# Quantification of non-linear effects in agglomeration economies for transport appraisals

Anupriya[1], Daniel J. Graham[1], and Prateek Bansal[2]

[1]Transport Strategy Centre, Department of Civil & Environmental Engineering, Imperial College London, London, SW7 2AZ, UK.
[2]Department of Civil & Environmental Engineering, National University of Singapore.

## Short summary

Agglomeration economies arising from the spatial concentration of economic activity have been known to exist and induce higher productivity for firms. The existing empirical evidence, however, has two key caveats. First, it mostly assumes a pre-specified (mostly log-log) functional form for the relationship between firm productivity and agglomeration. Second, it may lack valid instruments to adjust for potential confounding biases (for instance, from the omission of characteristics of local input and output markets) in the estimation of this relationship. This study adopts a flexible Bayesian Non-Parametric Instrumental Variables based approach to quantify non-linear effects in agglomeration economies. The approach uses innovative external instruments derived from traffic casualty data. We adopt a two-step framework: we first isolate the firm's total factor productivity from a Cobb-Douglas production function and thereafter estimate the non-linear effects of agglomeration on this productivity. Using data from a sample of firms classed into six key industry sectors in England, we present novel evidence that indicates the presence of significant non-linearities in agglomeration elasticities for most industry sectors. Our results provide critical inputs for the appraisal of transport investments.
**Keywords**: Agglomeration; Cost-benefit Analysis; Wider Economic Impacts; Elasticity; Non-parametric statistics; Bayesian machine learning.

## 1 Introduction

Transport investments are frequently aimed at bringing economic and social benefits to the economy. However, given their scale, it is important for policymakers to have a rigorous apriori understanding of the magnitude of potential benefits arising from the investment. Cost Benefit Analysis (CBA) provides a well-established theoretical basis to measure these benefits ex-ante and thus warrant such extensive investments (Graham & Gibbons, 2019; A. Venables et al., 2014; Mackie et al., 2012). CBA recasts the costs and benefits of the investment in monetary terms to estimate the net change in social welfare arising from the transport improvement.
The current appraisal process embraced in the UK's Transport Analysis Guidance (TAG) recognises two broad categories of such welfare impacts: (1) Direct user benefits (DUBs) and (2) Wider economic impacts (WEIs). DUBs consist of impacts on existing and new users of the transport system, generated via changes in the generalised cost of travel (say, via alterations in travel time or quality of service). Economic theory suggests that DUBs can capture all impacts under idealised economic conditions: under perfect competition, constant returns to scale, and in the absence of market failures. DUBs, therefore, formed the crux of the calculations in conventional CBA. Nevertheless, the above-described market conditions are seldom encountered in practice, thus, undermining the ability of DUBs to capture economic impacts in exhaustive detail. The scope of CBA has therefore been extended, primarily in the past decade, to include the economic impacts caused by market imperfections and externalities. Such *wider* group of impacts on the economy are manifested in WEIs. The overarching aim of this paper is to revisit the empirical evidence on WEIs of transport investments emerging via scale economies of agglomeration.

### *Background*

Agglomeration economies are spatial externalities that arise when economic agents (individuals and firms) locate in close proximity to each other, or in other words, locate within agglomerations of economic mass. Such proximity facilitates greater sharing, matching, and learning interactions between agents, which are known to be the key drivers of agglomeration (Duranton & Puga, 2004). For firms, agglomeration translates into benefits in form of improved labour market interactions, knowledge spill-overs, specialisation, and increased sharing of inputs and outputs (Marshall, 1920). Given these underlying benefits, economic theory predicts a positive impact of agglomeration on the productivity of firms. Theory suggests two forms of agglomeration economies: (i) economies of industry concentration or localisation economies, which includes benefits that occur through enhanced specialisation; and, (ii) economies of urban concentration or urbanisation economies, which result from the scale and diversity of markets (DfT, 2016). The former effects are external to the firm but internal to the industry, while the latter are external to the firm and the industry but internal to the urban area (or city) (Graham & Gibbons, 2019). Firms mostly experience the two forms of agglomeration economies simultaneously. It is, therefore, difficult to disentangle the two effects. The paper focuses on the quantification of urbanisation economies, which are typically considered in appraisal calculations.

The impact of transport investments on agglomeration economies emerge from transport's effect on economic geography (distribution of individuals and firms) and consequent proximity between economic agents (A. J. Venables, 2007). For fixed geography, transport improvements reduce the generalised costs of interaction between agents in a cluster, thereby assisting interactions and changing the effective density of the cluster. These effects are referred to as static agglomeration effects. Transport investments can also bring about changes in economic geography by making some locations more attractive to live and work, resulting in the relocation of individuals and firms. Such movements lead to changes in the physical density of the cluster and facilitate new and different interactions (DfT, 2016). These effects are termed dynamic agglomeration effects. Static effects generally appear as a subset of the overall dynamic effects of a transport provision. Interestingly, dynamic agglomeration effects may have a positive or negative impact on the productivity of firms in a cluster depending on whether the spatial density of the cluster increases or decreases as a consequence of the transport investment. In summary, the above discussion suggests that if increasing effective agglomeration results in productivity gains for firms, and if transport has an underlying role in determining effective agglomeration, then transport investments may induce productivity benefits or disbenefits for firms. Therefore, comprehension of these productivity effects is critical to developing a meticulous assessment of the impact of transport investments.

Consistent with the theory of urban agglomeration, the weight of empirical evidence in the literature supports a significant positive relationship between agglomeration (often represented by city size or the degree of access to economic mass (ATEM)) and productivity (measured by wages or by Total Factor Productivity (TFP)) (see, for instance, Lall et al., 2004; Au & Henderson, 2006; Rosenthal & Strange, 2008; Baldwin et al., 2010; Graham & Dender, 2011; Combes et al., 2012; Morikawa, 2011; Maré & Graham, 2013; Marrocu et al., 2013; Ahlfeldt et al., 2015). Elasticity estimates from the literature range between -0.800 and 0.658, with an unweighted mean of 0.046 and median value of 0.043 (refer to Graham & Gibbons, 2019, for a detailed review). The literature thus indicates that agglomeration economies exist and cause higher productivity for firms and workers. Nevertheless, the estimated magnitude of this effect varies substantially across studies. A meta-analysis of the empirical literature conducted by Melo et al. (2009) indicates that such variation results from contextual factors associated with the study design such as the nature of economies and urban systems and the type of industry sectors under study. Additionally, Graham & Gibbons (2019) highlights that the differences in estimated elasticities also result from the differences in methodological approaches adopted in previous studies. While most studies concurrently assume a log-log relationship between agglomeration and productivity, the approaches adopted to identify this relationship vary substantially. In particular, Graham & Gibbons (2019) note that there are considerable discrepancies in the extent to which these studies attempt to correct for potential confounding biases in the estimation of this relationship.

Relatedly, Graham & Gibbons (2019) identify six potential mechanisms via which such biases may emerge. First, confounding may occur due to the presence of unobserved firm-level sources of productivity that are not only crucial to the firm's choice of inputs, and thereby its TFP (see Van Beveren, 2012, for details), but may be determined by local technology factors such as agglomeration. Second, confounding may also occur due to the absence of knowledge on a firm's market exit decisions (see Ackerberg et al., 2006, for details), which may be determined by agglomeration. In particular, firms located in clusters of higher agglomeration may experience more competition, which could result in the exit of less productive firms from the market. Third, confounding biases may emerge via unobserved heterogeneity in output prices of firms, which have a systematic correlation with market competition, and thereby with agglomeration. Fourth, confounding may appear due to spatial sorting or self-selection of firms, which occurs when firms within the same industry derive unobserved productivity benefits by engaging in different activities across different locations. Such unobserved heterogeneity is often correlated with the level of agglomeration. Fifth, the relationship between agglomeration and productivity may be simultaneously determined. As shown by Graham et al. (2010), higher productivity locations may attract more private investment over time leading to larger agglomeration and a consequent increase in productivity. Failure to account for this reverse causality between productivity and agglomeration may produce biased and inconsistent estimates of agglomeration economies. Finally, additional confounding may emerge from unobserved components of local technology, such as specific characteristics of local input and output markets, that may be determine both agglomeration and productivity. One of the key objectives of this research is to deliver agglomeration elasticities that are robust to confounding, and as such suitable to inform the appraisal of transport investments.

### *Overview of the Analysis*

This paper estimates the relationship between agglomeration and productivity by adopting a two-step approach Combes & Gobillon (2015). The first stage model within this approach estimates TFP from the production function. The predicted values of TFP are then used as the dependent variable in a second-stage regression on agglomeration, which delivers the agglomeration elasticities. However, contrary to previous studies, we exploit the ability of this approach to model a flexible non-parametric relationship between agglomeration and productivity. In particular, we note that existing models in the literature mostly presume a Cobb-Douglas model for the agglomeration-productivity relationship. We argue that while economic theory suggests a positive impact of agglomeration on productivity, it does not necessarily imply that the relationship should be log-log. Furthermore, contextualisation of agglomeration elasticities across studies (as discussed in Section 1) may be indicative of the variation of these elasticities over agglomeration. Hence, we assert that parametric models with a predefined functional form may fail to capture the non-linearities in the agglomeration-productivity relationship, thus delivering estimates of agglomeration elasticities that may be biased.

To address this limitation, in this paper, we empirically estimate the relationship between agglomeration and productivity using a Bayesian non-parametric instrumental variables (NPIV) estimator (Wiesenfarth et al., 2014) that allows us to (i) capture non-linearities in the relationship with a non-parametric (NP) specification that does not require an assumed a-priori functional form, and (ii) adjust for any confounding bias using instrumental variables (IVs). As discussed in Section 1, such biases may emerge from reverse causality or from the omission of important covariates. Critical to the second point, we note that the literature may lack valid IVs for agglomeration, which could hinder the identification of the agglomeration-productivity relationship. To overcome this limitation, we recognise a novel external IV for agglomeration that is derived from traffic casualty records. In particular, we consider the severity of traffic accidents among active mode and motorcycle users during peak hours in a given location and time period as a relevant and exogenous instrument for agglomeration in that location and time.

We apply the proposed approach to a sample of firms in England, divided into six key industrial sectors: Manufacturing; Construction; Wholesale and Distribution; Transport; Information and Communication Technology; and Finance. To measure the TFP of these firms, we make use of an exhaustive panel dataset recorded by the Department of Trade and Industry. The data relate to annualised accounting information provided by all companies registered in the UK. For the purpose of this study, we consider the data between the years 2015 and 2019. Relevant measures of agglomeration for this period are formed using employment records maintained by the Office for National Statistics.

### *Contributions*

The major contributions of this study can be summarised as follows:

1. We derive a novel external instrument from traffic accident data to identify the relationship between agglomeration and firm productivity.
2. Our study delivers a novel comprehension of the non-linearities in agglomeration elasticities across key industry sectors in England. For instance, we find the finance sector to be associated with agglomeration diseconomies at lower levels of agglomeration and agglomeration economies at higher levels of agglomeration. This result indicates the presence of a critical economic mass beyond which productivity benefits set in for firms in this industry sector.

Interestingly, across all sectors, we note that the agglomeration elasticities take more extreme values than those from a linear model. For instance, while the linear model suggests agglomeration economies of magnitude 0.144 for firms in the finance industry, our non-linear model finds the elasticity estimates to go up to a level of 0.710.

3. Quantification of non-linearities in agglomeration economies facilitates a novel understanding of the spatial distribution of the productivity benefits of agglomeration across England. Such a mapping could be instrumental in identifying the potential gainers and losers of productivity benefits arising from a transport improvement.

The rest of this paper is structured as follows. Section 2 describes the data and the methodology used to estimate the agglomeration elasticities. Section 3 presents the results from the empirical study. Conclusions are drawn in the final section.

## 2 Model and Data

This section is divided into four subsections. The first subsection elaborates on the measure of agglomeration used in this study. The second subsection discusses the adopted two-step approach to estimate the relationship between agglomeration and productivity. The third subsection briefly describes the Bayesian NPIV method in the context of this study. The final subsection discusses the data used in this analysis.

### *Measure of Agglomeration*

A crucial prerequisite to understanding the WEIs that arise from agglomeration is to develop a suitable measure of agglomeration for each location or geographical zone. In line with the current CBA practice in the UK, we represent agglomeration using ATEM, or in other words, the Mean Effective Density (MED). The MED $\rho_j$ for zone, $j, j = (1, \ldots, n)$, is calculated as follows:

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

where $m_j$ represents a measure of economic activity in each zone $j$ and $f(.)$ denotes the deterrence function, which is a decreasing function of the cost of travelling from origin $j$ to destination $k$. The measure is designed to capture the effects of the geographic centrality of the zones, their size distribution, and the spatial distribution of economic mass (Graham & Gibbons, 2019). We consider the zonal employment level $E_{jt}$ as the measure of the economic activity of zone $j$ and year $t$ and the inverse Euclidean distance between the centroids of each zone $d_{jk}^{\alpha}$ for the construction of the deterrence function, where $\alpha$ is the distance decay parameter, generally assumed to take a value of 1.0. The resulting MED for zone $j$ in year $t$ is thus:

$$\rho_{jt} = \frac{1}{n}\sum_{k=1}^{n} \frac{E_{kt}}{d_{jk}^{\alpha}}. \tag{1}$$

### *Agglomeration and Productivity*

As mentioned in the Introduction, to quantify the impact of urban agglomeration on productivity, we adopt a two-step approach. The first step involves estimating the TFP of a firm by constructing its production function. The second step comprises regressing the estimated TFP values on the chosen measure of agglomeration (that is, MED) to derive the agglomeration elasticities $\delta^{\rho}$, given by

$$\delta^{\rho} = \frac{\partial \log \omega}{\partial \log \rho} \tag{2}$$

We emphasise that we choose TFP over labour productivity measures such as wage rate as the latter has the following disadvantages in the context of appraising WEIs. First, they can be determined by transport improvements via routes other than productivity (for instance, via shifts in labour supply). Second, while TFP is exclusively determined by local technology, output prices and average labour skills; wages are additionally influenced by the relative prices of other factors such as land and housing prices. Such dependability can introduce severe confounding biases in the estimation of agglomeration elasticities (Combes & Gobillon, 2015). Third, wage-based measures carry the assumption that the wage equals the value of the marginal product in competitive equilibrium. However, the equality assumption seldom holds in practice as wages are typically proportional to labour productivity (Combes & Gobillon, 2015). Lastly, wage-based measures only provide a partial representation of productivity as they are limited to the impacts on the labour input alone. Conversely, TFP provides a more comprehensive measurement of productivity with respect to all inputs, which in the agglomeration context is more critical as agglomeration may affect technology in several ways (Maré & Graham, 2009).

### Step 1: Estimating total factor productivity

Consistent with the literature, we assume that the production of outputs $Y_{it}^s$ by a firm $i$ in industry sector $s$ in year $t$ to follow a Cobb Douglas production function structure with inputs; capital $K_{it}^s$, labour $L_{it}^s$ and materials $M_{it}^s$; as covariates:

$$\log Y_{it}^s = \beta_k^s \log K_{it}^s + \beta_l^s \log L_{it}^s + \beta_m^s \log M_{it}^s + \omega_{it}^s + \gamma_t^s + e_{it}^s \tag{3}$$

where $\beta_k^s$, $\beta_l^s$ and $\beta_m^s$ are constants representing the elasticities of output with respect to the associated factor of production. $\omega_{it}^s$ is the unobserved efficiency or productivity of the firm, commonly referred to as its Total Factor Productivity (TFP). TFP represents the efficiency level that remains unobserved by the analyst, but is known to (or predicted by) the firm. $\gamma_t^s$ are year dummies that capture the year-specific effects on productivity and inflation. $e_{it}^s$ is a normally distributed idiosyncratic error term, or in other words, all random shocks to the outputs. From equation 3, the firm's TFP $\omega_{it}^s$ can be estimated as follows:

$$\hat{\omega_{it}^s} = \log Y_{it}^s - \hat{\beta_k^s} \log K_{it}^s - \hat{\beta_l^s} \log L_{it}^s - \hat{\beta_m^s} \log M_{it}^s - \hat{\gamma_t^s}. \tag{4}$$

Note that TFP affects the firm's choice of input factors and market exit decisions, thus rendering the variable factors of production, labour, and materials, endogenous in the model (De Loecker, 2007). Identification of the model parameters and estimation of TFP thus requires careful consideration of the potential confounding biases caused by the endogenous outputs. Following from the review of the literature on TFP estimation by Van Beveren (2012), we make use of a panel control function (CF) approach proposed by Ackerberg et al. (2006), which is an extension to Levinsohn & Petrin (2003). Ackerberg et al. (2006)'s CF approach uses a function with materials and capital as arguments to proxy for the endogenous unobserved productivity. This function is introduced into the production function (equation 3) as an additional model component to obtain consistent estimates of the model parameters.

### Step 2: Estimating the effect of agglomeration on productivity

To estimate the causal impact of agglomeration on productivity, we consider the estimated TFP $\hat{\omega_{it}^s}$ to be a function of the agglomeration measure $\rho_{it}^s$ indicating the MED of the zone $j$ where the firm $i$ is located.

$$\hat{\omega_{it}^s} = S^s(\rho_{it}^s) + \eta_{it}^s + \xi_{it}^s. \tag{5}$$

where $\eta_{it}^s$ consists of the unobserved characteristics of firm-level productivity. $\xi_{it}^s$ represents an idiosyncratic error term capturing all random shocks to the dependent variable. The exact structural form of how $\rho_{it}^s$ enters the equation is unknown, so we adopt a non-parametric specification $S^s(.)$ in which the shape of the relationship is delivered from the data and regression splines. Note that the percentage change in the estimated $S^s(.)$ with respect to the percentage change in the model covariate at any level of the covariate $\rho^s$ gives the corresponding value of agglomeration elasticity $\delta^{\rho,s}$.

We expect $\eta_{it}^s$ to be correlated with $\rho_{it}^s$. This correlation follows from the presence of omitted variables such as specific characteristics of local input and output markets, and functional or occupational differences caused by spatial self-selection by firms (see Section 1 for a detailed discussion). Further, the relationship between $\rho_{it}^s$ and productivity $\hat{\omega_{it}^s}$ may be simultaneously determined as higher productivity locations may attract a greater level of private investment over time leading to larger economic mass, which has a feedback effect on productivity. These estimation issues need to be carefully addressed to ensure that the agglomeration elasticity estimates are, as far as possible, causal rather than being simply associational. Therefore, we adopt a non-parametric instrumental variables (NPIV) regression, which not only enables non-parametric specification of $S^s(.)$ but also addresses potential endogeneity biases.

***Bayesian Nonparametric Instrumental Variable Approach***

IV-based estimators such as two-staged least square (2SLS) are widely adopted in applied econometrics to estimate parametric models that contain endogenous covariates. However, finite-dimensional parametric models (such as log-log models) for the relationship between agglomeration and productivity, are based on assumptions that are rarely justified by economic theories. The resulting model misspecification may lead to erroneous estimates of agglomeration elasticities. On the other hand, non-parametric methods have the potential to capture the salient features in a data-driven manner without making a priori assumptions on the functional form of the relationship (Horowitz, 2011). Therefore, a fairly growing strand in the econometrics literature proposes different approaches for NPIV regression, but such methods have not been considered in the estimation of the agglomeration-productivity relationship. Extensive reviews can be found in Newey & Powell (2003) and Horowitz (2011).

Classical (frequentist) NPIV regression approaches are popular in theoretical econometrics (Newey & Powell, 2003; Horowitz, 2011; Newey, 2013; Chetverikov & Wilhelm, 2017), but they are challenging to apply in practice due to two main reasons. First, tuning parameters to monitor the flexibility of $S(.)$ are often required to be specified by the analyst. Second, standard errors are generally computed using bootstrap, making these methods computationally prohibitive for large datasets. Therefore, we adopt a scalable Bayesian NPIV approach, proposed by Wiesenfarth et al. (2014), that can produce a consistent estimate of non-parametric $S(.)$, even if the analyst does not observe $\eta^s_{it}$. This Bayesian method addresses both challenges of the frequentist estimation because it *learns* tuning parameters related to $S(.)$ during estimation and uncertainty in parameters estimates is inherently captured by credible intervals (analogous to classical confidence intervals). In addition, it also enables nonparametric specification of the unobserved error component $\xi^s_{it}$, precluding the need for making additional assumptions.

We discuss the adopted Bayesian NPIV approach (Wiesenfarth et al., 2014) for a model with a single endogenous covariate, that is,

$$\hat{\omega} = S(\rho) + \epsilon_2, \quad \rho = h(z) + \epsilon_1 \tag{6}$$

Note that $\eta$ are encapsulated in $\epsilon_2$, and $z$ is an instrument for the endogenous regressor $\rho$. The relationship between $\rho$ and $z$ is represented by an unknown functional form $h(.)$ and $\epsilon_2$ is an idiosyncratic random error term. For notational simplicity, we drop the firm-year subscripts and sector superscripts. Bayesian NPIV is a control function approach, and assumes the following standard identification restrictions:

$$E(\epsilon_1|z) = 0 \quad \text{and} \quad E(\epsilon_2|\epsilon_1, z) = E(\epsilon_2|\epsilon_1), \tag{7}$$

which yields

$$\begin{aligned} E(\hat{\omega}|\rho, z) &= S(\rho) + E(\epsilon_2|\epsilon_1, z) = S(\rho) + E(\epsilon_2|\epsilon_1) \\ &= S(\rho) + \nu(\epsilon_1), \end{aligned} \tag{8}$$

where $\nu(\epsilon_1)$ is a function of the unobserved error term $\epsilon_1$. This function is known as the control function.

Conditional on the availability of a valid instrument (see Section 2), Bayesian NPIV can correct for confounding bias. To account for the nonlinear effects of continuous covariates, both $S(.)$ and $h(.)$ (refer to equation 6) are specified in terms of additive predictors comprising penalised splines. Each of the functions $S()$ and $h(.)$ is approximated by a linear combination of suitable B-spline basis functions. The penalised spline approach uses a large enough number of equidistant knots in combination with a penalty to avoid over-fitting. Moreover, the joint distribution of $\epsilon_1$ and $\epsilon_2$ is specified using nonparametric Gaussian Dirichlet process mixture (DPM), which ensures the robustness of the model relative to extreme observations. Efficient Markov chain Monte Carlo (MCMC) simulation technique is employed for fully Bayesian inference. The resulting posterior samples allow us to construct simultaneous credible bands for the non-parametric effects (i.e., $S(.)$ and $h(.)$). Thereby, the possibility of non-normal error distribution is considered and the complete variability is represented by Bayesian NPIV. We now succinctly discuss specifications of the kernel error distribution in Bayesian NPIV.

To allow for a flexible distribution of error terms, the model considers a Gaussian DPM with infinite mixture components, $c$, in the following hierarchy:

$$\begin{aligned}
(\epsilon_{1i}, \epsilon_{2i}) &\sim \sum_{c=1}^{\infty} \pi_c \mathrm{N}(\mu_c, \Sigma_c) \\
(\mu_c, \Sigma_c) &\sim G_0 = \mathrm{N}(\mu|\mu_0, \tau_{\Sigma}^{-1}\Sigma)\,\mathrm{IW}(\Sigma|s_{\Sigma}, S_{\Sigma}) \\
\pi_c = \upsilon_c &\left(1 - \sum_{j=1}^{c-1}(1-\pi_j)\right) = \upsilon_c \prod_{j=1}^{c-1}(1-\upsilon_j), \\
& c = 1, 2, ... \\
\upsilon_c &\sim \mathrm{Be}(1, \psi).
\end{aligned} \tag{9}$$

where $\mu_c$, $\Sigma_c$ and $\pi_c$ denote the component-specific means, variances and mixing proportions. The mixture components are assumed to be independent and identically distributed with the base distribution $G_0$ of the Dirichlet process (DP), where $G_0$ is given by a normal-inverse-Wishart distribution. The mixture weights are generated in a stick-breaking manner based on a Beta distribution with concentration parameter $\psi > 0$ of the DP. The concentration parameter $\psi$ determines the strength of belief in the base distribution $G_0$.

## Estimation Practicalities

We exclude discussion of the Gibbs sampler of Bayesian NPIV for brevity and focus mainly on implementation details and posterior analysis. Interested readers can refer to Wiesenfarth et al. (2014) for the derivation of conditional posterior updates.
We use the *BayesIV* and *DPpackage* in R to estimate the Bayesian NPIV. We consider 50,000 posterior draws in the estimation, exclude the first 15,000 burn-in draws and keep every $10^{th}$ draw from the remaining draws for the posterior analysis. The point-wise posterior mean is computed by taking the average of 3,500 posterior draws. Bayesian simultaneous credible bands are obtained using quantiles of the posterior draws. A simultaneous credible band is defined as the region $I_\theta$ such that $P_{S|data}(S \in I_\theta) = 1 - \theta$, that is, the posterior probability that the entire true function $S(.)$ is inside the region given the data equals to $1 - \theta$. The Bayesian simultaneous credible bands are constructed using the point-wise credible intervals derived from the $\theta/2$ and $1 - \theta/2$ quantiles of the posterior samples of $S(.)$ from the MCMC output such that $(1 - \theta)100\%$ of the sampled curves are contained in the credible band. A similar process is used to obtain the credible intervals of $h(.)$.

## Instrumental Variable

To satisfy the identification restrictions presented in equation 7, we need an instrumental variable (IV) $z$. The IV should be (i) exogenous, that is, uncorrelated with $\epsilon_2$; (ii) relevant, that is, correlated with the endogenous covariate $\rho$, conditional on other covariates in the model.
We derive valid external instruments from traffic casualty data. We consider the ratio of serious and severe traffic casualties to total casualties among active mode (pedestrians and cyclists) and motorcycle users during morning and afternoon peak hours (that is, 6:30-10:30 hours and 16:00-20:00 hours) in zone $j$ in year $t$ as an IV $z_{jt}$ for the MED (agglomeration) $rho_{jt}$ in zone $j$ and year $t$. We argue that as the MED of a city increases, peak-hour road network congestion in the city may also increase, and consequently, the average speed of travel in the network may decrease. As a result of slower vehicular speeds, the proportion of serious and severe traffic casualties to total casualties among active mode and motorcycle users during peak hours may decrease. Our argument follows from the traffic safety literature that suggests that a decrease in congestion may exacerbate the severity of peak-hour traffic casualties amongst active mode users and cyclists (Li et al., 2012; Noland et al., 2008). We thus expect a strong negative correlation between the chosen IV $z_{jt}$ and the endogenous covariate $\rho_{jt}$. Nevertheless, we argue that the chosen IV is exogenous because we do not expect the IV to scale with city size (population) and affect labour supply, and therefore, not directly determine the response variable (that is, TFP) of any firm $i$ located in zone $j$ and year $t$. In other words, we do not anticipate the chosen IV $z_{jt}$ to feature in a model for the response variable $\omega_{jt}$.

### *Data*

To gauge the presence of non-linearities in agglomeration elasticities, we consider a sample of firms in England as our case study. We consider the period between 2015 to 2019 as the study

period. For this period, we investigate the causal impact of MED and productivity in six most relevant industry sectors: Manufacturing (MAN), Construction (CON), Wholesale and Distribution (WAD), Transport (TRA), Information and Communication Technology (ICT), and Finance (FIN). As geographical regions or zones, we consider the Middle Layer 2011 Census Super Output Areas (MSOA11) in England, which includes a total of 6,791 units with a mean population of 8185 people.
The data sources for the key variables of interest are detailed in the next two subsections.

### Mean Effective Density

We obtain the data on annual employment levels in each MSOA11 unit from the Business and Employment Register available at Nomis[1] (official census and labour market statistics), a public repository maintained by the Office for National Statistics (ONS). To calculate the distance between the MSOA11 units, we extract the location information on MSOA11 units available in the ONS Postcode Directory[2], that is a detailed location database of all UK postcodes.
Traffic casualty data for the construction of IVs is obtained from the publicly available road safety data, maintained by the Department for Transport[3].

### Total Factor Productivity

The Department of Trade and Industry records all the accounting information provided by all companies registered in the UK. This information is available via the commercial software package Financial Analysis Made Easy (FAME)[4], co-hosted by Vistra and Bureau Van Dijk. To estimate the production function in equation 3, we extract the annual data on the following variables for each registered firm:

1. Turnover (output): The net income of the company.
2. Fixed Assets (capital): The depreciated value of buildings, plants and equipment.
3. Current Assets (materials): The current stocks and debt owned by the company.
4. Total Employees (labour): The total number of employees in the company.

To limit potential endogeneity biases emerging from spatial self-selection by firms (Graham & Gibbons, 2019), we remove firms with more than one trading address and those that have a registered office address different from the main trading address. We also filter out firms with international subsidiaries. Additionally, we only focus on small and medium-sized firms with a number of employees between 10 and 249 to reduce endogeneity from spatial self-selection of labour (Graham, 2009). Finally, we class the filtered data into the six industry sectors using their two-digits Standard Industrial Classification 2007 (SIC07). The resulting number of observations for each industry sector is reported in Table 1.

Table 1: Classification of firms into industry sectors.

| Industry Sector | SIC07 | Firms | Observations |
|---|---|---|---|
| MAN | 10-33 | 842 | 4210 |
| CON | 41-43 | 368 | 1840 |
| WAD | 45-47 | 688 | 3440 |
| TRA | 49-56 | 246 | 1230 |
| ICT | 58-63 | 357 | 1785 |
| FIN | 64-74 | 1452 | 7260 |

## 3 Results and Discussion

This section is divided into four subsections. In the first subsection, we describe our MED estimates for England. In the second subsection, we briefly visit the estimated parameters of the production function for various industry sectors and the estimated TFP values. In the penultimate subsection,

[1] Availble at `https://www.nomisweb.co.uk/`.
[2] Available at `https://geoportal.statistics.gov.uk/`.
[3] Available at `https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data`.
[4] Available at `https://fame.bvdinfo.com/version-202274/fame/1/Companies/Search`.

we discuss the agglomeration elasticities obtained from the Bayesian NPIV estimation. Using these estimates, we also describe the spatial distribution of agglomeration benefits in England. In the final subsection, we present the estimated kernel error distributions to illustrate the importance of the non-parametric DPM specification. The relevance of our instruments is also demonstrated in this subsection.

***Estimated Mean Effective Densities***

Table 2 presents the summary statistics for the estimated MED values for each MSOA11 in England. The table indicates that the distribution of MED values in England is positively skewed. While some zones have high levels of MED or agglomeration, most zones have low values. Thus, only a few zones in England show high values of agglomeration. This observation is further supported by Figure 1, which maps MED values constructed using total employment in 2019 as mass. This figure illustrates that whereas regions in and around cities like London, Manchester, and Birmingham correspond to higher levels of agglomeration, they only constitute a small geographical area in England.

Table 2: Summary of estimated MED for England.

| Statistic | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|
| Mean | 3934.64 | 4008.44 | 4071.00 | 4091.90 | 4154.83 |
| Median | 3375.18 | 3435.73 | 3480.50 | 3502.10 | 3550.63 |
| Std. dev. | 2272.62 | 2328.46 | 2374.13 | 2397.93 | 2450.14 |
| Max | 19706.27 | 20297.82 | 20745.36 | 21064.78 | 21694.94 |
| Min | 663.97 | 675.28 | 684.24 | 687.70 | 697.25 |
| Skewness | 2.43 | 2.44 | 2.44 | 2.48 | 2.50 |

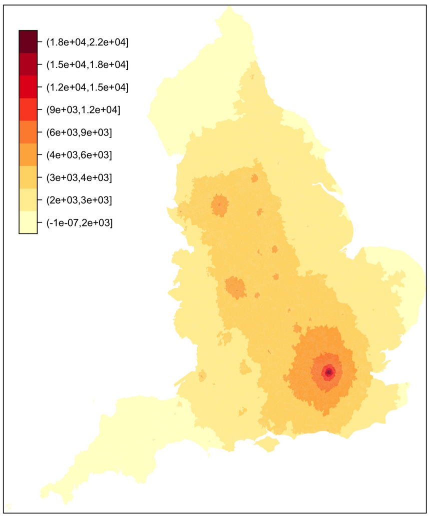



Figure 1: Map of MED values for England with total employment in 2019 as mass.

We complement the above-described statistics with Figure 2, which provides a histogram of the MED levels experienced by firms in each industry sector. We note from this figure that the majority of firms in the CON, MAN, TRA, and WAD sectors tend to locate in areas with MED values less than 6000. From a strategic point of view, firms in these sectors require large factories or warehouses. They may, therefore, prefer to locate these facilities in the periphery of cities where land prices, rents, and other costs are lower. Nonetheless, additional local maxima in their density plots at higher levels of MED also reveals the presence of a small number of firms in city centers, which may choose to locate their offices in central business district (CBD) for an easy commitment and location status. Conversely, firms in the ICT and FIN sectors, primarily tend to spread across the CBD, to avail the above-mentioned advantages. In the rest of this section, we quantify how these location choices translate into productivity benefits for firms.

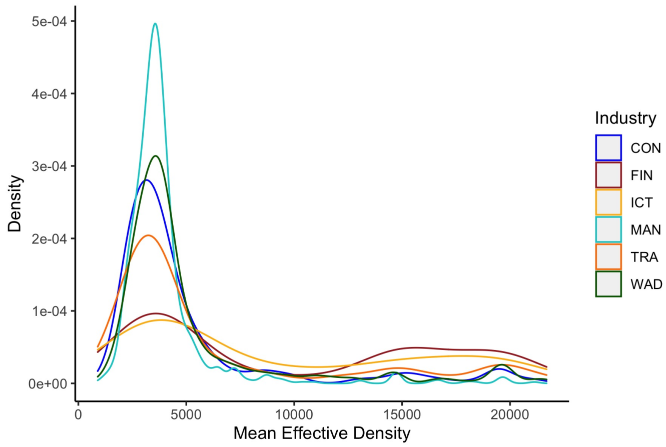



Figure 2: Histogram of agglomeration levels experienced by firms in 2019.

***Estimated Total Factor Productivity***

The parameter estimates of the production function, given by equation 3, for the six industry sectors are summarised in Table 3. We note that the FIN, ICT, MAN and WAD sectors are associated with returns to scale (RTS) values less than one, that is, decreasing RTS indicating that their output increases by less than the proportional change in all inputs. On the contrary, the CON and TRA sectors are associated with increasing RTS, implying a more than proportional increase in outputs with respect to inputs. These estimates summarise the technological advantages or disadvantages to firms in each industry.

Table 3: Parameter Estimates of the Production Function.

| Sector | No. of firms | $\beta_l$ | $\beta_k$ | $\beta_m$ | RTS |
|---|---|---|---|---|---|
| CON | 1841 | 0.695 | 0.467 | -0.040 | 1.123 |
| | | (0.010) | (0.014) | (0.012) | (0.021) |
| FIN | 7261 | 0.481 | 0.477 | 0.032 | 0.990 |
| | | (0.006) | (0.008) | (0.010) | (0.014) |
| ICT | 1786 | 0.323 | 0.630 | 0.006 | 0.959 |
| | | (0.008) | (0.008) | (0.020) | (0.023) |
| MAN | 4211 | 0.332 | 0.600 | 0.020 | 0.953 |
| | | (0.004) | (0.012) | (0.012) | (0.014) |
| TRA | 1231 | 0.439 | 0.539 | 0.026 | 1.005 |
| | | (0.012) | (0.050) | (0.016) | (0.054) |
| WAD | 3441 | 0.304 | 0.692 | -0.011 | 0.985 |
| | | (0.002) | (0.002) | (0.002) | (0.003) |

Table 4 presents the summary statistics for the estimated TFP values for each industry sector. The mean and median statistics suggest that firms in the WAD sector are associated with the highest level of unobserved productivity, closely followed by firms in the MAN and ICT sectors. The sectors FIN, TRA, and CON show lower levels of unobserved productivity.

Table 4: Summary of estimated TFP values.

| Statistic | CON | FIN | ICT | MAN | TRA | WAD |
|---|---|---|---|---|---|---|
| Min | -2.11 | -1.29 | 2.35 | -2.09 | 1.32 | 5.37 |
| Median | 6.93 | 7.41 | 8.02 | 8.07 | 7.26 | 8.80 |
| Mean | 6.94 | 7.45 | 8.04 | 8.10 | 7.19 | 8.90 |
| Max | 10.38 | 14.19 | 13.18 | 12.89 | 12.48 | 13.48 |
| Std. dev. | 0.90 | 1.09 | 1.084 | 0.74 | 1.11 | 0.92 |

***Results from Bayesian NPIV Estimation***

Figure 3 presents the estimates of $S(.)$ (see equation 6, second-stage) for the six industry sectors. The plots include the mean estimates and the 95 percent credible bands (shown by the dotted line). The density of tick marks along the X-axis represents the number of observations in the corresponding domain of agglomeration. The figure indicates the presence of significant non-linearities in the agglomeration-productivity relationship. This observation validates our hypothesis that presuming a log-log functional form may yield biased estimates of this relationship and associated agglomeration elasticities.

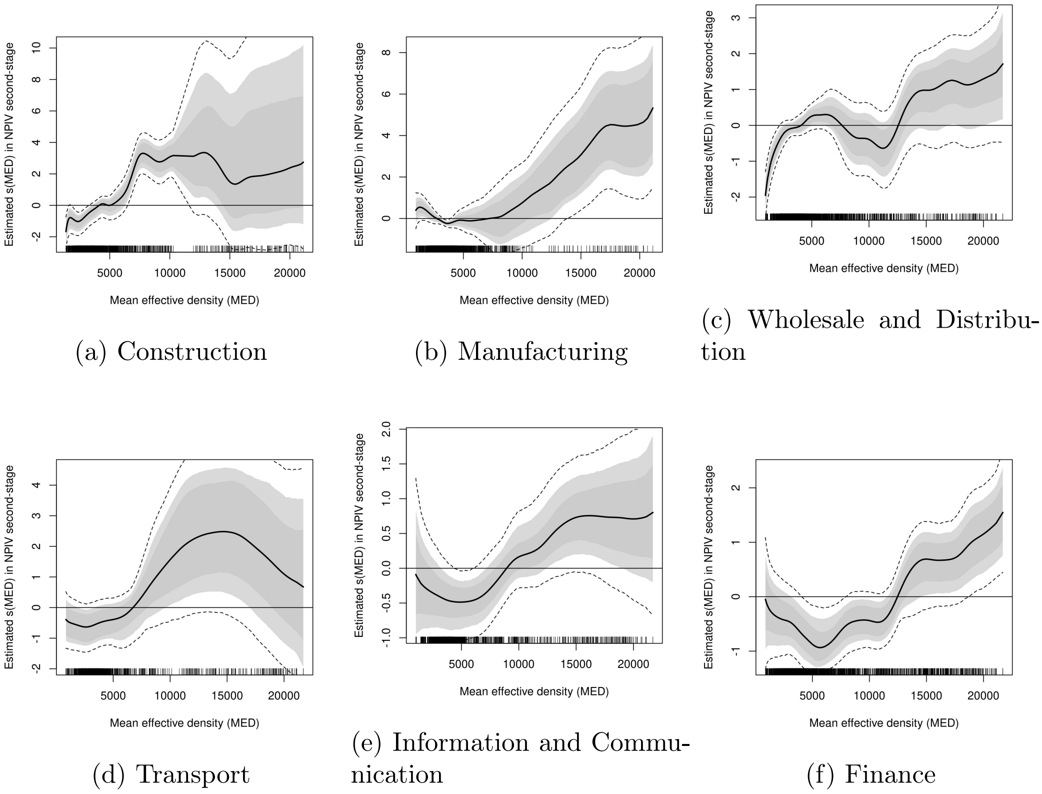
In Figure 4, we plot the agglomeration elasticities obtained at different levels of agglomeration and their corresponding credible bands. Note that these estimates are obtained using the 3,500 posterior draws (refer to Section 2) that are available at multiple points along the support of the model covariate, that is, MED. To obtain the elasticity at any point $\rho$, we identify a small interval $[\rho_1, \rho_2]$ surrounding $\rho$, where $\rho_1 = \rho - \Delta\rho$ and $\rho_2 = \rho + \Delta\rho$. We extract the 3,500 posterior draws at the two points $\rho_1$ and $\rho_2$ and calculate the change in elasticity for each draw. The mean of the resulting 3,500 samples gives the reported elasticity estimate at $\rho$ and the quantiles 0.025 and 0.975 give the corresponding lower and upper limits of the 95-percent credible bands.

For the CON sector, we note that agglomeration elasticities remain positive for MED levels below 8000, with values increasing from 0.15 to 0.72 and back to 0.48, and become statistically insignificant beyond that. For the MAN sector, we observe that the agglomeration elasticities are statistically significant between MED levels of 8000 to 1000 and 13000 to 18000, and statistically insignificant otherwise. The estimated elasticities increase from 0.19 to 0.78 until a MED level of 16000 and drop back to 0.17 before becoming statistically insignificant. Our estimates for the WAD sector suggest that the agglomeration elasticities fall from a value of 0.16 at MED level 2000 to a value of -0.20 at MED level 8000, while remaining positive (and statistically significant) between MED levels of 2000 to 5000, become negative (and statistically significant) between MED levels of 7000 to 10000. The agglomeration elasticities become positive (and statistically significant) again at a MED level of 11000, followed by a steep increase to the value of 0.75 at a MED level of 13000. The agglomeration elasticities, thereafter, fall to a value of 0.22 at a MED level of 15000, beyond which they become statistically insignificant. The estimated agglomeration elasticities for the TRA sector remain positive and statistically significant between MED levels of 6000 to 12000, and range between 0 to 0.50, the maximum being achieved at a MED level of 9000. Higher levels of agglomeration of the order of 18000 to 20000 MED are associated with negative (and statistically significant) values of agglomeration elasticities ranging between -0.75 to -0.50. Firms in the ICT sector are found to be associated with positive agglomeration elasticities in the interval [0.12,0.22] at agglomeration levels between 7000 to 15000 MED, the maximum being observed at a MED level of 13000. The agglomeration elasticities remain statistically insignificant otherwise. Finally, for the FIN sector, we first observe negative and statistically significant agglomeration elasticities between MED levels of 3000 to 5000. The estimated elasticities are of the order of -0.1. Nonetheless, the elasticities remain positive (and statistically significant) at MED levels between 7000 to 9000, 11000 to 1500, and also beyond MED levels of 18000. The positive values range in the interval [0.18,0.60], with the maximum occurring at a level of 13000 MED. Overall, Figure 4 indicates that barring the CON sector, the productivity benefits in all sectors comment into effect beyond a critical mass of agglomeration. This critical mass varies across industries.

Table 5 summarises the estimated agglomeration elasticities. The final column in the table reports the estimates from a one-step procedure where MED enters as a covariate in the production function (equation 3). The values in the final column are fairly consistent with the literature in which elasticity estimates have been derived by assuming the productivity-agglomeration relationship to be log-log (see Graham & Gibbons (2019) for a summary of 47 international empirical studies on the effects of agglomeration on productivity). Our results suggest that our non-linear agglomeration elasticity estimates take more extreme values compared to their log-log counterparts.

Next, we map the estimated agglomeration elasticities to the different zones (that is, MSOA11 units) in England using their MED values. This mapping allows us to understand the spatial distribution of the agglomeration impacts in England. Figure 5 shows these distributions for each industry sector. Note that we adopt the same colour key for each map in Figure 5 to allow comparison of agglomeration impacts across industries.

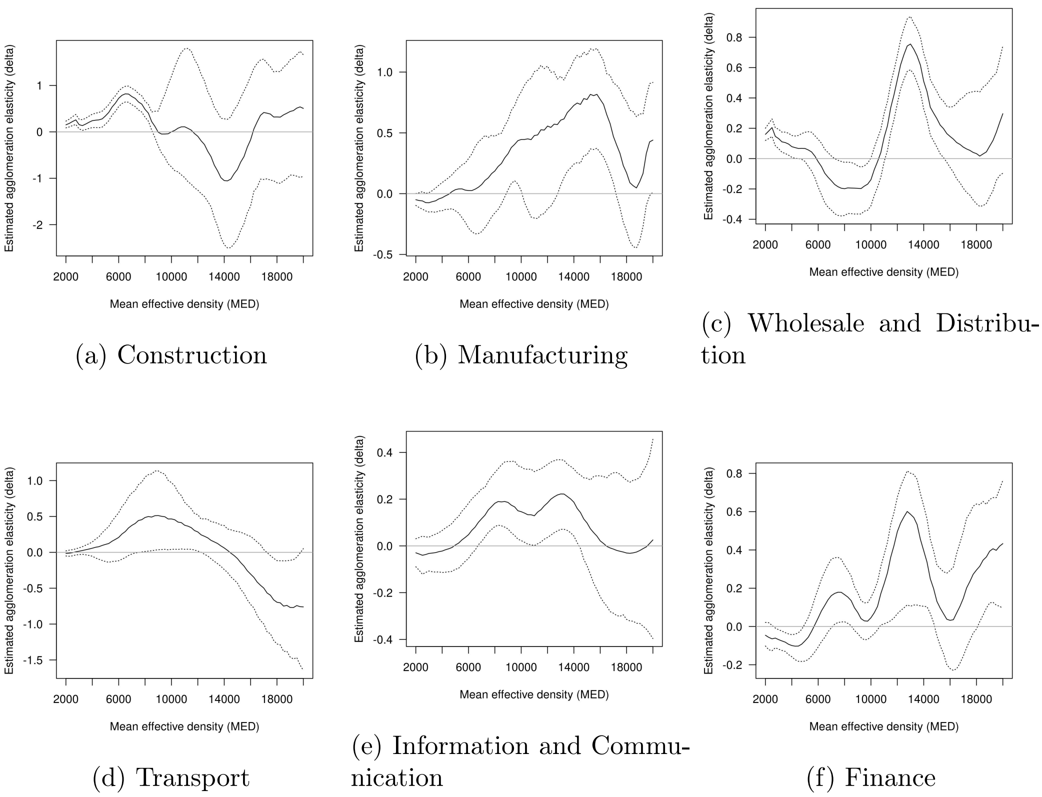
From Figure 5a, we note that the highest levels of agglomeration benefits (elasticities ranging in the interval [0.4,0.9]) in the CON sector can be observed in the peripheral regions of the Greater London Area (GLA) (for instance, Slough, Watford, and Loughton) and within cities of Manchester and Birmingham. Interestingly, the areas within the GLA are associated with statistically insignificant productivity effects of agglomeration. All other MSOA11 units are associated with agglomeration elasticities ranging between 0.1 and 0.3. Figure 5b suggests that the significant agglomeration benefits in the MAN sector remain confined within the GLA, while remaining the highest in the regions immediately surrounding the CBD of London (which includes the City of London, City of Westminster and Kensington and Chelsea, among others). Figure 5c indicates the firms in the WAD sector avail the highest benefits of agglomeration on their productivity by locating in the outskirts of the CBD of London. The impacts within the CBD of London and cities such as




(a) Construction (b) Manufacturing (c) Wholesale and Distribution

(d) Transport (e) Information and Communication (f) Finance

Figure 3: Estimated relationships between agglomeration and firm productivity.



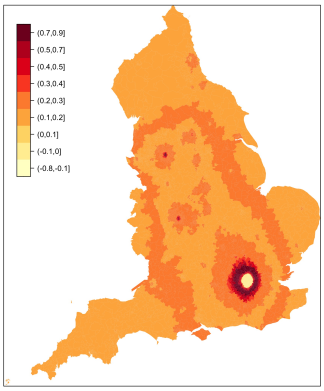

(a) Construction (b) Manufacturing (c) Wholesale and Distribution

(d) Transport (e) Information and Communication (f) Finance

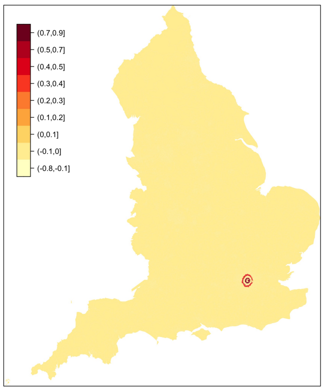
Figure 4: Estimated agglomeration elasticities.

Table 5: Summary of estimated agglomeration elasticities.

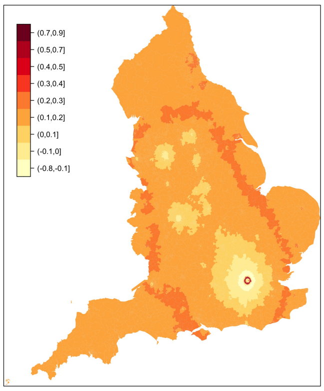
| Sectors | Mean Effective Density (MED) | | | | | | | | | One-step |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2000 | 4000 | 6000 | 8000 | 10000 | 12000 | 14000 | 16000 | 18000 | estimate |
| CON | 0.15 | 0.24 | 0.72 | 0.48 | 0 | -0.1 | -1.05 | -0.17 | 0.32 | 0.045 |
| | (0.04) | (0.07) | (0.09) | (0.09) | (0.48) | (0.69) | (0.68) | (0.68) | (0.57) | (0.009) |
| MAN | -0.05 | -0.03 | 0.03 | 0.19 | 0.45 | 0.56 | 0.73 | 0.78 | 0.17 | -0.002 |
| | (0.02) | (0.06) | (0.15) | (0.16) | (0.21) | (0.32) | (0.22) | (0.2) | (0.3) | (0.005) |
| WAD | 0.16 | 0.08 | -0.02 | -0.2 | -0.14 | 0.52 | 0.51 | 0.15 | 0.03 | -0.067 |
| | (0.02) | (0.04) | (0.08) | (0.1) | (0.08) | (0.08) | (0.09) | (0.1) | (0.19) | (0.002) |
| TRA | -0.02 | 0.06 | 0.21 | 0.49 | 0.47 | 0.28 | 0.06 | -0.3 | -0.7 | 0.115 |
| | (0.02) | (0.05) | (0.17) | (0.28) | (0.23) | (0.12) | (0.15) | (0.23) | (0.3) | (0.015) |
| ICT | -0.03 | -0.02 | 0.05 | 0.18 | 0.15 | 0.18 | 0.19 | 0.02 | -0.03 | 0.072 |
| | (0.03) | (0.04) | (0.06) | (0.07) | (0.07) | (0.08) | (0.07) | (0.13) | (0.15) | (0.009) |
| FIN | -0.05 | -0.1 | 0.04 | 0.17 | 0.05 | 0.5 | 0.4 | 0.03 | 0.3 | 0.144 |
| | (0.03) | (0.03) | (0.08) | (0.08) | (0.05) | (0.15) | (0.13) | (0.13) | (0.16) | (0.003) |



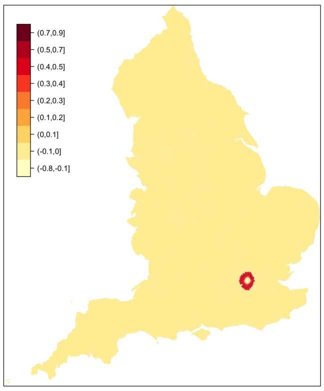
(a) Construction



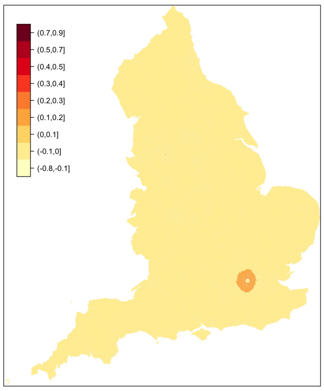
(b) Manufacturing



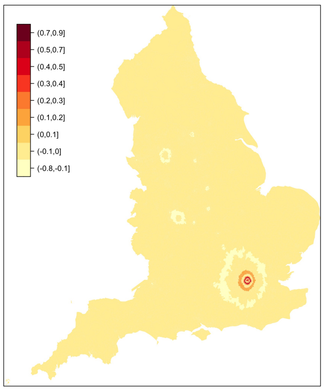
(c) Wholesale and Distribution



(d) Transport



(e) Information and Communication



(f) Finance

Figure 5: Spatial distribution of the agglomeration impacts in England in 2019.

Birmingham and Manchester either remain negative or statistically insignificant. Other regions show positive impacts of varying degrees as represented by the color key. Similar to the MAN sector, firms in TRA and ICT sectors (Figures 5d and 5e) observe agglomeration benefits in zones surrounding the CBD of London. From Figure 5f, we note that firms in the FIN sector derive the highest agglomeration benefit by locating within the CBD of London. Additionally, while most areas within the GLA observe positive productivity impacts of agglomeration, the peripheral areas of the GLA, Birmingham, and Manchester are associated with diseconomies of agglomeration. The effect in other areas remains statistically insignificant.

## 4 Conclusions

Transport accounted for 5 to 8 percent of the total public expenditure on services in the UK through the years 2017 to 2022[5]. In a typical year, the United Kingdom spends about £45 billion pounds on construction and maintenance of transport infrastructure. Understanding the economic and social benefits arising from investments of this order is thus important for policymakers. This paper contributes to the growing strand in the urban economic literature that focuses on measuring the Wider economic impacts (WEIs) of transport investments arising via scale economies of agglomeration.

We make two advances in this agenda. First, we develop a causal statistical framework to quantify the non-linearities in the relationship between agglomeration (represented by Mean Effective Density (MED)) and productivity (measured as Total Factor Productivity (TFP)). The estimated relationships, for the first time, provide a quantification of how agglomeration elasticities vary over different levels of agglomeration. Second, we determine a novel external instrument derived from traffic casualty data to identify the agglomeration-productivity relationship. Our study suggests the use of the severity of traffic casualties among active mode users and motorcyclists during peak hours as a relevant and exogenous instrument for agglomeration.

Our investigation of agglomeration elasticities in six key industry sectors in England suggests that agglomeration elasticities vary significantly vary over agglomeration levels. For the Construction Sector, we observe positive agglomeration elasticities only at low and mid levels of agglomeration. For the other five sectors which include, Manufacturing, Wholesale and Distribution, Transport, Information and Communication Technology, and Finance, we note the presence of a critical mass of agglomeration beyond which the positive benefits of agglomeration on productivity can be observed. Below this critical level, the agglomeration elasticities either remain negative (but statistically significant) or statistically insignificant at the 95-percent confidence level. Additionally, we note that agglomeration elasticities in the Transport sector become negative at extremely high levels of agglomeration, while very low levels of agglomeration show positive agglomeration elasticities for the Wholesale and Distribution sector, but of lower magnitude. Interestingly, the estimated agglomeration elasticities in this study take more extreme values than ones derived from a log-log model of productivity and agglomeration as adopted in the literature. Our estimates thus have crucial implications for the appraisal of transport investments.

Further, our exploration of the spatial distribution of the agglomeration impacts in England reveals that the highest levels of agglomeration benefits in the Construction sector are observed in the regions surrounding the Greater London Area (GLA) and within Manchester and Birmingham. For the Manufacturing, Wholesale and Distribution, Transport and Information and Communication Technology sectors, the largest productivity benefits of agglomeration are confined within the GLA, particularly along the fringes of its central business district (CBD). For the Finance Sector, the highest positive agglomeration elasticities are associated with the regions in the CBD of the GLA, while the outskirts of the GLA, Manchester and Birmingham see diseconomies of agglomeration. Our findings are unsurprising: these spatial patterns are consistent with the sector-wise preferences for office locations by firms. Refining this investigation with more data, particularly from other years or from other countries, is an important topic for further research and can provide an empirical basis for targetting transport investments in a manner that can spread productivity benefits more evenly.

## Acronyms

**2SLS** two-staged least square

**ATEM** access to economic mass

**CBA** Cost Benefit Analysis

**CBD** central business district

**CF** control function

**CON** Construction

**DP** Dirichlet process

**DPM** Dirichlet process mixture

[5] `https://www.gov.uk/government/statistics/public-expenditure-statistical-analyses-2022`

**DUBs** Direct user benefits

**FAME** Financial Analysis Made Easy

**FIN** Finance

**GLA** Greater London Area

**ICT** Information and Communication Technology

**IV** instrumental variable

**IVs** instrumental variables

**MAN** Manufacturing

**MCMC** Markov chain Monte Carlo

**MED** Mean Effective Density

**MSOA11** Middle Layer 2011 Census Super Output Areas

**NP** non-parametric

**NPIV** non-parametric instrumental variables

**ONS** Office for National Statistics

**RTS** returns to scale

**SIC07** Standard Industrial Classification 2007

**TAG** Transport Analysis Guidance

**TFP** Total Factor Productivity

**TRA** Transport

**WAD** Wholesale and Distribution

**WEIs** Wider economic impacts

## References


Ackerberg, D., Caves, K., & Frazer, G. (2006). Structural identification of production functions.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, *83*(6), 2127–2189.

Au, C.-C., & Henderson, J. V. (2006). Are Chinese cities too small? *The Review of Economic Studies*, *73*(3), 549–576.

Baldwin, J. R., Brown, W. M., & Rigby, D. L. (2010). Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science*, *50*(5), 915–934.

Chetverikov, D., & Wilhelm, D. (2017). Nonparametric Instrumental Variables Estimation under Monotonicity. *Econometrica*, *85*(4), 1303-1320.

Combes, P.-P., Duranton, G., Gobillon, L., & Roux, S. (2012). Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics*, *42*(6), 913–930.

Combes, P.-P., & Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics* (Vol. 5, pp. 247–348). Elsevier.

De Loecker, J. (2007). Do exports generate higher productivity? Evidence from Slovenia. *Journal of international economics*, *73*(1), 69–98.

DfT. (2016). TAG UNIT A2.4: Appraisal of Productivity Impacts.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063–2117). Elsevier.

Graham, D. J. (2009). Identifying urbanisation and localisation externalities in manufacturing and service industries. *Papers in Regional Science*, *88*(1), 63–84.

Graham, D. J., & Dender, K. V. (2011). Estimating the agglomeration benefits of transport investments: some tests for stability. *Transportation*, *38*(3), 409–426.

Graham, D. J., & Gibbons, S. (2019). Quantifying Wider Economic Impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation*, *19*, 100121.

Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P., & Noland, R. B. (2010). Testing for causality between productivity and agglomeration economies. *Journal of regional Science*, *50*(5), 935–951.

Horowitz, J. L. (2011). Applied Nonparametric Instrumental Variables Estimation. *Econometrica*, *79*(2), 347-394.

Lall, S. V., Shalizi, Z., & Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. *Journal of development economics*, *73*(2), 643–673.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, *70*(2), 317–341.

Li, H., Graham, D. J., & Majumdar, A. (2012). The effects of congestion charging on road traffic casualties: A causal analysis using difference-in-difference estimation. *Accident Analysis & Prevention*, *49*, 366–377.

Mackie, P., Graham, D., & Laird, D. (2012). Direct and wider economic benefits in transport appraisal. In *Handbook in transport economics. edward elgar, london* (pp. 501–526).

Maré, D. C., & Graham, D. J. (2009). Agglomeration elasticities in New Zealand. *Motu Working Paper No. 09-06*.

Maré, D. C., & Graham, D. J. (2013). Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics*, *75*, 44–56.

Marrocu, E., Paci, R., & Usai, S. (2013). Productivity growth in the old and new Europe: the role of agglomeration externalities. *Journal of Regional Science*, *53*(3), 418–442.

Marshall, A. (1920). *Principles of Political Economy* (8th ed.). London, Macmillan.

Melo, P. C., Graham, D. J., & Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional science and urban Economics*, *39*(3), 332–342.

Morikawa, M. (2011). Economies of density and productivity in service industries: An analysis of personal service industries based on establishment-level data. *The Review of Economics and Statistics*, *93*(1), 179–192.

Newey, W. K. (2013). Nonparametric Instrumental Variables Estimation. *American Economic Review*, *103*(3), 550-556.

Newey, W. K., & Powell, J. L. (2003). Instrumental Variables Estimation of Nonparametric Models. *Econometrica*, *71*(5), 1565-1578.

Noland, R. B., Quddus, M. A., & Ochieng, W. Y. (2008). The effect of the London congestion charge on road casualties: an intervention analysis. *Transportation*, *35*(1), 73–91.

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*(2), 373–389.

Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, *26*(1), 98–128.

Venables, A., Laird, J. J., & Overman, H. G. (2014). Transport investment and economic performance: Implications for project appraisal.

Venables, A. J. (2007). Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy (JTEP)*, *41*(2), 173–188.

Wiesenfarth, M., Hisgen, C. M., Kneib, T., & Cadarso-Suarez, C. (2014). Bayesian nonparametric instrumental variables regression based on penalized splines and dirichlet process mixtures. *Journal of Business & Economic Statistics*, *32*(3), 468–482.