UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 177 - Jan 12

Simulating the Spatial Distribution of Employment in Large Cities: with Applications to Greater London

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Simulating the Spatial Distribution of Employment in Large Cities: with Applications to Greater London

Duncan A. Smith, Camilo Vargas-Ruiz, & Michael Batty
duncan.a.smith@ucl.ac.uk, camilo.ruiz@ucl.ac.uk, & m.batty@ucl.ac.uk

Centre for Advanced Spatial Analysis (CASA)
University College London (UCL)
90 Tottenham Court Road, London WC1E 6BT, UK
http://www.casa.ucl.ac.uk/

2 January 2012

Abstract

In this chapter, we first review the development of employment location models as they have been developed for integrated models of land use transportation interaction (LUTI) where the focus is on the allocation of population and employment. We begin by sketching how employment models based on input-output and multiplier relationships are used to predict future employment aggregates by type and then we illustrate how these aggregates are distributed to small zones of an urban region in ways that make them consistent with the distribution of population and service employment allocated using spatial interaction-allocation models. In essence, the structure we are developing, which is part of an integrated assessment of resilience to extreme events, links input-output analysis to the allocation of employment and population using traditional land use transportation interaction models. The framework then down scales these activities which are allocated to small zones to the physical level of the city using GIS-related models functioning at an even finer spatial scale.

The crucial link in this chain is how we distribute detailed employment types generated from the input-output model to small zones consistent with the way population and services are allocated using the LUTI model. To achieve this, we introduce an explicit employment forecasting model in which employment types are related to functions of floorspace that they use. These are estimated using linear regression analysis which enables future predictions of their location to be scaled in proportion to the totals generated from the input-output framework. These future estimates of floorspace condition the supply side of the model, and combined with accessibility indicators, provide the heart of the employment location model. We have developed this model for London and its region – south east England – and after presenting the results of the model, we sketch how the integrated framework is being used to generate scenarios for future employment and population change.

Acknowledgements

This research was part funded by the ESRC Genesis Project (RES-149-25-1078), and the EPSRC Scale (EP/G057737/1) and Arcadia Projects (EP/G060983/1)

Introduction

Integrated land use transport interaction (LUTI) models originating from the notion that the spatial distributions of different geodemographic and economic activities define the way various processes of location operate in space. These are usually composed of activities that are coupled together in pairs with one activity determining another. These couplings form chains of activities that are sometimes entirely looped with no one activity taking precedence in the simulation, thus illustrating how urban activities determine one another simultaneously, but more usually, this chain is broken in some way, with specific activities being defined exogenously. In examining urban structure, it is not easy to determine where the causal chain of dependence can be broken but usually integrated land use transport models, define certain categories of employment, rather than population as being exogenous to the simulation, thus determining these variables in advance as initial conditions that drive the simulation. One of the most obvious distinctions is between employment which is clearly associated with export-orientated activities, and employment that is entirely dependent on population and other employment in the city. Export-orientated is usually difficult to predict with respect to the internal processes of location within the city, hence often being specified as exogenous to the simulation. This long-standing division is sometimes referred to as a distinction between 'basic' and 'non-basic' employment, which fifty years ago was the division between primary-manufacturing and service employment, but any division into activities that cannot be easily predicted and those which are easier to predict in terms of the particular city in question, are relevant to this definition.

Since these types of spatially disaggregate models were first developed following Lowry's (1964) pioneering model of Pittsburgh, there has been continued pressure to relax this distinction between exogenous and endogenous employment. Almost as soon as such model structures were generalised, the notion of building predictive models of different employment categories, traditionally considered exogenous, was broached. Putman (1983, 1991) was the first to develop explicit models of employment location which were embedded in land use transportation model structures, developing his EMPAL (**EMP**loyment **Al**location) suite of models that essentially fused spatial interaction with land development, based on econometric estimation with the use of accessibility potential functions reflecting spatial interaction in the manner first used by Hansen (1959). As with many LUTI models which have gradually moved away from their strict comparative static origins, the EMPAL model is now set within a semi-dynamic incremental framework that involves variables lagged in time but most models of this kind are still largely calibrated using trial and error estimation rather than within any strict econometric framework. More recently, individual firm location models based on agent-based and microsimulation approaches have found favour in the context of LUTI models; but progress has been slow due to the immense data requirements required for such models and problems over providing rich enough input data for scenario testing (see Maoh and Kanaroglou, this volume, and Moeckel, this volume).

Here we will focus on building an employment location model that is deeply embedded in the LUTI model where it is used to generate locational distributions that are used in other sectors such as residential and services location. This is then nested within a wider process of integrated modelling that embeds the LUTI model within a sequence that begins with demographic prediction, moves to aggregate employment prediction using input-output analysis and then predicts the employment distributions using simple regression. The results of this sequence are then used in the LUTI model which in turn generates predictions of population which are further scaled down to greater spatial detail using physical land development modelling based on GIS. The integrated assessment model was first developed for a project involving the evaluation of the impacts of long term climate change in the London region on the distribution of population and employment, particularly with respect to sea level rise over the next 100 years. The various models involved in the sequence were designed and built by different groups with commensurate expertise in demographics, input-output modelling, LUTI modelling (which represents our own contribution) and fine scale GIS[1] (Walsh, et al., 2011). The integrated assessment is now being simplified and continued with a stronger focus on assessing resilience not only over the long term but over the short term where

[1] The consortium involved in building the integrated suite of models comprised the University of Newcastle upon Tyne Earth Sciences group, responsible for the overall project, and for the transportation networks, GIS and flooding models, Cambridge Econometrics (University of Cambridge Land Economy Environmental group) responsible for the input-output model, and ourselves at CASA-UCL responsible for the LUTI model. Other groups at the Leeds University Institute of Transport Studies, Loughborough University Transport Group, and Manchester University Centre for Urban Regional Ecology were responsible for parallel pollution and energy studies. The project was organised under the auspices of the Tyndall Centre for Climate Change see http://www.tyndall.ac.uk/sites/default/files/engineeringcities.pdf

the emphasis is on the fracturing of urban networks and constraints on location of which climate and energy change are significant parts[2].

Here we will focus on the link between the input-output and the LUTI models, demonstrating the way we disaggregate the aggregate employment predictions for each employment type which emanate from the input-output model. These are then input to the LUTI model and various feedback loops are necessitated to ensure consistency and balance. In the next section, we discuss the sequence of aggregate, thence spatially disaggregate prediction in terms of the sequence of models used. We formally specify these models in generic and more formal terms, and then we develop the employment forecasting model that links the input-output and LUTI models. We calibrate this model to the London region (which nests the region used for the LUTI model – Greater London and the outer metropolitan area – within South East England) and we present a critique of the results. We then show how this model is used to couple the input-output and LUTI models together, commenting on potential feedback loops. We finally sketch possible scenarios with the model which are under current development and we conclude with an assessment of the state of the art in such modelling and suggest how this might be improved.

The Integrated Model Framework

The Generic Structure

In essence, the model sequence takes aggregate estimates of employment E and population P for the region, and first disaggregates these into m employment types E_m using demand from the population H_m for employment in these types. It organises these within an input-output framework and then allocates each employment type E_{im} to zones i of the urban region using the employment location model which is the main focus of this chapter. From these predictions, E_{im}, it simulates the location of populations P_j associated with these employments using spatial interaction models which are coupled together to form the land use transportation interaction (LUTI)

[2] ARCADIA: **A**daptation and **R**esilience in **C**ities: **A**nalysis and **D**ecision making using **I**ntegrated **A**ssessment, http://www.ukcip-arcc.org.uk/content/view/628/9/

model. This sequence is illustrated in the block diagram shown in Figure 1 which also indicates a number of key feedback loops that we will discuss in the following description. Before we do so however, we will describe the generic process that is implied in this structure.

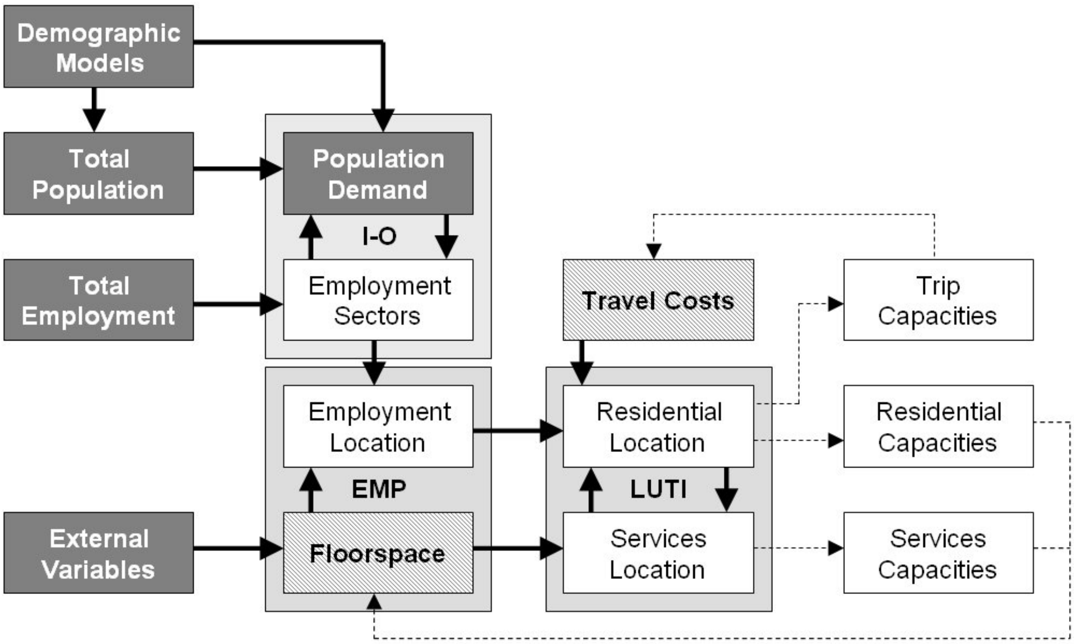


Figure 1: The Integrated Model Composed of Input-Output (I-O), Employment Location (EMP), and Land Use Transportation Interaction (LUTI) Models

The main drivers of the model framework are aggregate quantities of employment and population with population broken down into household demand associated with employment types m. These are illustrated by the dark boxes in the block diagram in Figure 1. They are linked to each other through the usual input-output coefficients that determine how much of one employment type m is used in the creation of another n and the resulting outputs – employment types E_m – are then input to the employment location model. The supply side of the urban system is driven by floorspace and travel cost variables which determine the attraction of locations and the deterrence to travelling between them. The employment models generate employment types in small zones E_{im} and these are distributed to residential zones using the residential location model which in turn generates population in zones j, P_j, from which the demand for service employment is generated. This is a final demand, consistent with one type of employment in the input-output model. This is constrained to the total required but all the service demand model does is distribute this quantity to the zones of the system.

The population and service demand models form the heart of the LUTI model although it might be argued that the employment model is also part of this extended framework despite the fact that in Figure 1, we show them as distinct from one another. There are many possible feedbacks in the system. First there are land and floorspace constraints on the generation of activities. Once employment and population have been predicted, it is possible to work out their associated land and floorspace requirements. If floorspace differs as it is likely to do so, it is possible to change the floorspace inputs to the employment and LUTI models and reiterate this sequence. If land supply constraints are breached which in turn are fixed by the subsequent GIS model which translates the employment and population activities to a finer spatial scale, then these models are also reiterated. Trip distributions are also central to the logic and these are assigned to the underlying networks; if capacities are breached in terms of such assignments, then travels costs are changed to reflect congestion and the model is reiterated to meet these constraints. These various loops, the exogenous (dark blocks in Figure 1) and endogenous variables (light blocks in Figure 1) and those that act in both ways (the stippled blocks in Figure 1) indicate the flow of simulation in Figure 1 with respect to various components of the integrated framework. The iterative loops that are used to balance the structure and bring the three models (I-O, Employment Location, and LUTI) into equilibrium are given by the broken lines and arrows.

The Input-Output Structure

Aggregate demand for population or employment is usually simulated using a linear structure in which exogenous demand X drives endogenous demand Y which is usually taken as a function of total or final demand Z. We can write this equation as $Z = X + Y$ and if we consider the function of total demand to be $Y = \lambda Z$, we are able to write the equation for total demand as

$$Z = X + \lambda Z \qquad , \qquad (1)$$

which simplifies to

$$Z = X(1-\lambda)^{-1} \quad . \qquad (2)$$

$(1-\lambda)^{-1}$ is the multiplier effect that scales exogenous to total demand. There is nothing in this structure that tells us what is total, exogenous or endogenous demand. It depends on the formulation of the problem. In traditional LUTI models, which are embedded within a loose input-output structure, demand is usually defined in terms of employment activities with Z as total, X export-orientated or basic, and Y services or non-basic employment dependent on the level of total employment. However in a traditional input-output framework, it is often argued that the exogenous input which drives the model is the demand by the population for employment which is usually called final or household demand, with the employment being that supplied to meet this demand. This difference in orientation implies that such linear input-output structures can be oblique to one other, and their precise form depends upon the way the system and problem are articulated.

In fact, the models in this integrated structure are what we might call 'loosely-coupled' in that different groups are responsible for their design and construction. The input-output model is a conventional structure where household demand drives employment structure rather than exogenous inputs of employment driving the resulting total employment. The model is part of a suite developed by Cambridge Econometrics and the one used for this application is disaggregated by UK regions with the East, London, and South East England representing the focus for the predicted employment structure[3] (Junankar, Lofsnaes and Summerton, 2007). It has not been possible to restructure the input-output model to reflect exogenous employment but in any case, in this application, it is preferable to consider future changes in household demand as having more significance with respect to the climate change scenarios which lie at the core of the application. Moreover it is somewhat easier to predict household demand than employment demand. Although our treatment here is generic and not quite the precise form that is used, the model is structured along the following lines.

[3] http://www.camecon.com/ModellingTraining/suite_economic_models/MDM-E3/MDM-E3_overview.aspx

The model is formulated in terms of predicting total demand for employment in category m, E_m, from the final household demand for employment in each category n, H_n, which is calculated as

$$H_n = \omega_n \alpha^{-1} P = \alpha^{-1} P_n \quad , \tag{3}$$

where ω_n is the ratio of employment in category n to total employment E, α is an activity rate defined as P/E and P_n is the actual household demand measured in terms of population. We then define the coefficients of dependence between the employment sectors as λ_{mn} and the input-output equation can then be written as

$$E_m = H_m + \sum_n \lambda_{mn} E_n \quad . \tag{4}$$

We can write the equivalent of the multiplier relationship in matrix form as $\mathbf{e} = \mathbf{h}(\mathbf{I} - \mathbf{\Lambda})^{-1}$ where $\mathbf{e}$ and $\mathbf{h}$ are appropriate row vectors for E_m and H_m and $(\mathbf{I} - \mathbf{\Lambda})^{-1}$ is a matrix inverse where $\mathbf{I}$ is the identity matrix and $\mathbf{\Lambda}$ the matrix of the technical coefficients λ_{mn}. Equation (4) predicts the employment by category/type which is then input to the employment location model that scales these quantities to employment in zonal locations E_{im} which we will sketch in the next section.

The Employment Location Model

Before we introduce the model, it is worth noting that there are several developments of LUTI models that extend the spatial interaction structure of these models to embrace spatial interactions between different employment sectors; in short, disaggregating the classic input-output model in equation (4) into zonal as well as sectoral categories. Following earlier work by Macgill (1977), one of the authors (Batty, 1986) has developed a generic framework for this but the one that is being currently used in operational practice is the extended MEPLAN model which is built around a spatially disaggregate input-output structure (Echenique, 2004). In fact, Echenique's model is rather more general than any of those that we have detailed here in that he defines two types of sector – production and consumption, or supply and demand, not specifying formally whether these sectors are measured in terms of population or employment. The model as it has been applied, defines production as employment and consumption as household demand, measured in units of employment and thus this structure is similar to the input-output model used here. As the model is iterative in structure, what begins the process can be either inputs or outputs and there are many possible variants. However, the key difference is that the model structure here separates aggregate predictions for employment from their spatial allocation which in turn are separated from the residential and service locations and interactions. In the Echenique and related models, all these stages are fused.

For the rapid assessment of scenarios, we need to develop a model whose independent variables are relatively easy to forecast in their own right. To this end, we first predict floorspace as a linear function of various independent variables that are easy to specify for future scenarios and then use these predictions of floorspace to produce employment by category and location using a second linear prediction model. Immediately an issue arises as to the transmission of errors in this two stage process but although error is clearly passed on from the first to the second models, both models are heavily constrained to lie within certain limits. It is not possible in these kinds of model to work with negative quantities and moreover, we need to make sure that the total employment which is predicted for each category sums to the known totals generated from the input output model. These complicate the structure and in general, compromise the strict statistical interpretations of the model fits.

To provide some sense of these two models, the first activity that is predicted – floorspace y_i' say – is estimated as a function of various independent variables x_{ig} where the coefficients $a_o ... a_g$ are estimated using least squares regression. These variables are then input to a second regression which predicts an employment category z_i' as a function of another set of independent variables v_{ig} with coefficients $b_o ... b_g$ and the floorspace variable from the first stage $\vartheta y_i'$ where ϑ is the appropriate weight from the floorspace model. These two models are written as

$$y_i' = a_0 + \sum_g a_g x_{ig} \quad , \tag{5}$$

$$z_i' = b_0 + \sum_g b_g v_{ig} + \vartheta y_i' \quad . \tag{6}$$

It might be remarked that these equations could be collapsed into one but at each stage we constrain the estimates to meet minimum values that are greater than or less than zero. There are two ways of doing this. First if the estimates that are below zero are small in number and value, we simply exclude the variable from the model and replace the value of the predicted floorspace or employment by its observed equivalent. That is, for equations (5) and (6) respectively

$$\text{If} \quad y_i' \le 0 \quad \text{then} \quad y_i' = y_i \tag{7}$$

$$\text{If} \quad z_i' \le 0 \quad \text{then} \quad z_i' = z_i \quad , \tag{8}$$

where the primed variable denotes the predicted value and the non-primed variable the observed value. The second method is to simply scale the value by adding the minimum value to each prediction; that is

$$\text{If any} \quad y_i' \le 0 \quad \text{then} \quad y_i' = \min_{\ell} y_\ell' + y_i' \tag{9}$$

$$\text{If any} \quad z_i' \le 0 \quad \text{then} \quad z_i = \min_{\ell} z_\ell' + z_i' \quad . \tag{10}$$

In fact we have used the first method for the first model whereas for the second model where we predict employment by category, we scale the estimates so that the sum of these employments meet the predetermined input-output totals.

We can now state the model as follows. First we predict the floorspace F_{im}' associated with employment category m as

$$\left.\begin{array}{ll} F_{im}' = a_0 + \sum_g a_g x_{ig}, & if \quad F_{im}' > 0 \\ F_{im}' = F_{im}, & if \quad F_{im} \le 0 \end{array}\right\} \quad , \tag{11}$$

and then we use this floorspace to predict the employment in category m as E'_{im} from

$$\left.\begin{array}{ll} \hat{E}_{im} = b_0 + \sum_g b_g v_{ig} + \vartheta F'_{im}, & if \quad \hat{E}_{im} > 0 \\ \hat{E}_{im} = E_{im}, & if \quad \hat{E}_{im} \le 0 \end{array}\right\} \quad . \qquad (12)$$

The final employment location is scaled to ensure that

$$E'_{im} = E_m \frac{\hat{E}_{im}}{\sum_m \hat{E}_{im}} \quad . \qquad (13)$$

Equations (12) and (13) that define the model are applicable to each employment category associated with the input-output model. When we detail the calibration below, we will present the actual independent variables and categories used.

The Land Use Transportation Interaction (LUTI) Model

We can state the land use transportation models quite succinctly. We allocate employment E_{im} to residential areas using a spatial interaction model that computes the work trips T_{ij}^q by travel mode q from workplace zone i to residential zone j as a function of the residential floorspace R_j, the modal travel cost c_{ij}^q from i to j, and the friction parameter β^q. We state this model as

$$T_{ij}^q = \left(\sum_m E_{im} \right) \frac{R_j \exp(-\beta^q c_{ij}^q)}{\sum_\ell R_\ell \sum_q \exp(-\beta^q c_{i\ell}^q)} \quad . \qquad (14)$$

Note that the script ℓ is used throughout the text as a floating index pertaining summation over zones, employment sectors or floorspace types to make a distinction from the actual flows or volumes. Population in zone j is computed by summing trips over modes q and employment zones i and scaling the result by the activity rate α as

$$P_j = \alpha \sum_i \sum_q T_{ij}^q \qquad . \tag{15}$$

We now compute the demand for employment in category n at zone c as a function of the household demand in j, thus repeating in some way the sort of structure that is represented at both the input-output and employment location models stages. Then

$$S_{jcn}^q = \xi_n P_j \frac{F'_{cn} \exp(-\phi^q c_{jc}^q)}{\sum_\ell F'_{\ell n} \sum_q \exp(-\phi^q c_{j\ell}^q)} \qquad . \tag{16}$$

where the service centre zone is notated as c, ξ_n is the employment demand coefficient for category n and ϕ^q the modal friction parameter. We are then able to predict the employment of each category in zone c by summing equation (16) over j and q as

$$E'_{cn} = \sum_j \sum_q S_{jcn}^q \qquad . \tag{17}$$

We see an immediate simultaneity in equations (14) to (17) where an element of the employment input which we take from the prior employment allocation models is also predicted by the LUTI model. Strictly this leads to iteration over equations (14) to (17) until balance occurs (Wilson, 1970). However what we effectively do is divide the categories of employment into two sets: those that we consider cannot be predicted as a function of population which is akin to true exogenous employment in the traditional basic sense, and employment in that category that is clearly a function of population – in short, service or non-basic employment.

Essentially we apply equations (16) and (17) to the retail category of service employment, that is $n = retail\ employment$ and we aggregate this to one class. Employment which is not retail employment is predicted using the employment location model in the previous section and thus the two employment models are quite separate. However within the overall structure, there exists the potential for iterating between the

LUTI and employment location models if this appears appropriate. Other loops involving these various models relate to a) capacity constraints and b) locational attractors. First there are capacity constraints on the floorspace and land areas consumed by population and employment. We convert population and employment into residential and commercial floorspace as $R'_j = \eta P_j$ and $F'_{cn} = \gamma E'_{cn}$ where η and γ are appropriate conversion coefficients. Then if $R'_j > R_j^{\max}$ and $F'_{cn} > F_{cn}^{\max}$, we scale these attractors to produce less activity in the next iteration of equations (14) to (17). This reduces the attraction until the constraints are met. We should note that these floorspace variables are also used in the employment location model and thus another iteration can be set up which involves both the LUTI and its precursor, the employment model. Finally if the trip distributions T_{ij}^q and S_{jcn}^q exceed their link capacities, this sets up yet a further iteration which involves ensuring that these link capacities are met and this involves altering the travel cost matrices accordingly. There are many possible loops of this kind and we are exploring all of these within the wider project (Batty et al., 2011).

Applications to London and South East England

Defining the Regional Model and Classifying Employment

Our suite of integrated models is being developed for the urban region comprising Greater London and the Outer Metropolitan Area which is a region containing 6.82 million (m) jobs and 13.42m population. It is broadly circular whose radius varies between 40 and 50 miles from central London. It includes all the major London airports and green belts but does not extend to include the university towns of Oxford and Cambridge, the new town of Milton Keynes, or the south coast ports and resort towns. This region however cuts across three official English regions – London, the East, and the South East – and this complicates the employment modelling for the input-output model is built at this regional level. The employment model which disaggregates these regional totals for the finer scale population zones is thus built for these three regions which we refer to as the Greater South East Region. This larger region is about twice the size of the actual LUTI model region with a population of 20.56m and employment of some 10.09m (which is composed of 3.84m in London, 2.82m in the East and 4.15m in the South East). The LUTI model region is composed of some 1767 zones based on wards and this is nested within the larger employment model region which contains some 3202 zones. The zoning systems for the three models are illustrated in Figure 2. In fact it is worth noting that the ratio of population to employment – the activity rate α – in the larger region to the smaller is quite close, 1.967 compared to 2.038 which indicates that the London region is only slightly more active than its wider periphery.

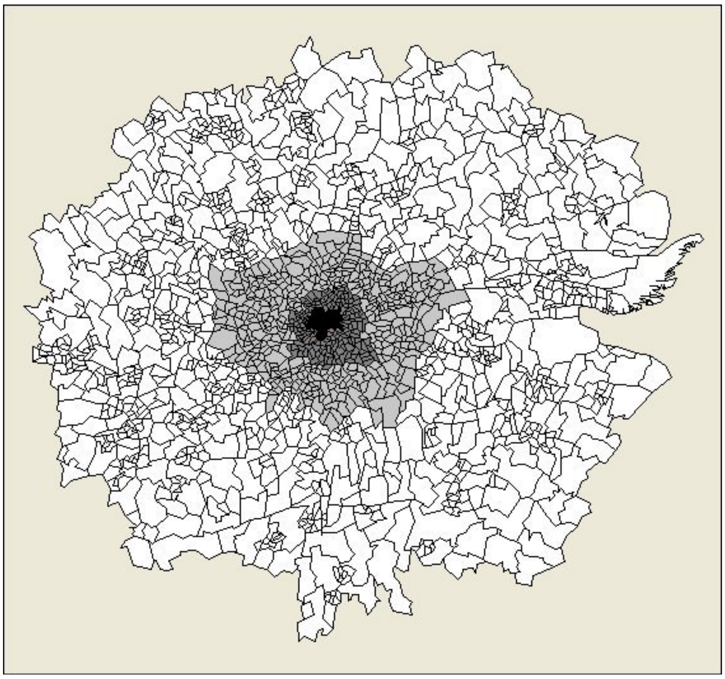


Figure 2: The Input-Output, Employment and LUTI Model Regions

Left: the three regions used in the I-O and Employment models, with the inset → Right: the London and Outer Metropolitan Area, with the Greater London Authority area (mid grey), Inner London (dark grey) and the Central Boroughs (black) nested within one another.

To achieve consistency between the input-output and the LUTI models, the employment model will be developed for the larger region but only applied to the smaller region which is a proper subset of the larger without any iterative couplings between the sectors. In short this is only possible because the employment model is composed of separable and distinct sectors. We now need to define these sectors. There are 42 sectors in the input-output model including miscellaneous but to give an idea of what the region would look like if employment were distributed uniformly between these sectors and over all 3202 zones, we would only have 75 employees in each sector in each zone. In short, many of these sectors are far too small to be capable of being simulated in locational terms. 24 of these sectors or about 60% account for only a tiny fraction, some

5% of the employment, and we thus need to collapse these according to some obvious logic that takes account of their magnitude and spatial distribution. The services sector in this region is enormous and constitutes over 80% of all employment. Many of these services are highly clustered. 2% of all employment, each in different types, is in just 4 categories x zones which is 0.001% of the number of categories x zones while 10% of all employment is located in just 0.01% of the categories x zones. Indeed if we plot the spatial distribution of employment across the larger region, it is very clear that this is highly polarised with some 10%, 20%, 30% 40% and 50% of all employment in 0.4%, 1.5%, 3.6%, 6.9% and 11.6% of all zones. This is far more extreme than the oft-quoted skew distribution of income which suggests that 80% of all wealth is in the hands of 20% of the population (the so-called 80-20 rule which describes the long tail)..

A good measure of inequality of the distribution of employment is given by the entropy function defined as

$$H = -\sum_i \frac{E_i}{E} \log \frac{E_i}{E} \tag{18}$$

which varies from 0 to a maximum value of $\log(n) = \log(3202) = 3.505$, where 0 denotes a situation where all employment is in one zone and $\log(n)$ a situation where all the employment is evenly spread. A measure of inequality is given by the redundancy which is defined as

$$R = \frac{\log(n) - H}{\log(n)} = 1 - \frac{H}{\log(n)} \tag{19}$$

which varies from 1, complete inequality to 0, complete uniformity or equality. The entropy H in the employment distribution is 3.179 and thus the redundancy R is 0.093, apparently much closer to uniformity than extreme inequality, notwithstanding the apparently highly skewed nature of the distribution. In Figure 3, we show the employment distribution in terms of its rank size in both its raw and its log transformed plots and it is clear that there is substantial polarisation. But given the number of zones, this distribution is moderated by the fact that there is employment in all zones, hence the value of the redundancy statistic. Were we to aggregate the zones, the distribution would become more extreme.

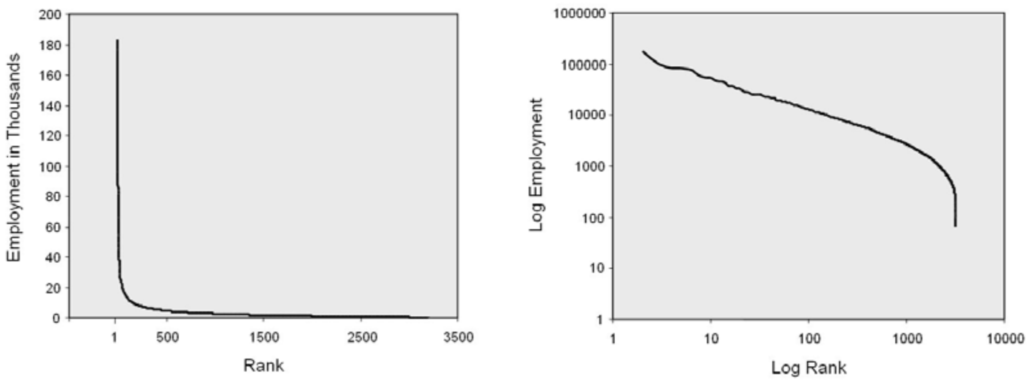


Figure 3: Inequalities in the Employment Size Distribution by Ranked Locations

Another perspective on the structure of employment in the region and one that is required in grouping the 42 sectors from the input-output model into a more manageable and representational set of classes involve examining the clustering of employment types. The services sector is enormous constituting over 80 percent – manufacturing and transport only 10 percent, and thus we need to disaggregate different kinds of services – financial and business and probably IT from retail and from public services. It is clear that the locational demands of these various services are quite different as our analysis of clustering reveals. A good measure of spatial clustering is the Getis-Ord G statistic (Getis and Ord, 1992) which essentially determines how close a particular activity in a particular zone is to all other activities of that type in all other zones. This statistic has been computed for all the zones in the system with respect to each of the 42 categories of employment that are defined in the input-output model. We use an inverse distance weighting with a 7 mile radius cut-off to encapsulate relevant zones near to those which we consider clusters might appear in. The observed employment at zonal (ward) level is taken from the Annual Business Index from the Office of National Statistics which is a 10% yearly sample, and we combine multiple years (2005-2007) to produce reliable estimates of the actual distribution.

The distribution of these cluster statistics for the 42 sectors is highly polarised and we use these to provide indicative measures of the extent to which sectors need to be specific. The largest cluster values which are more than twice the average are of two kinds: banking and finance, large in employment, and specific one-off activity locations such as oil, gas and water utilities which tend to be rather small in employment. Banking and finance are heavily concentrated in the core of the region. The second less intense set of clusters involve professional, miscellaneous, public, and hotel-restaurant services which form more diffuse clusters. These are still concentrated in Greater London but are much larger in scale constituting some 1.5m jobs in the region. In short, we see two kinds of services – those which are very niche and concentrated in the core and those which are concentrated in urban areas which define the polycentric structure of towns that form the wider region but are spread more diffusely as the population is spread throughout the region. One-off locations do also lead to clusters but it is very clear that many of the clusters are heavily orientated to the size and scale of the agglomerations of population in the region. In terms of other significant but smaller groups, only about 5% of employment is in manufacturing which is about the same as in transport, with the transport activity concentrated around airports and central city hubs. Utilities, primary industry and construction are much more idiosyncratic in their locations, either spread fairly evenly in association with population or in very small one-off concentrations. In fact the scale of this activity is so small and the locational requirements so diffuse or special that we will treat these sectors rather differently as we illustrate below when we design the employment models.

In summary then, clustering is either caused by the agglomeration of activities in attractive business areas (overwhelmingly in central London), or the clustering of activities within and around large facilities such as airports and industrial parks. As we have noted, the most highly clustered activities are finance, with insurance closely behind, and these activities are dominant within Central London and thus good contenders to be aggregated together. Business services are significantly less spatially clustered and, while still centralised, are not concentrated to the same degree in central London, and thus should remain separate. In terms of clustering at large facilities, this can be seen in the air transport, which is highly clustered, overwhelmingly around Heathrow in outer London. Motor vehicle, electronic and textile manufacturing are moderately clustered, and are located in outer London and the wider region. The majority of manufacturing activities however do not show high degrees of clustering. The sub-regional geography also indicates some misclassifications in the proposed groups. Oil and gas is concentrated in central and inner London, and is surely service jobs rather than primary sector activities. Similarly the high degree of printing and publishing jobs concentrated in inner London shows that these are service jobs, rather than manufacturing. These groups need to be split into their component parts or simply classed in a service group, as the service proportion is likely to greatly exceed the manufacturing proportion. The utilities groups are highly clustered, and are located mainly outside Greater London. Construction jobs are not clustered, and other than air and water transport, other transport jobs are also widely spread.

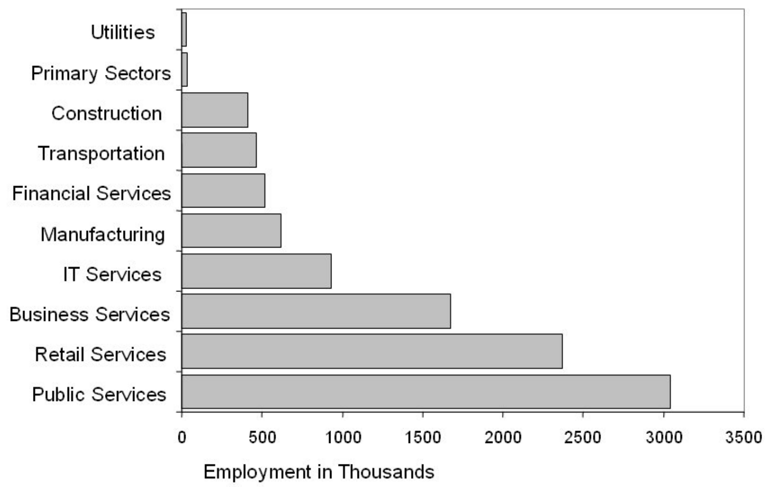


Figure 4: The Distribution of Employment Groups for the Location Model

A process of to-ing and fro-ing in examining these patterns led us to aggregate the 42 sectors from the input-output model to 10 distinct sectors which we define in terms of their size as Public Services (30%), Retail Services (24%), Business Services (17%), IT Professional Services (9%), Manufacturing (6%), Financial Services (5%), Transportation (4%), Construction (4%), Primary Sectors (<0.5%), and Utilities (<0.5%). We show these in terms of employment totals in the bar graph in Figure 4, from which it is quite clear that the 'services' categories completely dominate the region with 84% of the activity in these groups. Although we do not show this here for the region, this mirrors the polarisation of employment that we commented on above as implied by Figure 3. As a structure for the ten sectors, it represents as good an aggregation reflecting all these diverse factors as we are likely to get. It is on this basis that we have developed the employment location model.

Specifying and Fitting the Floorspace Location Models

Of the ten sectors, four have very little spatial similarity to the other six which are the core sectors. In fact for each one of these four, they do not have any cross-correlations with any of the other nine sectors with values greater than 0.4 and most are below 0.1. Of the remaining six sectors, all the service sectors are highly correlated with one another, in contrast to the manufacturing sector which is spatially quite distinct. We now define these sectors in terms of employment E_k where we use the index k to distinguish these new sectors from the original 42 input-output sectors m, n. In terms of possible explanatory variables, in particular floorspace, these six sectors have quite strong relationships with the three kinds of floorspace, each having a correlation of at least 0.66 with total, industrial, retail and office floorspace. Another interesting feature of this data, is that of the two accessibility variables measured using population potential from Hansen's (1959) measure for public and private transport networks, the correlations with these variables are all less than 0.2. This prompts us to compare these distributions with population itself and these correlations are also rather low, thus implying that the employment base is very different spatially from the distribution of population. The relevant comparisons between employment and floorspace categories shown as correlations are given in Table 1 where the key correlations are in bold italics (between the six sectors) and underscored for the floorspace comparisons.

Table 1: Correlations (x100) between Six Key Employment Sectors and Floorspace

	Manu-facturing	**IT Services**	**Financial Services**	**Business Services**	**Public Services**	**Retail Services**
Manufacturing	100					
IT Services	24	100				
Financial Services	5	42	100			
Business Services	18	***81***	***46***	100		
Public Services	16	65	25	71	100	
Retail Services	***33***	75	33	***77***	***67***	100
All-Floorspace	47	75	49	78	65	81
Retail-Floorspace	23	59	26	63	59	<u>85</u>
Office-Floorspace	14	<u>79</u>	<u>65</u>	<u>89</u>	<u>71</u>	72
Industrial-Floorspace	<u>66</u>	18	2	11	10	29

As we noted in the previous sections, the employment model which is based on predicting the employment in each of these six sectors is divided into two stages. Clearly from Table 1, correlations between employment and floorspace are quite high and as floorspace is a critical supply side variable that can be manipulated to determine future scenarios that we might test, we decided to first build a predictive model of the three floorspace categories in Table 1, without the global category All-Floorspace which is in fact the sum of Retail-Floorspace, Office-Floorspace, and Industrial-Floorspace. This first model uses these various measures of floorspace as independent variables together with a series of other variables that reflect other key factors that appear important. In fact we used a stepwise procedure to generate the best set of variables for the three categories. The other variables are of two kinds; accessibility to existing office, retail and industrial floorspace (to generate clustering and agglomeration effects) where the attractor is the relevant floorspace variable, general accessibility to population, accessibility to particular infrastructure facilities such large airports based on network travel time to the facility; and then potential space for urban development, such as existing land for commercial uses which is linked to agglomeration and brownfields redevelopment potential, and open space with potential for development. This was calculated as total greenspace less protected greenspace crucial to existing and future planning policies.

Our initial models regressed the three categories of floorspace (which we now define as F_j^z, $z = 1$: retail floorspace, $z = 2$: office, $z = 3$ industrial) in different combinations of these variables and we immediately found that there were strongly contrasting patterns for urban and rural areas. These differences were difficult to capture in a single regression model with low values of the variance explained. This problem was tackled by introducing two sets of dummy variables in of each floorspace models. The objective of the first set is to differentiate wards with no commercial floorspace (isolated rural wards) from those with commercial floorspace (urban wards) and those with high commercial floorspace (town centre wards). The objective of the second set is to include the distinct conditions in Inner London and this is achieved by introducing three different subregional effects that partition the region into three distinct sets of areas based on nearness to the central core. The models which were ultimately selected incorporate these dummies and this does weaken the model's ability to generate completely new urban developments in rural areas. Arguably the majority of such development takes place at 'seed points' in small settlements that become towns, and on brownfield sites that are redeveloped and thus this should not be a major shortcoming. The three models that have been fitted are variants of the generic structure that follows:

$$
\begin{aligned}
F_j^z = a_0 + \sum_{\substack{\ell=1,2,3 \\ \ell \neq z}} a_k F_j^\ell + a_4 B_j + a_5 D_j + a_6 A_j^z + a_7 T_j + a_8 G_j + \\
+ a_9 H_j + a_{10} Q_j + a_{11} M_j + \sum_{\ell=1,2,3} s_\ell \delta_j^\ell + \sum_{\ell=1,2} f_\ell \delta_j^\ell
\end{aligned}
\tag{20}
$$

where the coefficients $a_o ... a_{11}$ are the usual regression weights that imply a degree of significance for the variables to which they are ascribed, s_ℓ and f_ℓ are the coefficients of the dummy variables δ_j^ℓ which take on values of 0 or 1 for specific zones. The three dummies associated with s_ℓ, $\ell = 1,...,3$ are those for the three subregional effects noted above while the two dummies associated with f_ℓ, $\ell = 1, 2$ are those which determine thresholds on the relevant floorspace category. $B_j, D_j, A_j^z, T_j, G_j, H_j, Q_j, M_j$ are respectively non-domestic buildings, domestic buildings, accessibility to floorspace of type z, accessibility by public transport to population, both based on Hansen's (1959) potential measure, greenspace, distance to Heathrow, distance to other airports, and distance to major motorway junctions. This model has been fitted to the three floorspace distributions and the results are presented in Table 2.

The performance of these models is quite good. Some 70 percent of the spatial variation over a very large region with many small zones is good enough for forecasting purposes where we are very largely concerned with specifying future changes in floor space in 'what if' contexts. Moreover we will use these models as inputs to the employment model where we will bring other variables to bear as well as these various constraints on ranges of values that we indicated in an earlier section. It is to these models that we now turn. Concern about error propagation is worth noting but the rather inelegant fitting and manipulation of these models is achieved in the context of a deeper learning about the region.

Table 2: Estimates (and Standard Errors) for the Three Floorspace Models

Independent Variable	Retail Floorspace $\{F_j^1\}$	Office Floorspace $\{F_j^2\}$	Industrial Floorspace $\{F_j^3\}$
Constant a_0	-3992.257 (908.641)	-13763.129 (2517.158)	-21825.274 (2465.222)
Retail Floorspace F_j^1	*na*	0.771 (0.034)	-0.138 (0.026)
Office Floorspace F_j^2	0.125 (0.005)	*na*	-0.162 (0.012)
Industrial Floorspace F_j^3	-0.023 (0.008)	-0.278 (0.021)	*na*
Nondomestic Building B_j	56.984 (5.640)	226.331 (15.232)	450.912 (8.434)
Domestic Building D_j	25.062 (4.198)	*na*	-23.834 (8.604)
Access to Retail Floorspace A_j^1	0.070 (0.004)	*na*	*na*
Access to Office Floorspace A_j^2	*na*	0.045 (0.003)	*na*
Access to Industrial Floorspace A_j^3	*na*	*na*	0.125 (0.005)
Public Transport Accessibility T_j	*na*	128.338 (33.297)	47.792 (14.114)
Greenspace G_j	-0.149 (0.018)	-0.173 (0.050)	-0.364 (0.037)
Time to Heathrow H_j	*na*	-121.947 (38.358)	*na*
Time to All Airports Q_j	27.001 (7.939)	-178.775 (58.570)	*na*
Time to Motorways M_j	*na*	*na*	-74.595 (32.856)
Subregional Effect Core δ_j^1	-89153.674 (5057.273)	94686.907 (15593.312)	-62219.581 (8014.538)
Subregional Effect Inner Areas δ_j^2	-29533.271 (2213.944)	-56162.792 (5268.918)	-60027.959 (3446.372)
Subregional Effect Outer Areas δ_j^3	-6657.748 (1065.701)	-12246.272 (2551.055)	-27674.705 (2273.093)
High Impact Floorspace for z	111907.227 (2580.984)	69355.676 (6664.709)	*na*
No Impact Floorspace for z	-6583.713 (958.500)	*na*	*na*
Correlation Squared	0.719 (14712.458)	0.677 (41189.450)	0.679 (29902.481)

The Employment Location Models

As we argued above, we do not intend to develop separate models for all ten sectors of employment. Four of these sectors – the primary, utilities, construction and transport sectors are too specific in their locational requirements, too diffuse across the region, and/or simply too small to be significant for such forecasting. When we forecast the future locations of these sectors, we will simply scale them to their baseline values at 2005, the year at which the employment models have been calibrated and for which data exists, or we will pre-specify their values in terms of the scenarios we develop; or more likely a combination of these two strategies will be used.

The six sectors k that we formally model are all functions of their relevant floorspace which has been modelled in the previous section. We also add a number of key variables based on accessibility and travel time/distances to these models, developing a generic form that is applied differentially to each sector dependent on what we consider to be significant causal drivers of their respective locational patterns. The model is similar to that for floorspace and we state it as

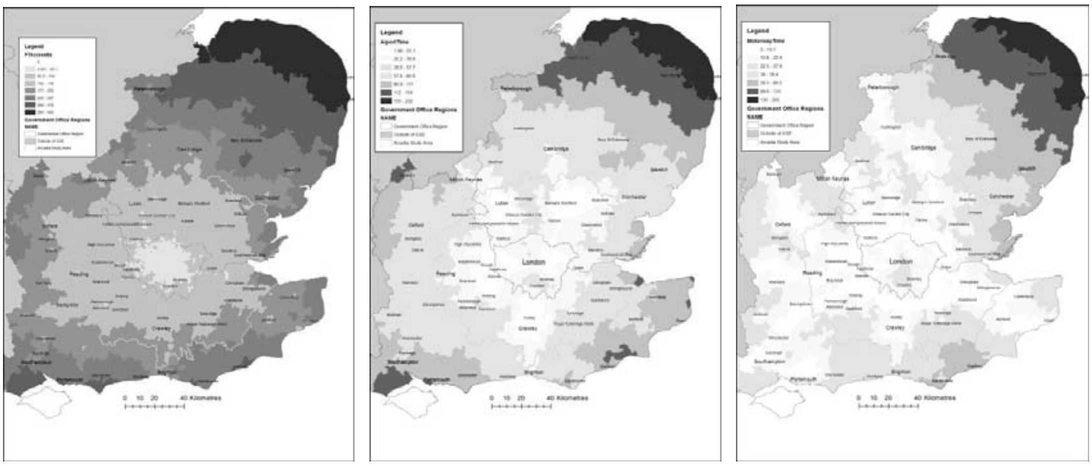
$$E_{jk} = b_0 + \sum_{z=1,2,3} b_k F_j^z + b_4 A_j^{PT} + b_5 A_j^R + b_6 M_j + b_7 J_j + b_8 H_j + b_9 h_j + b_{10} Q_j \tag{21}$$

where $E_{jk}, k = 1,...,6$ or k={financial services, manufacturing, IT services, business services, retail services, and public services}, $b_0 ... b_{10}$ are the respective regression coefficients, A_j^{PT} and A_j^R are the public transport and road accessibility potentials for the respective employment variables, J_j is distance/travel time to rail hubs, and h_j is the distance/travel time to A-Class roads. All other variables are as defined previously.

Even though the employment models are largely driven by floorspace from the previous model, and observed floorspace is used to fit the models in equation (21), these are also functions of the various accessibility potentials and direct distance measures to facilities. The generic Hansen (1959) measure A_j^k is defined for an activity k in zone j as

$$A_j^k = K \sum_i E_{ik} d_{ij}^{-2} \qquad (22)$$

where we show these as accessibilities to employment type E_{ik} and where the distance/travel time d_{ij} is that associated with the particular mode of travel, that is public transport PT or car/road transport R. To provide some sense of the distribution of these variables, we show three accessibility patterns for the Greater South East Region in Figure 5. In Figure 6, we also show the distribution of employment for the ten categories, six of which we are modelling using equation (21). It is clear that for the distributions we are not intending to model, the patterns are much more diffuse than those for the services and manufacturing that we are modelling. These latter distributions are very highly concentrated and the affect of the accessibility variables in equation (21) is to diffuse these concentrations a little when these variables are embedded into the employment model.



Access by Public Transport   Accessibility to Airports   Accessibility to Motorways

Figure 5: Sample Accessibility Surfaces Computed Using Various Weighted Indices

In Table 3, we present the results of fitting the six models. The performance of these models is good and we have included only those variables which are significantly different from their exclusion from the models at the 5% level. There has been considerable to-ing and fro-ing in developing these models, examining regional distributions such as those in Figures 5 and 6, and then examining significance and the causal logic of activities and indices that might be related to one another. In terms of Table 3, each of the categories is associated with its most immediate floorspace – that

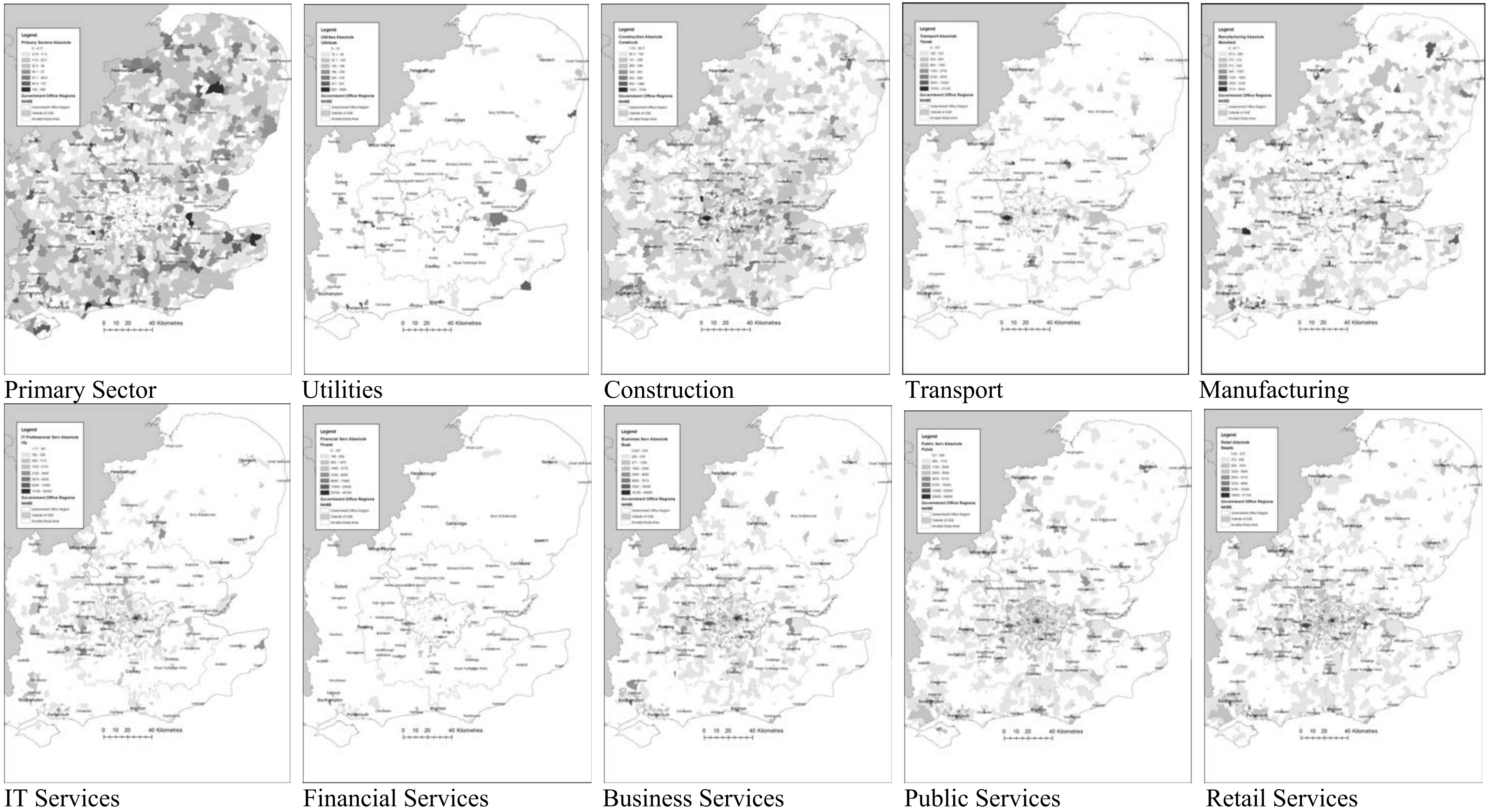


Figure 6: The Distribution of the Ten Employment Types Across the Greater South East Region
(Readers should zoom in on these figures to extract the detail: the figures are located at http://www.complexcity.info/the-scale-project/)

Table 3: Estimates of the Coefficients and Performance of the Employment Location Models (Standard Errors in (•))

	Financial Services $\{E_{j1}\}$	**Manufacturing** $\{E_{j2}\}$	**IT Services** $\{E_{j3}\}$	**Business Services** $\{E_{j4}\}$	**Retail Services** $\{E_{j5}\}$	**Public Services** $\{E_{j6}\}$
Constant b_0	-405.715 (61.877)	44.992 (14.903)	333.043 (40.553	268.647 (40.092)	113.798 (60.795)	1145.055 (70.777)
Retail Floorspace F_j^1	*na*	*na*	*na*	*na*	0.054 (0.001)	*na*
Office Floorspace F_j^2	0.011 (0.000)		0.010 (0.000)	0.023 (0.000)	*na*	0.019 (0.000)
Industrial Floorspace F_j^3	*na*	0.005 (0.000)	*na*	*na*	*na*	*na*
Employment Accessibility (within 25 minutes by public transport)	-4.408 (0.222)	*na*	1.069 (0.117)	2.721 (0.185)	0.791 (0.095)	2.843 (0.251)
Employment Accessibility (within 25 minutes by road)	0.461 (0.015)	*na*	-0.086 (0.008)	-0.188 (0.013)	*na*	-0.287 (0.017)
Time to Motorway Junctions M_j	23.519 (3.491)	*na*	-4.617 (1.354)	-7.140 (1.990)	-13.383 (2.805)	-7.971 (2.732)
Time to Rail Hubs J_j	*na*	*na*	*na*	*na*	29.918 (6.472)	-34.275 (9.126)
Time to Heathrow Airport H_j	*na*	*na*	-1.656 (0.693)	*na*	*na*	*na*
Time to Nearest A-Class Roads h_j	*na*	-4.709 (1.565)	*na*	*na*	*na*	*na*
Time to All Airports Q_j	-5.593 (2.109)	0.870 (0.364)	*na*	*na*	-3.846 (1.705)	*na*
Correlations Squared	***0.637***	***0.653***	***0.795***	***0.899***	***0.826***	***0.713***

is financial, business, IT and public services with office floorspace, retail with retail floorspace and manufacturing with industrial. Manufacturing is only associated with A-Class road and airport accessibility, while the other five services sectors are influenced by accessibility to motorways, and accessibility to public and car/road transport (with the exception of retail which is only public transport). Access to Heathrow airport appears to only influence IT services, while retail and public services are influenced by access to rail hubs. All this seems to make sense although there is substantial correlation between the various accessibility variables. In short, with the floorspace models, there is substantial implicit weighting through double counting. We make no apology for this, for we want to optimise the fit and as we will illustrate below, short of a perfect fit, we need to be very specific about how we use these models in prediction to reduce the error that is contained in the model estimates of the base year calibrations.

Using the Employment Models within the Integrated Framework

Error Propagation in Comparative Static Models

Our model framework is in the comparative static tradition. The input-output, employment, and LUTI models are all estimated to a cross section of urban and spatial structure at the base year, in this case 2005. The various data that we use have culled from various series and censuses from 2001 to 2007 but we have normalised these to ground all the data in the common year. There is very little discussion of how to use cross-sectional forecasting models in predictive contexts. The notion of comparative statics which is an old concept in economics, is assumed to be one where a model of a system produces a static equilibrium and in forecasting, a new static equilibrium is assumed to occur at some point in the future where the actual change is the difference between the two equilibria. This, of course, assumes that whatever changes that take place in the future work themselves out to the new equilibrium by the end of the forecasting period. In short, it is assumed that any motions that take place between the equilibria are not required to be known for the prediction to be useful.

Lowry (1965) in his seminal article on model design says of comparative statics: “The process by which the system moves from its initial to its terminal state is unspecified.

Alternatively, comparative statics may be used for impact analysis, where no target date is specified. Assuming only one or a few exogenous changes, the model is solved to indicate the characteristics of the equilibrium state toward which the system would tend in the absence of further exogenous impacts". In fact, Lowry (1964) in his **Model of Metropolis**, makes the further point that such comparative statics produces an instant metropolis, an emergent structure rather than the actual urban structure that we currently observe. It is on this premise that the baseline calibration of any comparative static model is 'what would happen' if the forces at work in the data describing the current urban structure were to work themselves out, rather than 'what has actually happened'. This has a profound impact on how such models should be used in prediction.

The problem with any model in terms of its calibration is that no calibration can give perfect estimates of the observed situation, whether it be a comparative static or a dynamic model. The problem is how the errors which are part of the calibration are to be treated in using the model in prediction. If one simply makes predictions with the same model that has been used in calibration, then these errors will be transmitted through to the future state. If the errors are greater than the differences that are associated with the future state, then what is error and what is new are completely confounded. Indeed, this is intrinsic to the entire prediction in that it will always be possible to find a forecasting period in which the overall change will be less than the overall error. It is somewhat remarkable that there has been so little discussion of this issue but in practice what is usually done is to generate differences rather than absolutes. In fact this is no easier with dynamic models that are fitted to differences than to static models that simulate the entire structure because both contain errors. It might even be argued that static models are preferable in that it is the difference between the two equilibria that is significant, and any prediction is then simply treated as a difference between the existing situation and the future state.

We can give this situation more clarity in terms of the input-output and employment models which are both comparative static. Adding time to the variables, we define the total population at the base year time t as $P(t)$ and total employment as $E(t)$. These quantities are external to the entire framework; in fact they can be predicted using aggregate demographic and employment models and one alternative we are considering is to predict population using the MoSeS model (Birkin, et al., 2009). However assuming we have predictions of these totals as $P(t+1)$ and $E(t+1)$ at time $t+1$, then we clearly know the increments

$$\left.\begin{aligned} \Delta P(t) = P(t+1) - P(t) \\ \Delta E(t) = E(t+1) - E(t) \end{aligned}\right\} \quad , \qquad (23)$$

which may be positive of negative. These are net values, for population and employment may grow and decline in different areas and this represents an intrinsic problem to most modelling frameworks that accept exogenous totals of this kind. Nevertheless, the input-output model takes total employment at times t and $t+1$ and generates totals $E_k(t)$ and $E_k(t+1)$, the differences $\Delta E_k(t) = E_k(t+1) - E_k(t)$ also being either negative or positive.

The employment model distributes these totals to the zones of the system, but before we broach the problem of dealing with net totals, let us consider the predictions of the estimated model at the calibration year $E'_{ik}(t)$ and predictions at the same year but with input variables for the future year defined as $E'_{ik}(t:t+1)$. Now the sum of these employments across sectors and zones equals observed total employment at time t

$$\sum_i E'_{ik}(t) = \sum_i E'_{ik}(t:t+1) = E_z(t) \quad , \qquad (24)$$

$$\sum_i \sum_k E'_{ik}(t) = \sum_i \sum_k E'_{ik}(t:t+1) = E(t) \quad , \qquad (25)$$

and to model where the increment of employment is located, we simply take the ratios $E'_{ik}(t:t+1)/E'_{ik}(t)$ and apply these to the increment $\Delta E_k(t)$. Then the employment model for each sector k is

$$\Delta E_{ik}(t) = \Delta E_k(t) \frac{E'_{ik}(t:t+1)}{E'_{ik}(t)} \quad , \qquad (26)$$

and the new employment at time $t+1$ is

$$E_{ik}(t+1) = E_{ik}(t) + \Delta E_{ik}(t) \quad . \qquad (27)$$

It is easy to show that these employments in equation (27) sum to the category totals produced by the input-output model. In short, the errors in the actual calibrated model are not transmitted to the new totals but only to the increment of employment due to the fact that we only allocated the increment using the model, not the original employment structure which is simply used as the baseline.

Problems with Predicting Total Aggregate Activity

The problems of specifying aggregate totals which begin the chain of prediction involve the fact that these totals reflect net rather than gross change. For example, a sector might be declining overall but growing in some places with decline elsewhere cancelling this out. This might be caused by simple differential growth *in situ* or in the case of population it can be a complex concatenation of changes in fertility, mortality and migration. However if we simply take the aggregate quantities and these are negative, then the largest negative values will be allocated to the areas which are most attractive to employment. In short, these kinds of models tend to work with positive quantities and when we feed them with negative, then the logic will simply be reversed. Strictly speaking, we need to separate out growth from decline – positive from negative. If we have some sense of this, then we probably need to input this distinction externally and for each quantity, $\Delta E_k(\tau)$, we need to break this into positive and negative components $\Delta E_k \tau) = \Delta E_k^+(\tau) + \Delta E_k^-(\tau)$ and allocate these quantities directly using the logic which is embodied in equations (26) and (27). In this case, we would then be somewhat more confident that the mechanisms of growth and decline are being dealt with appropriately in the model framework. But in the last analysis, the models need to be much more disaggregate spatial at all stages and the logic of the integrated input-output, employment and spatial interactions models developed by Echenique (2004) amongst others is unassailable. Models of the kind developed here tend to be based on the expediency of coupling models together which have been designed with different purposes in mind and in any further developments, the need to develop the integrated framework with strong coupling between the submodels would be critical.

Using the Model Framework for Scenario Generation

The integrated model is currently being used to explore and test the impact of three rather different scenarios for the future of the London region. We have a baseline scenario referred to as 'business as usual' which is essentially a trend projection based on current planning policies and infrastructure developments that are already being implemented. Current policies also imply a degree of compaction mirroring growth in the last two decades, and this is linked to the growth in services and the knowledge economy based in London and larger cities. The second scenario is one of 'decarbonisation' with low carbon cities being the dominant vision. This is implied in more concentrated, higher density forms and restricted car-based growth. In planning policy terms, it is an exaggerated version of 'business as usual, and it has implications for transport in terms of higher fuel prices/taxes promoting greener alternatives. The third scenario is one of 'deregulation' particularly with respect to compaction and the force of the greenbelt. It provides a direct contrast to the second.

All these scenarios imply different levels of employment. None of the ten categories of employment involve a decrease overall but this does not mean the problems with the net effects of change have been removed. In terms of the operation of the floorspace predictions, then starting with new estimates of floorspace which reflect these scenarios, we need to iterate equations (20) and also add new capacities, change accessibilities and add in anticipated changes in the building stock. In fact there are many locations that we need to specify in terms of these predictions and we will concentrate on making general changes across the board so that we might evaluate sufficiently different kinds of scenarios in terms of the anticipated physical changes. After we have made predictions of floorspace, we can make explicit changes in floorspace in the employment equations. In fact, we tend to reserve the employment model for handling one-off changes in physical and transport capacities while leaving the floorspace models to deal with more general changes which affect all zones within the wider region. In short, the floorspace and employment models have been designed so that we can intervene and formalise scenarios in this way through physical changes to the future region.

Conclusions and Next Steps

There are many problems in developing an integrated framework of models based on a loose-coupling of existing structures. In theory, an integrated strategy that weaves employment into population modelling with consistency at all scales is required but the problems of doing this at a disaggregate level and in a temporal, non-equilibrium context are formidable. The debate over cross-sectional or dynamic models or some fusion of both does not address the two issues that are relevant to using any model in prediction: the need to begin the process somewhere with assumptions about aggregate quantities and the need to deal with perpetuating errors in the calibrated outputs of models when they are used in prediction. Neither of these problems has been addressed much, somewhat remarkably given that there is nearly 50 years of sustained effort with these kinds of model. This must be due to the fact that the scale of these efforts has tended to end before their intensive use in prediction, often because of the sheer scale of the exercise in getting to the point where predictions are possible, and/or in terms of the organisations involved in building the model whose expertise and goals are often very different from those organisations involved in using it in prediction.

Computation has now reached the point where many of these problems can now be resolved. Models like this one can be quickly designed, changed, and implemented with stakeholders and other scientists in ways that enrich their use in prediction. We need a sustained attack on the problem of prediction in urban systems, given that during the time when these kinds of models have been developed, our view of the predictability of human systems and what predictive models actually mean has changed quite radically. The notion of providing predictions but for the shortest intervals of time, has changed quite dramatically. Models such as this and others in this book, now need to be used in conditional prediction of many kinds, to generate 'what if' scenarios and inform the debate about how these tools might be continually modified in response to what we learn about the problem and the future. In this context, there is a still a massive dilemma between ever more detailed models, and ever more scepticism about what we can predict and how we might use prediction in helping us inform these dialogues about the future. Integrated modelling is important in this but so also is the need to develop fast, visually accessible models that we can change quickly in response to our learning. This implies that the argument about simple, static, aggregate versus complicated, dynamic, disaggregate models structure is far from over: it is just beginning.

References

Batty, M. (1986) Technical Issues in Urban Model Development : A Review of Linear and Non-Linear Model Structures, in B. G. Hutchinson and M. Batty (Editors) **Advances in Urban Systems Modelling**, North Holland Publishing Company, Amsterdam, 133-162.

Batty, M., Vargas-Ruiz, C, Smith, D., Serras J., Reades, J., and Johansson, A. (2011) Visually-Intelligible Land Use Transportation Models for the Rapid Assessment of Urban Futures, **Working Paper 163**, Centre for Advanced Spatial Analysis, University College London, http://www.bartlett.ucl.ac.uk/casa/pdf/paper163.pdf

Birkin, M., Townend, P., Turner, A., Wu, B., and Xu, J. (2009) MoSeS: A Grid-Enabled Spatial Decision Support System, **Social Science Computing Review**, **27**, 4, 493-508.

Echenique, M. H. (2004) Econometric Models of Land Use and Transportation, in D. A. Hensher, K. J. Button, K. E. Haynes, and P. R. Stopher (Editors) **Handbook of Transport Geography and Spatial Systems**, Pergamon, Amsterdam, 185-202.

Getis, A. and Ord, J. K. (1992) The Analysis of Spatial Association by Distance Statistics, **Geographical Analysis**, **24**, 189-206.

Hansen, W. G. (1959) How Accessibility Shapes Land Use, **Journal of the American Institute of Planners**, **25**, 73–76.

Junankar, S., Lofsnaes, O., and Summerton, P. (2007) **MDM-E3: A Short Technical Description**, Working Paper, March 2007, Cambridge Econometrics, Cambridge, UK, available at http://www.camecon.com/

Lowry, I. S. (1964) **Model of Metropolis**, Memorandum RM-4035-RC, Rand Corporation, Santa Monica, CA.

Lowry, I. S. (1965) A Short Course in Model Design, **Journal of the American Institute of Planners**, **31**, 158-165.

Macgill, S. M. (1977) The Lowry Model as an Input-Output Model and Its Extension to Incorporate Full Intersectoral Relations, **Regional Studies**, **11**, 5, 337-354.

Maoh, H., and Kanaroglou, P. (2012) A Firm Failure Microsimulation Model, in F. Pagliara, M. de Bok, D. Simmonds and A. Wilson (Editors) **Employment Location in Cities and Regions: Models and Applications**, Springer, Berlin and New York, this volume

Moeckel, R. (2012) Firm Location Choice Versus Employment Location Choice in Microscopic Simulation Models, in F. Pagliara, M. de Bok, D. Simmonds and A. Wilson (Editors) **Employment Location in Cities and Regions: Models and Applications**, Springer, Berlin and New York, this volume

Putman, S. H. (1983) **Integrated Urban Models, Policy Analysis of Transportation and Land Use**, Pion Press, London.

Putman S. H. (1991) **Integrated Urban Models 2: New Research and Application of Optimization and Dynamics**, Pion Press, London.

Walsh, C. L., Dawson, R. J., Hall, J. W., Barr, S. L., Batty, M., Bristow, A. L., Carney, S., Dagoumas, A. S., Ford, A. C., Harpham, C., Tight, M. R., Watters, H., and Zanni, A. M. (2011)Assessment of Climate Change Mitigation and Adaptation in Cities, **Urban Design and Planning** (Proceedings ICE), **164**, 2, 75–84.

Wilson, A. G. (1970) **Entropy in Urban and Regional Modelling**, Pion Press, London.