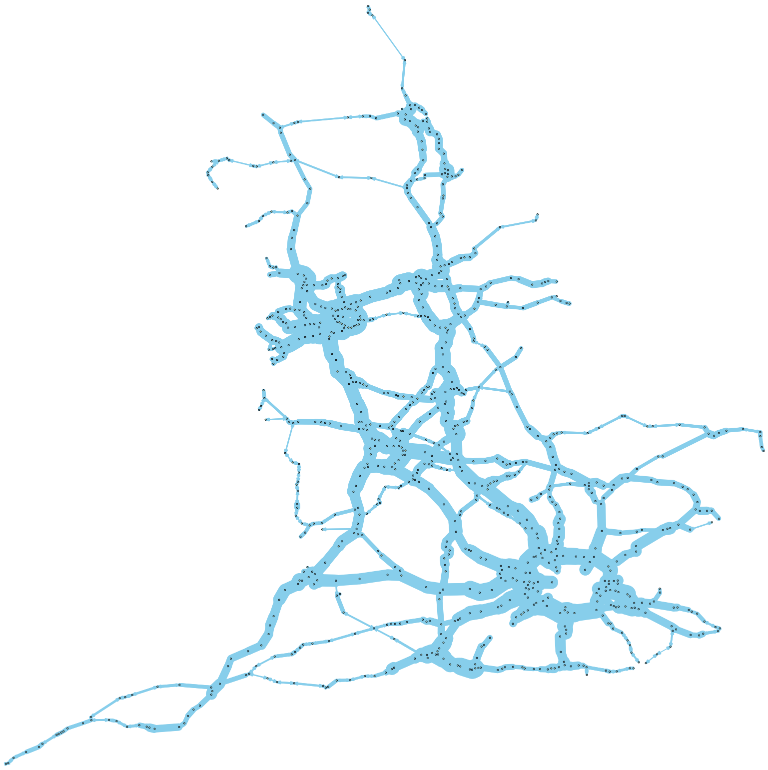


Figure 4: Highway England's routable graph network 2011 (width of edges indicate average yearly traffic flow)

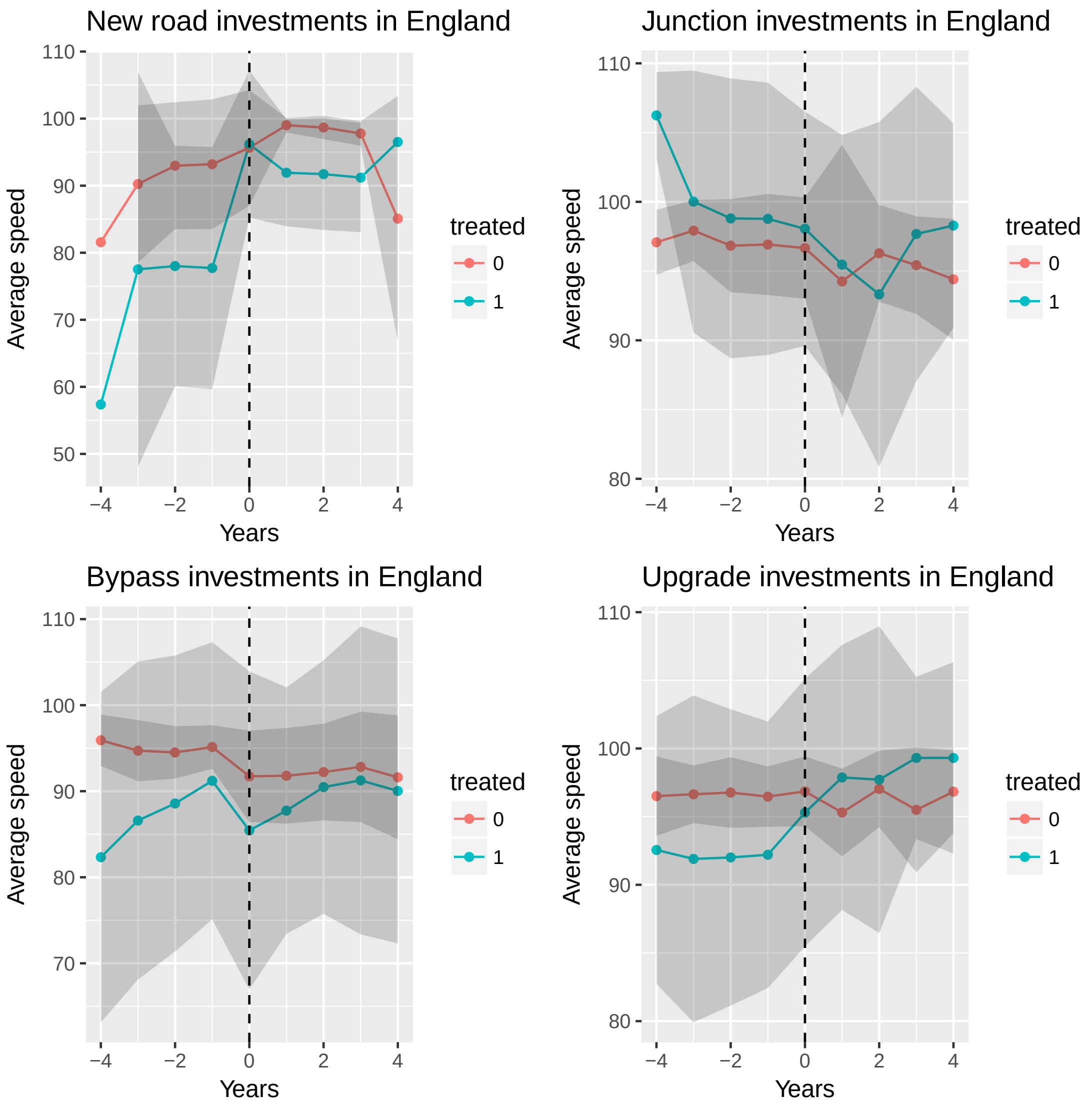


Figure 5: The effect of road projects on traffic speed in England 2009-2015 (SE around data points)

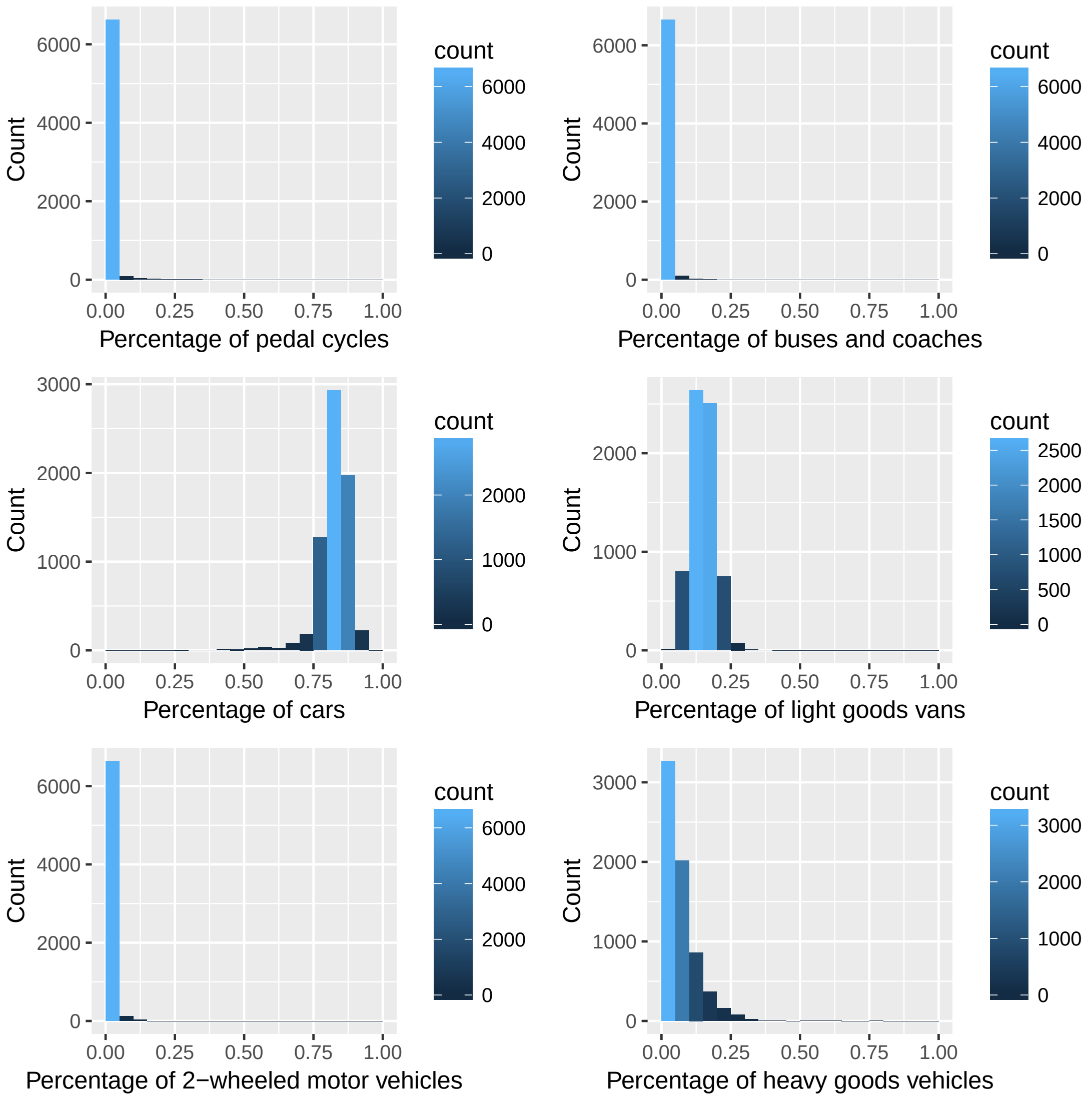


Figure 6: Vehicle type distribution of traffic on Highway England's network in 2014

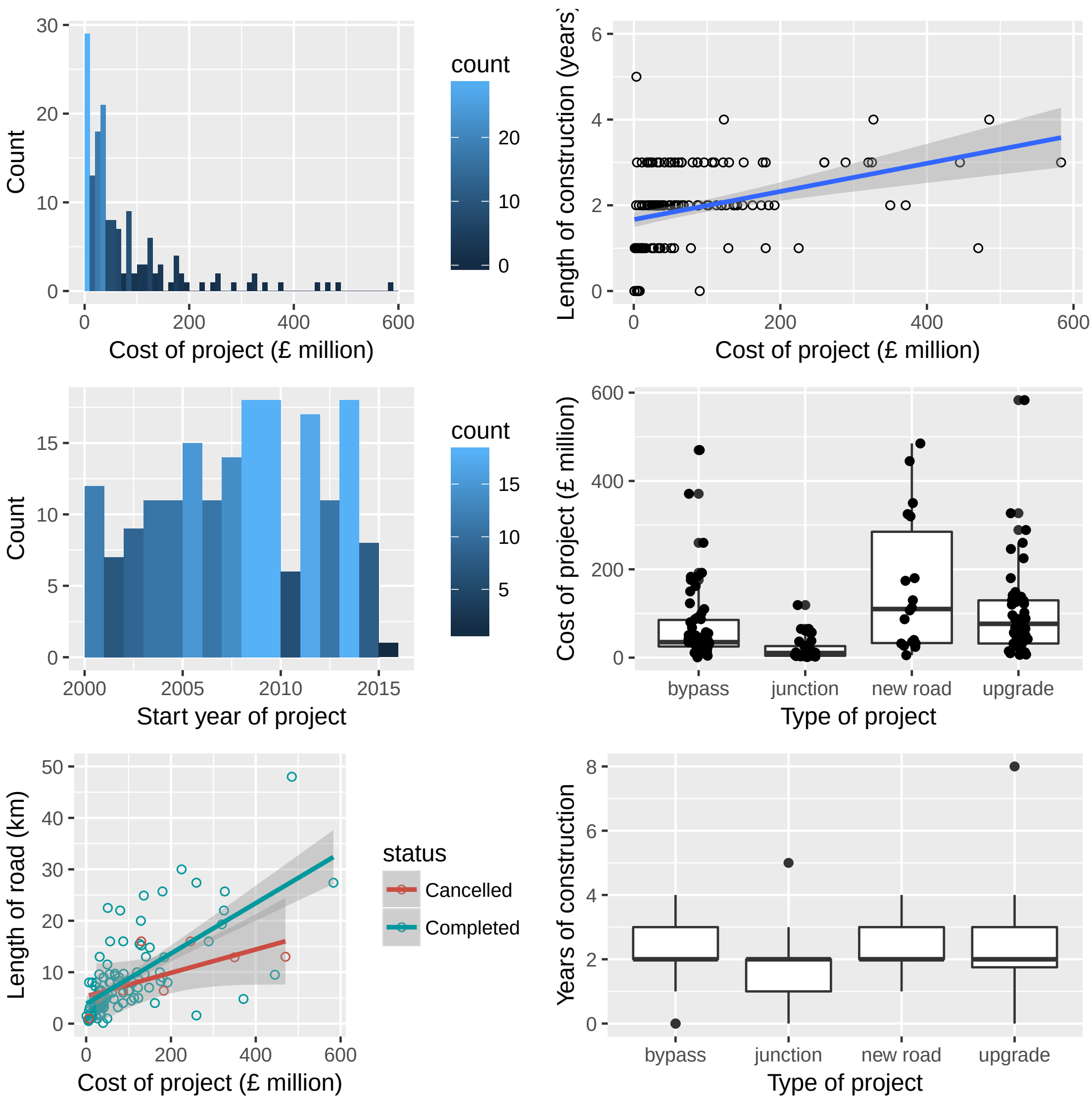
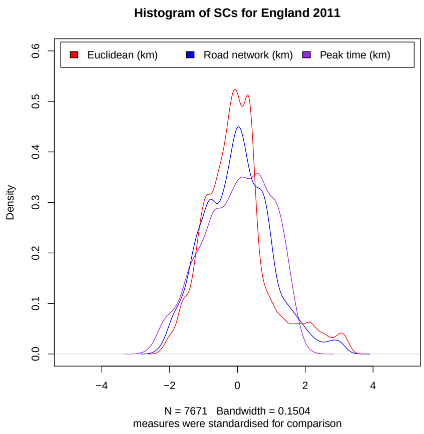
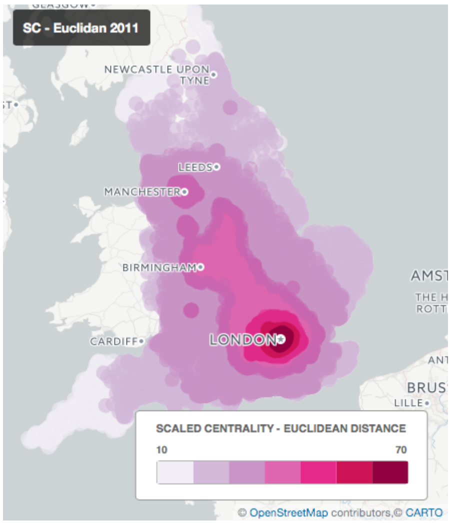


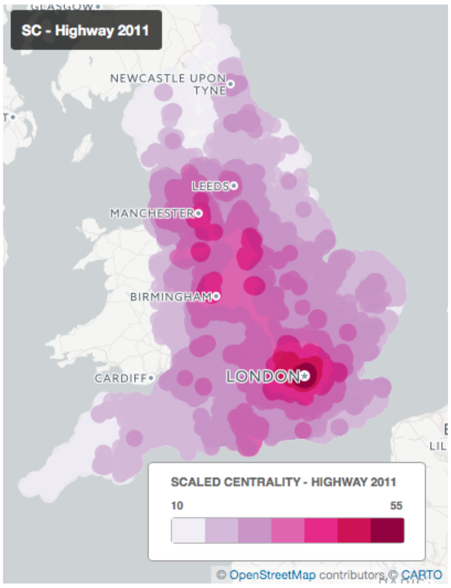
Figure 7: Detailed statistics of road investments in England 2000-2016



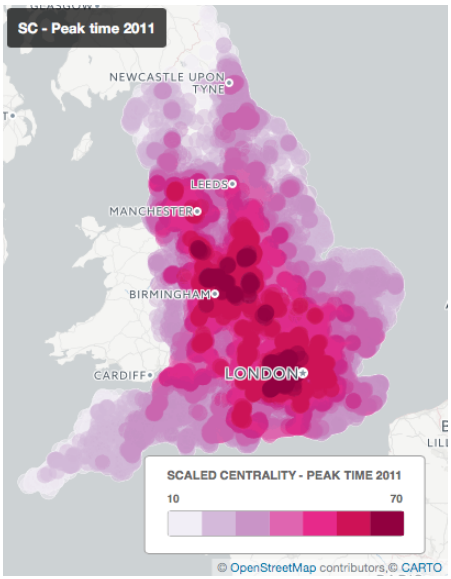
(a) Histograms of all SCs



(b) SC – Euclidean 2011

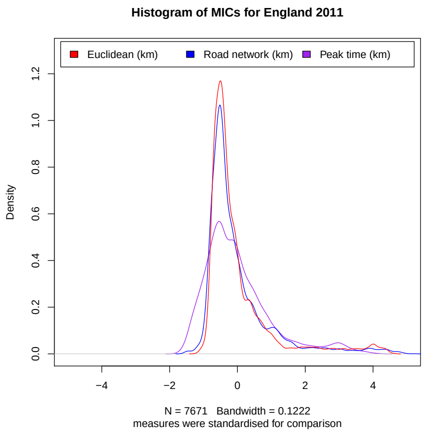


(c) SC – Highway network 2011

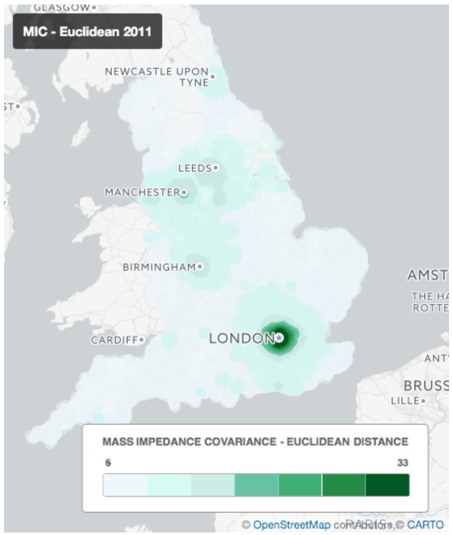


(d) SC – peak time 2011

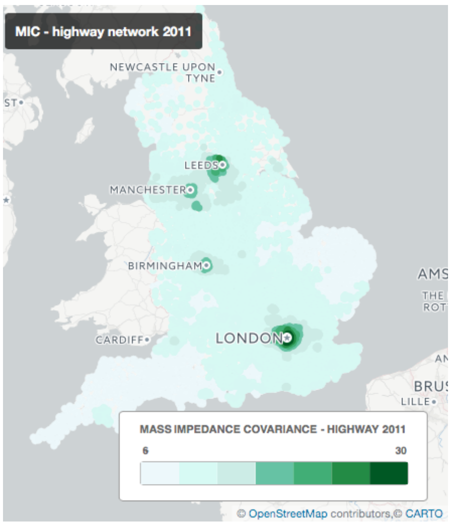
Figure 8: Mass Impedance Covariance (MIC) for England 2011



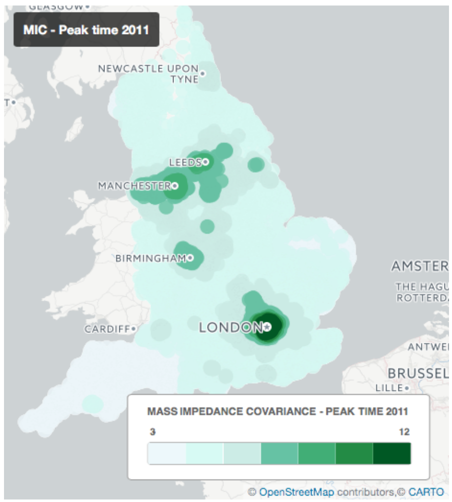
(a) Histograms of all MICs



(b) MIC – Euclidean 2011



(c) MIC – Highway network 2011



(d) MIC – peak time 2011

Figure 9: Mass Impedance Covariance (MIC) for England 2011