[[Draft. Incomplete]]

Spatial Cluster Empirics

by

Danny Quah and Helen Simpson*
LSE Economics Department
and
Institute for Fiscal Studies
June 2003

* We thank Gilles Duranton and Henry Overman for help with the data and the MacArthur Foundation and the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS for financial assistance. Comments from Steve Redding and ISOM Conference participants have been useful.

Spatial Cluster Empirics
by
Danny Quah and Helen Simpson
LSE Economics Department
and
Institute for Fiscal Studies
June 2003

ABSTRACT

A useful model of economic geography should determine not just how much economic activity occurs at a given location but also where that location is relative to others. In empirical studies, those latter features appear nowhere in measures of concentration or in cross-section and panel data regression. Indeed, they are absent in many analytical models of economic geography. We provide a new econometric method for studying clustering that takes explicitly into account such spatial relations, motivating the analysis with a theoretical model of dynamically evolving spatial distributions that generates a spatial law of motion. Applying our techniques to geographically disaggregated data on UK manufacturing, we find ...

Keywords: agglomeration, manufacturing, point process, spatial concentration, spatial distribution, spatial law of motion, spillover

JEL Classification: C21, D92, L60, O18, R12

Communications to: Danny Quah, Economics Department, LSE, Houghton Street, London WC2A 2AE

Tel: +44/0 20 7955.7535, Email: `dq@econ.lse.ac.uk`
(URL) `http://econ.lse.ac.uk/staff/dquah/`

Spatial Cluster Empirics

by
Danny Quah and Helen Simpson*
LSE Economics Department
and
Institute for Fiscal Studies
June 2003

1 Introduction

It has become a commonplace to assert that the outstanding feature of economic activity is clustering, both in time and in space. But while the former—business cycles—has empirical techniques for study that are legion and varied, correspondingly rigorous methods are sparser for the latter—economic geography.

This absence has implications that are not just methodological but substantive in at least three dimensions: First, characterizing spatial concentration can shed light on interesting economic hypotheses regarding the nature of increasing returns. Second, policies for economic growth and development often involve ideas relating geographical clustering and productivity. Whether those policies achieve what they intend depends on the empirical validity of the motivating ideas. Third, spatial clustering is just another way of saying regional inequality. Inequality across space matters for the same reasons as do inequalities across people and across nations.

---

* We thank Gilles Duranton and Henry Overman for help with the data and the MacArthur Foundation and the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS for financial assistance. Comments from Steve Redding and ISOM Conference participants have been useful.

This paper describes a class of spatial probability models useful for analyzing geographical clustering in economic activity. These structures are motivated by our hunch that useful models of economic geography should determine not only how much economic activity occurs at a given location but also where that location is relative to others. Our models take such spatial relations explicitly into account.

The goal in this paper is not to propose yet another index of concentration or agglomeration. Such calculations are informative, but we seek instead to provide empirical and theoretical descriptions of economic geography that can be matched explicitly in structural estimation. To do that, the analysis needs to give a law of motion in space, analogous to the laws of motion in time used to study business cycles.

Section 2 describes the connection between this paper and related literature. It also describes, relative to that literature, the basic issues with which the current paper will be concerned. Section 3 provides an explicit skeleton framework that highlights our distinguishing, on the one hand, models that describe how much economic activity happens at given locations and, on the other, models that determine simultaneously both the location and the intensity of economic activity.

Section 4 presents a class of statistical models suggested by the model of Section 3 and used for the empirical analysis that follows in Section 5. The paper concludes with Section 6, summarizing our principal findings and describing extensions and directions for future work. The Technical Appendix Section 7 provides additional technical discussion.

2 Related work and basic issues

With rigorous general equilibrium modelling of Marshall's insights on 19th-century industrial clustering in place (among others, Fujita, Krugman and Venables (1999) and Krugman (1991)), the ensuing empirical analysis has, in our interpretation, taken two principal forms.[1]

[1] Since Marshall also used vivid language in describing how industrial centers ferment ideas so that "mysteries of the trade become no

First are attempts to examine directly the underlying economic mechanism, using the spatial dimension primarily as a source of apposite data. Examples of these include Ciccone and Hall (1996), Jaffee, Trajtenberg and Henderson (1993), Rauch (1993), and Henderson (2003), among others. Panel data or cross-section regressions, using observable covariates related by space and geography, are useful and revealing here. Such regression analyses end up constructing a hypothetical representative unit—a firm, a location, a city—and trace through the impact of differing covariate values on the performance of that unit.

Second are attempts to characterize the entire spatial distribution of economic activity, relative to a set of hypotheses. An economic theory might, for instance, predict a power-law distribution of production across space, evolving in particular ways over time. Or, theory might suggest certain patterns of industrial concentration over an appropriate spatial measure, possibly different from geographical area or Lebesgue measure.[2] The empirical research then seeks to document such regularities. Examples of this second approach include Devereux, Griffith and Simpson (1999), Duranton and Overman (2002), Ellison and Glaeser (1997), and Ioannides and Overman (2000). Here,

mystery; but are as it were in the air", and how "if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas", it is natural to draw a connection from spatial clusters to economic performance where knowledge matters importantly—examples might include Internet industries and Silicon Valley (see, e.g., David, Foray and Dalle, 1998; Jaffee, Trajtenberg and Henderson, 1993; Kolko, 2002; Quah, 2000; Rauch, 1993). In such analyses, however, proximity can easily mean something other than physical distance in Euclidean space.

[2] Localization—concentration in specific industries that end up more skewed, say, than that overall across all industries—can be viewed as simply a property relative to a distortion of the underlying space or, alternatively, a spatial feature suggesting the use of some relative measure different from Lebesgue measure.

interest rests not in the characteristics of a representative unit, but instead in the joint behavior of all the different units distributed across space. The current paper belongs to this second form of analysis. For want of a better terminology, we find it useful to think of this as a *macro* empirical analysis.

However, while most others using this macro empirical analysis have sought out particular implications of that spatial distribution—an index of concentration or localization, say, which is just a functional of an underlying distribution that strictly collapses the information in that distribution—our goal in this paper is to estimate directly a complete structured probability model for that distribution. Our direct line of attack has three advantages. First, it removes the need for axiomatic description of what yet another proposed index of concentration should do: with an explicit model, one simply calculates any desired feature of interest. Second, it automatically imposes internal consistency in one's understanding of what the data say. Third, it is a first step for further analysis relating interpretable structural economic parameters to observable data. In contrast, looking at only concentration indexes or other functional characterizations does not extend naturally to structural estimation.

Before proceeding to a more elaborate discussion of our models, it will help to set out some basic issues. This will also permit more precise description of how our work departs from others.

Fig. 1 shows, pictorially, one of the datasets we will analyze below. Each dot on the map of the UK in the Fig. shows the location of a computer industry plant. These location data represent a merging of latitude-longitude information in Duranton and Overman (2002) and postcode data for the manufacturing sector from the UK's Office of National Statistics 1996 list of all companies. The data are accurate to within a distance of 100m.

Distinct clusters of dots appear around London in the southeast and, proceeding northwards, around Birmingham and Manchester in turn, before another visible concentration at the Edinburgh-Glasgow corridor.

For comparison Fig. 2 shows the location of cutlery establishments in the UK. The most obvious cluster now appears around Sheffield,

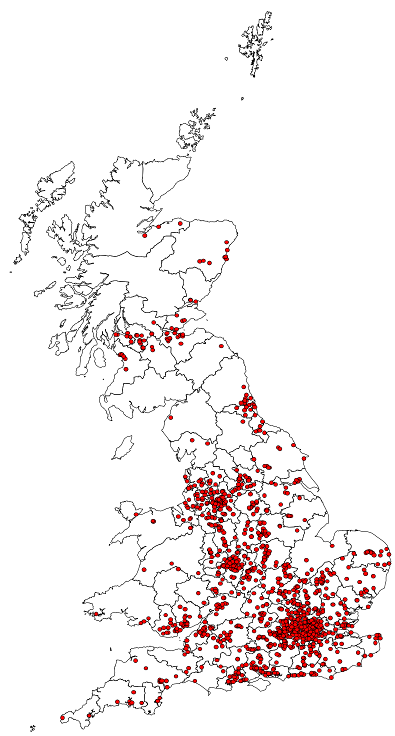

Figure 1: **Location of computer industry establishments in the UK**

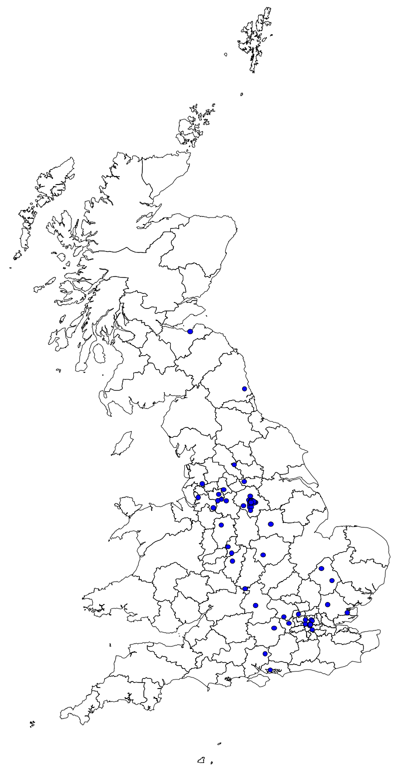

Figure 2: **Location of cutlery industry establishments in the UK**

although London again clearly remains a center of activity. Comparing Figs. 1 and 2, the cutlery industry has a point pattern in locations visibly different from that for the computer industry.

A spatial cluster might be taken to be just where a higher concentration of dots appears. On the map, there could be a single such concentration or there could be multiple concentrations over different locations—London might constitute a cluster, as might Silicon Glen in Scotland. (Section 5 will, when we turn to it later, give sharper characterization to these statements.) Not shown in Fig. 1 but associated with each dot might be potentially observable data on number of employees, value added, and so on, indicating the amount of economic activity occuring at a given establishment. Continuing the graphical encoding, we might think of such data represented by heights over each dot in Figs. 1–2.

To introduce the notation and framework used in the remainder of the paper, let $\mathcal{S}$ denote an underlying geographical space, fixed throughout the analysis. In Figs. 1 and 2, $\mathcal{S}$ is the collection of all points in the British Isles. More abstractly, $\mathcal{S}$ might be taken to be some subset of finite-dimensional Euclidean space.

A *spacetime stochastic process on* $\mathcal{S}$ is a collection of real-valued random variables

$$Y = \{ Y(z,t) : z \in \mathcal{S},\ t \in \mathbb{R}_+ \}, \tag{1}$$

where z indexes space and t indexes time. The variable $Y(z,t)$ might be observed employment in the computer industry [or value added or some other measure of a specific economic activity] in location z at time t.

For most points z in $\mathcal{S}$ quantity $Y(z,t)$ is zero. For instance, from extensive fieldwork we know that no computer industry establishments have located anywhere within 100m of the Starbucks where this paper's coauthors regularly drink coffee. Thus, at that Starbucks's z, there is coffee but no computer commercial activity. Define the subset of points

$$\mathcal{G}(t) = \{ z \in \mathcal{S} : Y(z,t) > 0 \} \tag{2}$$

where *some* economic activity of the kind we wish to study occurs at time t. It is $\mathcal{G}(t)$ at a given t that Fig. 1 depicts.

In our interpretation, a leading goal of macro empirical analysis in economic geography is to produce a probability law of motion for Y in equation (1). As time evolves, not only do the values of Y change, but so does set $\mathcal{G}$ in equaton (2). Theoretical modelling should generate laws of motion for Y and $\mathcal{G}$ that can be matched to data.

This analysis thus partly concerns how, given $\mathcal{G}(t)$ why $Y(z,t)$ for $z \in \mathcal{G}(t)$, takes a certain height over maps like Figs. 1 and 2. This would be research that takes locations as conditionally exogenously given and asks how much activity locates at a given site instead of at another given location. Most analytical research, in our view, has taken this approach, asking whether of two possible sites, say, as much activity locates at the first site as at the second.

But that is only part of the modelling problem. Theoretical modelling should also give insight on the determination of the set of points $\mathcal{G}(t)$. Since, formally, this just asks where $Y(z,\cdot)$ equals zero, it might seem to be only a special case of the previous analysis. However, posing this question relative to observed data such as Figs. 1–2 clarifies that the substantive issues are two-fold: First, does theoretical modelling consider a set of possible locations sufficiently rich so its predictions can be meaningfully compared to realizations such as we see in the Figures? Second, what kind of empirical analysis will such theoretical modelling motivate?

Figs. 3–4 suggest some stylized answers. The left side of Fig. 3 shows, in the top and bottom panels, two hypothetical configurations of economic activity in space, each a snapshot at a fixed point in time. Both panels graph, against space $\mathcal{S}$ on the horizontal axis, levels of economic activity Y on the vertical. In the top panel, high and low levels of activity alternate across space. By contrast, in the bottom panel, all low levels of activity cluster on one side of a fixed point; all high levels, on the other side. However, both top and bottom panels imply the identical cross-section distribution, shown on the right side of Fig. 3: One half of the observations are recorded to have high activity; one half, low.

We conclude from Fig. 3 therefore that cross-section distributions

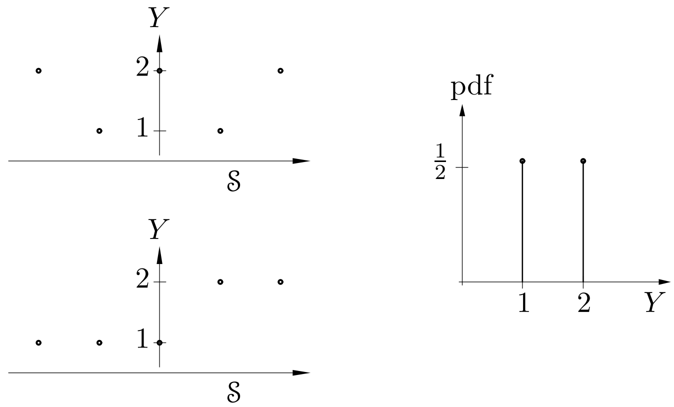

Figure 3: Y **uninformative for** $\mathcal{G}$ Different spatial configurations of activity on the left alias to an identical cross-section distribution on the right

or indexes calculated from them such as Gini or Herfindahl coefficients, by ignoring the spatial dimension in the underlying observations, are uninformative on spatial clustering in the sense described in the left bottom panel of the Figure.

But perhaps this difficulty arises from incorrectly looking at completely disaggregated point data. If in Fig. 3 the space $\mathcal{S}$ were divided into left and right halves, then the implied regional distribution would appropriately reveal no clustering for the top panel and extreme clustering for the bottom.

Aggregating into larger regions, however, raises yet other problems. Fig. 4 illustrates where four locations are invariant in physical space but arbitrary re-drawing of regional boundaries can lead to either extreme clustering—with all four active locations in a single region—or, the opposite, no clustering at all, with each region holding only a single active location. Geographers call this disfigurement of the underlying reality the *modifiable areal unit problem* or MAUP; in timeseries econometrics, aliasing problems like this are described in, among others, Sims (1971).

We can summarize the discussion as follows. Empirical analysis

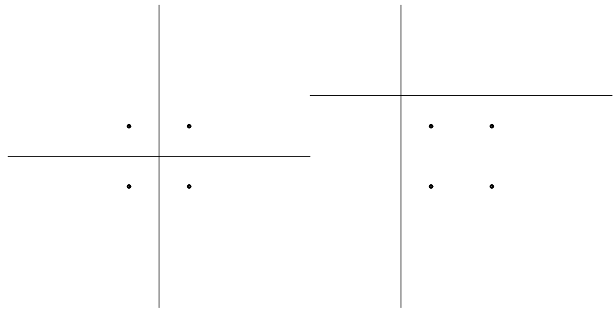

Figure 4: **Arbitrary discrete boundaries** On the left, economic activity is recorded to be located uniformly across regions. On the right, there appears to be clustering in just a single, distinguished region. However, the underlying points are invariant.

of location data such as in Fig. 1 needs to be explicit on how it uses information in the spatial dimension if it is to formalize usefully the notion of spatial clusters. Moreover, spatial clusters can take different forms, with two logically distinct and opposite forms usefully identified. First, a cluster might refer to a high concentration of dots in a map like Fig. 1. Second, given a set of dots, regularly spaced or otherwise, a cluster might refer a high level of economic activity over only one dot or only a relatively few dots. A complete empirical analysis would consider simultaneously these two extreme perspectives as well as all others in between—this is indeed what the analytical model in Section 3 attempts. Many other studies—empirical and theoretical—consider only the second. In the empirical analysis of Sections 4–5, we consider the first, to emphasize how our focus differs from most of the literature and to clarify the importance of the spatial relations dimension in empirical economic geography modelling.[3]

[3] Empirical studies using measures of market access (see, e.g., the references surveyed in Overman, Redding and Venables, 2001) study how given configurations across geographical space determine

To conclude this section, we should be clear about what we do not do. Business cycle models determine downturns and upturns in economic activity not in the sense of precise dating ("a downturn for the second quarter of 1992") but in the sense of a dynamic law of motion for economic variables of interest. We take the same approach here when we say an economic-geography model determines locations—the model should provide a spatial law of motion, not pronounce regarding, say, a particular street corner in London.

3 A Model of Activity and Location

This section describes a model determining a spatial law of motion in the level of activity, evolving through time. In the language of Section 2, the model endogenously determines both Y over $\mathcal{G}$ and $\mathcal{G}$.

The model is close to those in Quah (2000, 2001, 2002) which, in turn, draw on ideas in Krugman and Venables (1997) and Turing (1952) but with two significant departures, one theoretical and the other empirical: First, perfect foresight dynamic optimization is assumed and turns out to be crucial to developing the spatial law of motion.[4] Second, the underlying geography is not taken to be a fixed finite set of points, but instead a continuous space so that spatial clusters—local peaks in the density of economic activity in space—can be clearly distinguished. By contrast, with exogenously-specified discrete points on an continuous space, trivially, everything is a local peak.

economic activity at a location. What we aim for instead, in the highly stylized setup that follows, is the endogenous determination of those configurations themselves.

[4] Krugman and Venables (1997) provide a much richer description of their economic environment than we give below of ours. But the equilibrium there is static; the added-on dynamics are not based on any explicit economic reasoning. Interest there lies in divergent timepaths, leading to extreme spatial concentration. By contrast, in this paper, dynamics are central to the equilibrium concept and we seek instead to describe behavior along convergent timepaths.

To focus on these relatively technical, predictive features, the model is stripped down to a barebones, equilibrium structure. None of the rich layers of complexity—of the externalities from workers and jobs co-locating, from skills complementarity, from upstream-downstream inter-connected chains of production, and so on—that are described in thoughtful work elsewhere on this topic will play a role here.

Equilibrium in the model will be described by a partial differential equation in economic activity, one dimension in time, the other in space. The economic structure of the model determines how local peaks in space appear along the dynamic equilibrium path in spatial distributions. Clusters are not located potentially only over a small set of specified locations but emerge endogenously as peaks in an equilibrium spatial law of motion. Technically, those spatial clusters arise as specific linear combinations of sinusoidal eigenfunctions, activated along a saddlepath-stable trajectory to steady state equilibrium. They therefore trace out a tradeoff between the given historical distribution of activity across geography, on the one hand, and adjustment towards the steady-state distribution, on the other.

3.1 Space

Let the underlying space $\mathcal{S}$ be the continuous set of locations on the unit circle, $\mathcal{S} = \{\, z : |z| = 1 \,\}$. Identify each location $z \in \mathcal{S}$ with the corresponding angle $\omega \in (-\pi, \pi]$, so that $z = e^{i\omega}$, as in Fig. 5.

Assume economic activity Y sees productivity spillovers across space, i.e., there is a technology spatial-shift factor, describing how activity Y in one location affects output W elsewhere. Quantity Y can also be read as employment but we will continue to refer to it as economic activity, mainly to suggest its connection with the discussion elsewhere in this paper but also to denote its more general intent.

Represent the technology shift factor by a non-negative bounded function

$$\mathsf{A} : \mathcal{S} \times \mathcal{S} \to [0, 1]$$

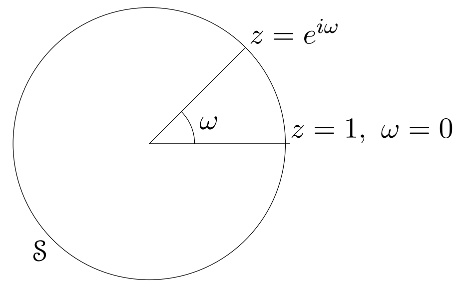

Figure 5: $\mathbb{S}$ **continuous set of locations on unit circle** Each location $z \in \mathbb{S}$ can be identified with angle $\omega \in (-\pi, \pi]$.

that achieves its maximum on the principal diagonal, i.e., $\mathsf{A}(z, z) = 1$ for all $z \in \mathbb{S}$. The map A might, additionally, be taken to evolve through time but that provides no additional insight so we assume A time-invariant. The value $\mathsf{A}(z', z)$ parameterizes how activity Y at location z' affects output W at z, i.e., let

$$W(z, t) = \left[\int_{\mathbb{S}} \left[\mathsf{A}(z', z) \cdot Y(z', t) \right]^{\gamma} dz' \right]^{1/\gamma}, \quad \gamma \in (0, 1), \qquad (3)$$

where $(\gamma - 1)^{-1}$ is the elasticity of substitution across spatial activity.

With no spatial spillovers, function A is zero everywhere except on the principal diagonal so that then $W = Y$. With spillovers highly localized in space, function A is small except in a neighborhood surrounding the principal diagonal. This specification bears some resemblance to standard iceberg transportation costs in economic geography but, in general, $\mathsf{A}(z', z)$ might equal 1 not only on $z' = z$, and A need be neither symmetric nor monotonically declining in Euclidean distance between z' and z.

A further property useful in A is radial homogeneity, i.e., $\mathsf{A}(z', z)$ varies only with $z' - z$, not with each argument separately. This says that all locations are, ex ante, identical; economic interaction varies with spatial relations, but not with specific location names. Note the variation is in the vector separation $z' - z$, not distance $|z' - z|$, as

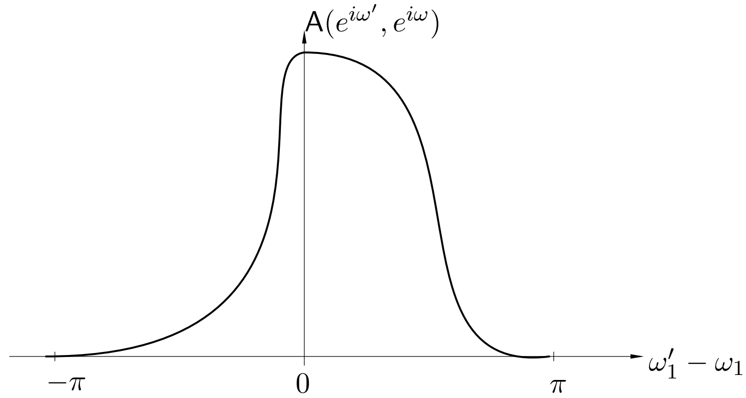

Figure 6: **Spillover — Technology shift factor** Spatial spillover $0 \leq \mathsf{A}(z', z) \leq 1$, and $\mathsf{A}(z, z) = 1$. In this example A is not symmetric. However, it does satisfy radial homogeneity, i.e., $\mathsf{A}(e^{i\omega'}, e^{i\omega})$ depends only on $\omega' - \omega$, not $|\omega' - \omega|$.

will happen with the definition of isotropic in Section 7 to follow.[5] A radially homogeneous spillover function is graphed in Fig. 6.

3.2 Individual optimization

The economy begins at time 0 with some arbitrary spatial density $Y(\cdot, 0)$ and proceeds forever. The evolution of economic activity Y across space will be determined in equilibrium by optimizing, forward-looking agents choosing how much activity to have where.

Without loss, let Y also denote the spatial density of economic activity by requiring for all t:

$$\forall z \in \mathcal{S} : \; Y(z, t) \geq 0 \quad \text{and} \quad \int_{\mathcal{S}} Y(z, t) \, dz = 1. \tag{4}$$

At each location $z \in \mathcal{S}$ agents receive returns according to their contribution to output not just where they are but across all $z' \in \mathcal{S}$; there

[5] Thus, if A were a matrix, it would be Toeplitz but not Hermitian.

are no unacknowledged externalities. The marginal contribution to $W(z', t)$ of $Y(z, t)$ is

$$\frac{\partial W(z', t)}{\partial Y(z, t)} = \mathsf{A}(z, z')^{\gamma} \left[\frac{Y(z', t)}{Y(z, t)} \right]^{1-\gamma},$$

so that integrating over all locations $z' \in \mathcal{S}$ gives

$$\begin{aligned} w(z, t) &= \int_{\mathcal{S}} \frac{\partial W(z', t)}{\partial Y(z, t)} \, dz' \\ &= \int_{\mathcal{S}} \mathsf{A}(z, z')^{\gamma} \left[Y(z', t)/Y(z, t) \right]^{1-\gamma} \, dz'. \end{aligned} \tag{5}$$

Equation (5) carries the intuitive feature that since $\gamma \in (0, 1)$, other things equal, higher $Y(z)$ lowers the return to further activity locating at z. At the same time, however, if $Y(z')$ is high in the appropriate neighborhood around z dictated by spillover $\mathsf{A}(z, z')$, so too the reward at the margin from economic activity clustering in that neighborhood.

The average return across space at time t is

$$\overline{w}(t) = (2\pi)^{-1} \int_{\mathcal{S}} w(z', t) \, dz'. \tag{6}$$

Agents at z take as given the payment $w(z, t) - \overline{w}(t)$ per unit of economic activity $Y(z, t)$. While, other things equal, agents seek to locate where this quantity is largest, spatial economic activity will optimally change only gradually through time because of adjustment costs:

$$C(\, \dot{Y}(z, t) \,) = \frac{1}{2} \zeta \, \dot{Y}(z, t)^2, \quad \zeta > 0, \tag{7}$$

where $\dot{Y}$ denotes the partial derivative with respect to time, $\partial Y / \partial t$. The form of costly adjustment in equation (7) also implies that at any time t the current spatial density $Y(\cdot, t)$ is taken as given.

Agents are forward-looking, discount the future at a constant rate $\rho > 0$, and choose a timepath in economic activity for where they locate, taking as given the reward and cost structure in equations (5)–(7) and their expectations of the choices made by all other agents elsewhere in space. Equilibrium is when every selected timepath is consistent with expectations and economic activity across space constitutes at each timepoint a density—the counterpart in this model of factor input market-clearing.

At time t agents at $z \in \mathcal{S}$ solve

$$\sup_{\{\,Y(z,s):s\geq t\,\}} \int_{s\geq t} e^{-s\rho} \left[\, (\, w(z,s) - \overline{w}(s)\,)\, Y(z,s) - C(\,\dot{Y}(z,s)\,)\,\right] \, ds \tag{8}$$

$$\begin{gathered}\text{s.t. equations (5)–(7),}\\ \left\{\, w(z',s) : s \geq t,\ z' \in \mathcal{S}\,\right\},\\ \left\{\, Y(z',t) : z' \in \mathcal{S}\,\right\},\ \text{and}\\ \left\{\, Y(z',s) : s \geq t,\ z' \neq z\,\right\}.\end{gathered}$$

To simplify and close the system, assume that output W is sold on international markets outside of $\mathcal{S}$ at some exogenous price and so can hereafter be ignored.

From the usual "stable roots backwards, unstable roots forwards" solution procedure (Sargent, 1987), the decision rule solving problem (8) has

$$\dot{Y}(z,t) = \zeta^{-1} \int_0^\infty e^{-s\rho} \left[\, w(z,t+s) - \overline{w}(t+s)\,\right] \, ds. \tag{9}$$

Equation (9) says that economic activity in any location changes with the expected present discounted value of returns. Since ζ is positive, the adjustment $\dot{Y}$ is more muted, the larger is ζ. Conditional on ζ, the rule is to raise $\dot{Y}$ whenever compensation at that location is high relative to average compensation elsewhere, with the comparison undertaken in expected present discounted value.

3.3 Equilibrium distribution dynamics

At time 0 the spatial density $Y(\cdot, 0)$ is given. Combining this initial condition with expectations for

$$\left\{\, Y(z',t) : z' \in \mathcal{S},\ t > 0\,\right\}$$

determines $\dot{Y}(z,t)$ from equation (9), using equations (5) and (6).

Equilibrium is a mapping Y from $[0,\infty]$ to the space of probability densities on $\mathbb{S}$, such that when $Y(\cdot,t) = \mathsf{Y}(t)$, using $Y(\cdot,0)$ and equations (5), (6), and (9) recovers the mapping Y as a fixed point. In words, equilibrium describes a timepath for the spatial density of economic activity such that all locations optimize program (8) and undertake actions consistent with one another. A *steady-state equilibrium* is an equilibrium that is constant, i.e., $\mathsf{Y}(t+s) = \mathsf{Y}(t)$ for all $s \geq t$.

In this model, no clustering is always a steady-state equilibrium. The constant uniform density for Y implies $Y(z,t) = Y(z',t)$ so that, in equation (5),

$$w(z,t) = \int_{\mathbb{S}} \mathsf{A}(z,z')\, dz' = \int_{|z'|=1} \mathsf{A}(z-z')\, dz'.$$

But then, for all $z_0, z_1 \in \mathbb{S}$ we have $w(z_0,t) = w(z_1,t) = \overline{w}(t)$ reproducing, from decision rule (9), an unchanging uniform density Y over $\mathbb{S}$. This no-clustering steady state implies also that any clustering that arises does not do so purely mechanically from, say, the shape of spatial technology shifts assumed in A.

Outside of steady state, the situation becomes both more difficult and more intricate. While many other interesting equilibria might exist, we consider only those that satisfy a Markov property and that we can characterize using the method of undetermined coefficients.

Assume equilibrium is Markov, i.e., confine attention to mappings Y where there exists an operator T mapping the space of densities on $\mathbb{S}$ to the space of their time derivatives,

$$\dot{\mathsf{Y}}(t) = (\mathsf{T}\,\mathsf{Y})(t) \quad \forall t \in [0,\infty). \tag{10}$$

In equilibrium, with $Y(\cdot,t) = \mathsf{Y}(t)$, the stochastic kernel representation of equation (10) is:

$$\dot{Y}(z,t) = \int_{\mathbb{S}} M_{t,Y(\cdot,t)}(z',z) Y(z',t)\, dz' \quad \forall\ z \text{ and } t, \tag{11}$$

where $M_{t,Y(\cdot,t)}$ is the adjoint of operator T, and we have subscripted M explicitly to denote the nonlinearity on the right side of equation (11).

Use equation (10) to generate a candidate equilibrium timepath and plug it into the reward equations (5) and (6). By the Markov property the resulting timepaths for w and $\overline{w}$ depend only on the initial spatial density $Y(\cdot, t)$. Substituting these timepaths into the right side of the decision rule (9) gives an equation for $\dot{Y}(z,t)$ that again depends only on the spatial density at time t, $Y(\cdot, t)$. In this solution procedure, the right side of (9) has changed from an integral over time $s \geq t$ to an integral over $z' \in \mathcal{S}$ at t fixed. Comparing this outcome with the operator adjoint equation (11)—the method of undetermined coefficients—allows us to restrict M and thereby the Markov operator T.

Models of this kind have been studied in Quah (2000, 2001, 2002), adapting ideas from the discrete versions considered in Krugman and Venables (1997) and Turing (1952). Radial homogeneity in A and symmetry across $\mathcal{S}$ imply that operator T has a Toeplitz property: If T were a $p \times p$ matrix, its spectrum would equal the discrete Fourier transform of any fixed row of T; its eigenfunctions would be the orthonormal set of complex exponentials $(2\pi)^{-1/2}e^{i\omega j}$, with ω evaluated on the collection of discrete Fourier frequencies $k2\pi/p$, integer k from 0 to $p-1$ (Grenander and Szegö, 1958). In the general case operator T continues to have for its spectrum a Fourier transform and its eigenfunctions the (invariant) orthonormal set of complex exponentials (Quah, 2000, 2001, 2002).

To be clear about what has been achieved here, it is useful to observe that the model has, in equations (5), (6), and (9), extensive inter-relations across space and time. The decision rule (9), in particular, holds dynamically not only for a single fixed location z but simultaneously across the continuum $z \in \mathcal{S}$. The model therefore produces a real-valued partial-differential equation (PDE) describing equilibrium—a law of motion across both space and time. What the Markov operator representation in equations (10)–(11) allows is analysis of the equilibrium as a single ordinary differential equation (ODE) in time. Instead of a PDE, we analyze only an ODE but one taking values in an infinite-dimensional state space of probability densities on $\mathcal{S}$.

Although a complete characterization of the spatial distribution dynamics is unavailable at this level of generality, the analysis provides two statements.[6] First, the spectrum of T can be written as the difference between a complex-valued function σ and a positive constant c depending on the model parameters (γ, ζ, ρ) and increasing in the adjustment cost coefficient ζ (Fig. 7). Second, the eigenfunctions of T are complex exponentials—invariant to the precise form of T, instead following only from T Toeplitz—so that nontrivial, nondegenerate linear combinations of the eigenfunctions generate sinusoidal functions on $\mathcal{S}$ or, more vividly, clusters in space (Fig. 8).

An implication of these two statements is useful to note for understanding the resulting dynamics. As the system (10)–(11) evolves and the state—the spatial density—changes through time, the stochastic kernel M in equation (11) will, in general, also evolve. But although its spectrum changes, the operator's eigenfunctions always remain invariant. Thus, the active eigenfunctions—those corresponding to the parts of the spectrum that are not nullified to preserve saddlepath stability—are simply multiplied into different scalars to produce the transition dynamics in the spatial density, a snapshot of which is given in Fig. 8.

For saddlepoint stability, portions of the spectrum with positive real parts—unstable eigenvalues—have to be nullified by appropriate choice of the state variable Y. Fig. 7 shows how, with changing c, those unstable parts of the spectrum will expand or contract, and complementarily the stable parts of the spectrum, denoted in the Fig. by $\mathbb{J}'$. Spatial densities Y that imply convergence to the uniform steady state are nondegenerate linear combinations in the operator T's eigenfunctions and therefore show clusters, simultaneously determining $(Y, \mathcal{G})$, as in Fig. 8.

[6] For brevity, the calculations are not given here; they would simply repeat those in Quah (2000, 2001, 2002).

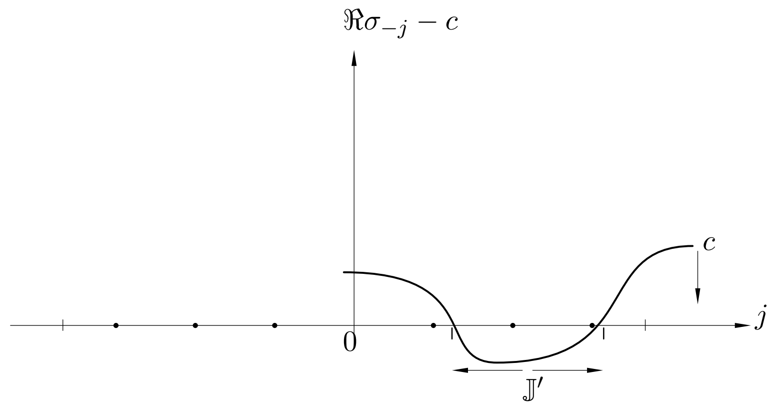

Figure 7: **Real part of the spectrum of T** Displaced downwards with increasing c, the resulting components, $\mathbb{J}'$, activate associated complex exponentials in the invariant family of eigenfunctions.

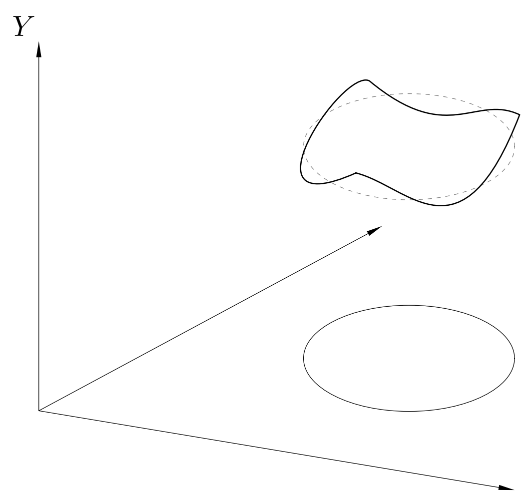

Figure 8: **Waves in space** Saddlepoint stability in equilibrium determines $(Y, \mathcal{G})$ jointly evolving through time.

4 Statistical framework

Models such as that in Section 3 lend themselves naturally to interpreting the data in Fig. 1 and Fig. 2. Potential clusters are not fixed prior to the analysis and the densities of points observed in those UK maps lend themselves naturally to interpretation as empirical counterparts of unobserved theoretical densities.

An appropriate framework then for empirically modelling spatial clusters should address how the spatial density changes, conditional on the realizations of events within arbitrary subsets of the underlying geographical space. But this exactly describes a spatial *hazard function*. The simplest hazard function is just a constant, and that correctly describes a situation of no clustering.

Notice that, appropriately, a constant hazard will not produce point events that are appear regularly distributed across space. Indeed, such regularity will be evidence not of the absence of clustering in the underlying model, but instead, the opposite, it will indicate inhibition, or locations repelling one another. Similarly, apparent spatial clustering need not be evidence of clustering in an underlying model—a non-constant hazard—but instead simply the expected outcome of a constant hazard spatial probability model.

To describe the constant hazard basecase and the extensions that we use in our empirical analysis, establish the following notation (some of which conforms to casual everyday terminology and so was already used earlier). Without loss, suppress the time argument, and say that whenever $Y(z) > 0$ location z is an *event*; call $\mathcal{G} \subset \mathcal{S}$ from equation (2) the set of events. For arbitrary region $S \subset \mathcal{S}$ let $N(S)$ denote the number of events contained in S, and write $|S| = \nu(S)$ for the area or Lebesgue measure of S. Denote by dz an infinitesimal region containing z. The random measure N is sometimes what is referred to as the point process (e.g., Cressie, 1993) but in this paper we employ the more direct reference to the collection of events themselves. The two are formally equivalent.

A constant spatial hazard function describes a complete spatial randomness. An alternative description useful for empirical analysis is the following.

Definition 4.1 *For λ a fixed positive real number, the collection of events $z_1, z_2, z_3, \ldots$ in $\mathcal{S}$ is a* ***stationary spatial Poisson process with intensity*** *λ if:*

1. *for every finite region $S \subset \mathcal{S}$ the number of events $N(S)$ is distributed Poisson with mean $\lambda \cdot \nu(S)$; and*
2. *for every finite region $S \subset \mathcal{S}$, conditional on $N(S) = n$, the events $z_1, z_2, z_3, \ldots, z_n$ constitute an iid sample from the uniform distribution over set S.*

Properties 1. and 2. in Defn. 4.1 imply that for every disjoint pair of regions S_1, S_2 the random variables $N(S_1)$ and $N(S_2)$ are independent.

As in time series, while stationarity is precisely defined as the invariance of joint probability distributions across displacements, many different departures producing nonstationarity are possible. One natural alternative has the intensity varying:

Definition 4.2 *Suppose $\lambda : \mathcal{S} \to \mathbb{R}_+$ is a given non-constant function on $\mathcal{S}$. The collection of events $z_1, z_2, z_3, \ldots$ in $\mathcal{S}$ is an* ***inhomogeneous spatial Poisson process*** *if:*

1. *for every finite region $S \subset \mathcal{S}$, the number of events $N(S)$ has Poisson distribution with mean $\int_S \lambda(z)\, dz$; and*
2. *conditional on $N(S) = n$, the locations $z_1, z_2, \ldots, z_n$ in $\mathcal{S}$ form an independent random sample from that distribution on S with density proportional to $\lambda(z),\ z \in S$.*

Call λ the ***intensity function*** *of the Poisson process.*

As suggested by the terminology, the inhomogenous processes in Defn. 4.2 are closely related to the stationary spatial Poisson processes of Defn. 4.1. The critical difference is that for the former the expected number of events in space varies as the intensity function does, whereas for the latter, the expected number of events is constant per unit area. In both, however, spatial independence is maintained, and so neither captures any intricate web of spatial relationships as might appear in a model like that in Section 3.

Despite this, however, inhomogeneity is useful in that it allows explicit conditioning, i.e., a researcher can model the intensity function as depending on covariate explanatory variables

$$\lambda(z) = \lambda\left(X(z)'\theta \right), \quad z \in \mathcal{S},\ \theta \in \mathbb{R}^{\dim(\theta)}, \tag{12}$$

where $X(z)$ is a $\dim(\theta)$-dimensioned vector of explanatory variables, observed in location z, and θ is a unknown parameter vector to be estimated. Clustering then is generated when explanatory variables take on values implying an especially high intensity. Figs. 9–11 and Table 1 in Section 5 to follow present estimates to illustrate in UK manufacturing spatial inhomogeneity and clustering, and potential explanations for them.

In the inhomogeneous Poisson process of Defn. 4.2, clustering and therefore conditional spatial dependence arise from specific variation in the intensity function λ. For time-series business cycles, this would be equivalent to explaining economic fluctuations using only contemporaneous movements in exogenous explanatory variables. More insightful in the study of business cycles has been to allow endogenous propagation mechanisms. The counterpart for spatial empirics is true endogenous spatial dependence, dropping the Poisson feature shared by both Defns. 4.1 and 4.2.

Suppose that the intensity function λ in Defn. 4.2 is itself stochastic. The resulting point process is then a doubly stochastic process driven by an underlying probability mechanism that can be used to model explicitly spatial dependence, through cross-correlation in the random λ.

Definition 4.3 (Cox process) *Let*

$$\{\, \Lambda(z) : z \in \mathcal{S} \,\}$$

be a non-negative valued stochastic process on $\mathcal{S}$. *The collection of events* $z_1, z_2, z_3, \ldots$ *in* $\mathcal{S}$ *is a* ***Cox process*** *driven by* Λ *if conditional on* $\Lambda = \lambda$, *the collection constitutes an inhomogeneous Poisson process with intensity* λ.

While it is intuitive that the Cox process in Defn. 4.3 just given can produce clustering and spatial dependence in the data Y without exogenous covariates X, it is nonetheless useful to see this explicitly. To this end we define second-order intensity and introduce an explicit cluster (Neyman-Scott) process.

From Defns. 4.1–4.2 the intensity function is also:

$$\lambda(z) = \lim_{|dz| \to 0} \frac{N(dz)}{|dz|} = \lim_{\nu(dz) \to 0} \frac{N(dz)}{\nu(dz)}, \quad z \in \mathcal{S}. \tag{13}$$

Thus, we can usefully define *second-order intensity* by:

$$\lambda_2(z, z') = \lim_{\substack{\nu(dz) \to 0 \\ \nu(dz') \to 0}} \frac{E\,[\,N(dz)N(dz')\,]}{\nu(dz)\nu(dz')}, \quad z,\ z' \in \mathcal{S}. \tag{14}$$

For the Cox process in Defn. 4.3 we have:

$$\lambda(z) = E\,[\,\Lambda(z)\,] \quad \text{and} \quad \lambda_2(z, z') = E\left[\,\Lambda(z)\Lambda(z')\,\right].$$

More generally, interpreting λ in equation (13) as the mean or first-order moment function, λ_2 in equation (14) is the (uncentered) covariogram or second-order moment function, defined on $\mathcal{S} \times \mathcal{S}$. Similarly, just as intensity λ describes the expected number of events occurring over a given region, second-order intensity λ_2 characterizes the expected additional events relative to a given event. To see this explicitly, define:

Definition 4.4 (Ripley's K-function) *For $r \in \mathbb{R}_+$ and $z \in \mathcal{S}$ define the* ***ball at z of radius*** *r:*

$$B(r, z) \stackrel{\text{def}}{=} \{\, z' \in \mathcal{S} : \|z - z'\| \le r \,\} \tag{15}$$

(where $\|\cdot\|$ denotes Euclidean distance) and

$$\begin{aligned} K(r) &\stackrel{\text{def}}{=} E\,[\,N(B(r, z)) - 1 \text{ for randomly selected event } z\,] \\ &= E\,[\,E\,[\,N(B(r, z)) - 1 \mid z \text{ event}\,]\,]. \end{aligned} \tag{16}$$

Ripley's K-function in equation (16) calculates the expected number of additional events located in a ball surrounding a randomly chosen event. The function is sometimes defined to be scaled by λ^{-1} for homogeneous Poisson processes (e.g., Diggle (1983) and Cressie (1993)). Function K explicitly quantifies spatial dependence and clustering.

Related to covariance stationarity in timeseries analysis, suppose that second-order intensity λ_2 depends only on the Euclidean distance between its two arguments, i.e.,

$$\lambda_2(z, z') = \lambda_2(\|z - z'\|).$$

Then, under regularity conditions given in the technical appendix Section 7, second-order intensity and Ripley's K are related in an intuitive way:

$$K(r) = \frac{2\pi}{\lambda} \int_0^r s\lambda_2(s)\, ds \tag{17}$$

or

$$\lambda_2(r) = \frac{\lambda}{2\pi} \times \frac{K'(r)}{r},$$

i.e., up to proportional scaling, second-order intensity is just the slope of the K-function. For example, the stationary Poisson process of Defn. 4.1 has $\lambda_2(r) = \lambda^2$ so that its K-function is just

$$K(r) = \frac{2\pi}{\lambda}\, \lambda^2 \int_0^r s\, ds = \pi\lambda r^2.$$

Because the analytical form to K can be found for several interesting cases and its empirical counterpart readily estimated, we can use K both to estimate underlying parameters (through a version of nonlinear least squares) and, through equation (17), explicitly analyze spatial dependence. We do this in Section 5 for the following process that explicitly generates spatial clusters:

Definition 4.5 (Poisson cluster process) *Let μ be a positive real number,* Pr *be a pdf over the non-negative integers, and* f *a probability density over* $\mathcal{S}$*.*

1. *Generate unobserved primary events $\zeta_1, \zeta_2, \zeta_3, \ldots$ on $\mathcal{S}$ as a stationary spatial Poisson process with intensity μ;*

2. *let each* ζ_j *seed a cluster comprising* J_j *secondary events*

$$z_1(\zeta_j), z_2(\zeta_j), \ldots, z_{\mathsf{J}_j}(\zeta_j) \in \mathcal{S},$$

3. *where* J *is iid* Pr*; and*

4. *conditional on* (ζ_j, J_j) *secondary events locate on* $\mathcal{S}$ *as* J_j *iid draws centered on* ζ_j *with clustering probability density* f.

Then the collection of events $z_1, z_2, z_3, \ldots$ *is a* ***Poisson cluster process*** *with parameters* $(\mu, \mathrm{Pr}, \mathsf{f})$.

By its definition the Poisson cluster process generates spatial clusters explicitly. What is of interest to us is quantification of the parameters Pr and f—the number of events in a typical cluster and the spatial spread of those events. Section 5 presents estimates of these for a number of UK manufacturing industries.

Although we will not use them in this paper, obvious generalizations of the process given in Defn. 4.5 are possible: the draws on $(\mathrm{Pr}, \mathsf{f})$ can be made general rather than iid; primary events can be drawn from inhomogeneous spatial processes; and so on. In spatial statistics, it is these generalizations that are properly called Poisson cluster processes. For our purposes, however, the special case will already be intricate enough to analyze, and will produce sufficiently rich clustering possibilities. As a side-issue, it is worth noting also that, as intuition might already suggest, certain Cox processes are Poisson cluster processes and certain Poisson cluster processes are, in turn, Cox processes (Cressie, 1993, Ch. 8).

5 Empirical application

We begin the empirical analysis by provisionally adopting the framework of inhomogenous spatial Poisson processes given in Defn. 4.2. From the data such as in Figs. 1–2 we can estimate intensity functions for locations in different industries. We present these in Figs. 9–11.

[TBW]

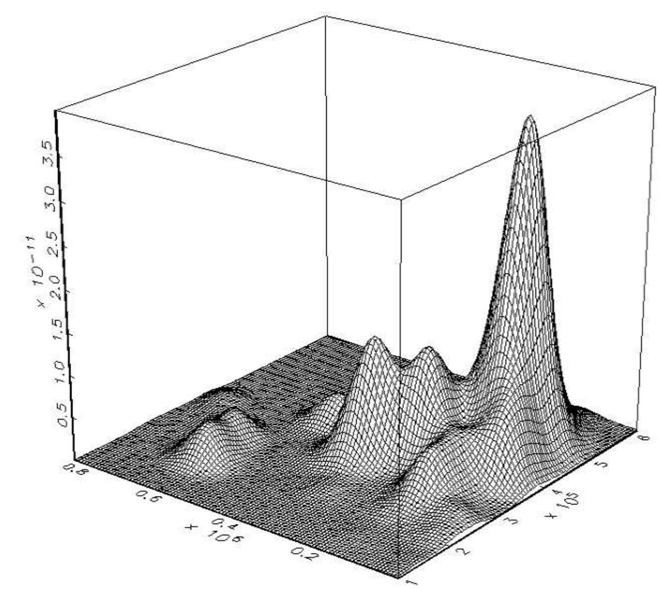

Figure 9: **Estimated intensity of computer industry establishments in the UK** Perspective is from the southwest. The base of the Figure includes some ocean surrounding the British Isles. The highest peak denotes the London cluster; smaller peaks in the north indicate Silicon Glen in Scotland

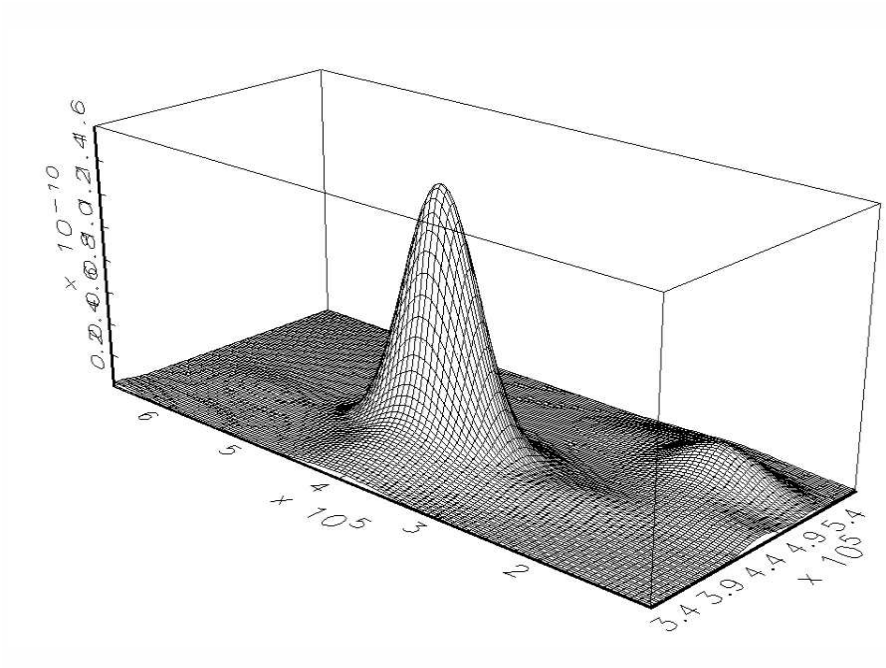

Figure 10: **Estimated intensity of cutlery industry establishments in the UK** Perspective is again from the southwest. See caption to Fig. 9. The peak is Sheffield.

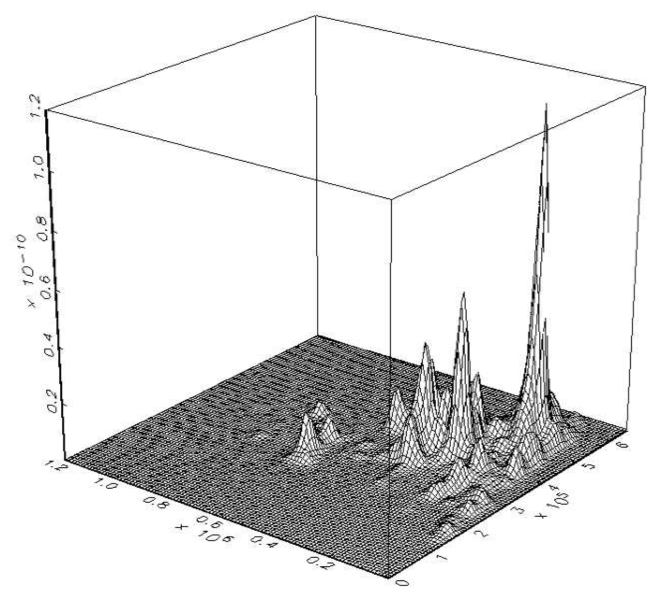

Figure 11: **Estimated intensity of all manufacturing industry establishments in the UK** Perspective is again from the south-west. See caption to Fig. 9

6 Conclusion and extensions

In this paper we have provided theoretical and empirical models for analyzing fluctuations in economic activity across space.

Our work departs from earlier related analyses in economic geography in two respects: First, we sought explicit spatial laws of motion across geography—both analytical and empirical—so that, simultaneously, both how much and where clustering occurs are jointly determined. Second, instead of providing yet further indexes to capture this or that aspect of clustering and localization, the explicit spatial laws of motion we analyze facilitated directly comparing theoretical predictions and empirical descriptions. They are therefore useful for future work intending structural estimation.

In our empirical work we found ...

[TBW]

Section 3 developed an analytical model that, although abstract and simple, generates predictions richer than can be captured in our probability models of Sections 4. In particular, the theoretical model provides statements not just about location but also on quantity of economic activity and on dynamics. Thus, a more appropriate statistical model is one of spacetime marked point processes. The introduction of marks, i.e., quantity indicators on point events, cannot be modelled just by hypothesizing multiple events piled on top of a singular location, for the process would then no longer be orderly (see the Technical Appendix, Section 7). Allowing that multiplicity would invalidate both the Poisson-based distributions that have been assumed throughout and the simple relationship between spatial dependence and K-functions. Secondly, the dynamic dimension will grow in importance as more data of the kind we have studied in this paper are collected over time. These more intricate statistical modelling are areas we intend to study in future research.

7 Technical appendix

This appendix has four goals: First, it establishes the explicit relation between spatial dependence and Ripley's K-function given in equation (17). Second, it provides some more explicit technical discussion on the almost-equivalence of Cox processes and Poisson cluster processes in Defns. 4.3 and 4.5. Third, it describes the estimation procedures we used for Section 5. Last, it provides additional technical references.

7.1 Spatial dependence and Ripley's K-function

Following equations (13)–(14) call a spatial process *second-order stationary* when its intensity λ is constant and its second-order intensity varying only with the vector difference between its arguments, i.e.,

$$\forall z, z' \in \mathcal{S} : \quad \lambda_2(z, z') = \lambda_2(z'') \text{ for } z'' = z - z',$$

or, that its second-order intensity is translation-invariant. Call a second-order stationary spatial process *isotropic* when its second-order intensity depends only on the scalar distance between its arguments, i.e.,

$$\forall z, z' \in \mathcal{S} : \quad \lambda_2(z, z') = \lambda_2 \left(\|z - z'\| \right).$$

Finally, say that a spatial process is *orderly* if every point in $\mathcal{S}$ almost surely carries at most a single event, i.e.,

$$\text{Prob} \{ N(dz) > 1 \} = o(|dz|).$$

For an orderly process,

$$E[N(dz)] - \text{Prob} \{ N(dz) = 1 \} = o(|dz|)$$

and

$$E \left[N(dz)N(dz') \right] - \text{Prob} \left\{ N(dz) = N(dz') = 1 \right\} = o \left(\max \left(|dz|, |dz'| \right) \right);$$

and we can calculate its conditional intensity, conditional on any random variable or event X, as

$$\begin{aligned}\lambda(z \mid X) &= \lim_{\nu(dz)\to 0} \frac{E\,(\,N(dz) \mid X\,)}{\nu(dz)} \\ &= \lim_{\nu(dz)\to 0} \frac{\text{Prob}\,(\,N(dz) = 1 \mid X\,)}{\nu(dz)}\ .\end{aligned}$$

Consider then the probability of an event at z conditioned on z' already holding an event:

$$\begin{aligned}&\text{Prob}\left\{\,N(dz) = 1 \mid N(dz') = 1\,\right\} \\ &\qquad = \frac{\text{Prob}\,\{N(dz) = N(dz') = 1\}}{\text{Prob}\,\{N(dz') = 1\}}\ .\end{aligned}$$

Therefore,

$$\begin{aligned}&\lambda(z \mid N(dz') = 1) \\ &\quad = \lim_{\nu(dz)\to 0} \left[\frac{\text{Prob}\,\{N(dz) = N(dz') = 1\}}{\text{Prob}\,\{N(dz') = 1\}} \times \frac{1}{\nu(dz)}\right],\end{aligned}$$

so that the conditional intensity

$$\begin{aligned}&\lambda(z \mid z' \text{ holds an event}) \\ &\quad = \lim_{\substack{\nu(dz)\to 0\\ \nu(dz')\to 0}} \frac{E\,[\,N(dz)N(dz')\,]}{\nu(dz)\nu(dz')} \times \left(\frac{E\,[N(dz')]}{\nu(dz')}\right)^{-1} \\ &\quad = \lambda_2(z, z') \cdot \lambda(z)^{-1}.\end{aligned}$$

Thus, when the spatial process is orderly and isotropic on $\mathbb{R}^2$, we have:

$$\begin{aligned}K(r) = E\left[\int_{z'\in B(r,z)} \lambda(z' \mid z)\, dz'\right] &= \int_0^r \frac{\lambda_2(s)}{\lambda} \cdot (2\pi s)\, ds \\ = \frac{2\pi}{\lambda} \int_0^r s\lambda_2(s)\, ds,\end{aligned}$$

implying

$$\lambda_2(r) = \frac{\lambda}{2\pi} \times \frac{K'(r)}{r},$$

as in equation (17) of Section 4.

7.2 Cox processes and clustering

To see the relation between Cox processes and Poisson cluster processes, suppose that a Cox process has the random driver intensity:

$$\Lambda(z) = \omega \sum_{j=1}^{\infty} \mathsf{f}(z - z_j), \quad z \in \mathcal{S},$$

where $z_1, z_2, z_3, \ldots$ are the events of an inhomogeneous Poisson process with intensity $\lambda : \mathcal{S} \rightarrow \mathbb{R}_+$ and f is a pdf on $\mathcal{S}$. Then the probability characteristics of this Cox process are identical to those of a Poisson cluster process that has primary events with intensity $\mu = \lambda$ and observed secondary events with clustering probability density f and number generating probability Pr that is Poisson with mean ω. Cressie (1993, pp. 663–664) provides calculations for this example and discusses further the almost-equivalence between Cox and Poisson cluster processes.

7.3 Estimation

For Section 5 nonparametric estimates of intensity λ were obtained using a kernel-smoothed two-dimensional density estimator. It is these that are presented as Figs. 9–11.

Given covariates X and data $z_1, z_2, \ldots, z_n \in \mathcal{S}$, the inhomogeneous Poisson process of Defn. 4.2 has log-likelihood:

$$\ell(\theta) = \sum_{j=1}^{n} \log \lambda_\theta(z_j) - \int_{\mathcal{S}} \lambda_\theta(z') d\nu(z'),$$

where in Section 5 we used the exponential function

$$\lambda_\theta(z) = \exp\left(X(z)'\theta \right).$$

Under regularity conditions, the maximum-likelihood estimator is consistent and asymptotically normal.

For the Poisson cluster processes given in Defn. 4.5, maximum-likelihood estimation is usually intractable in practice, although conceptually straightforward. Let the typical realization λ_θ of the Λ process be parameterized by vector $\theta \in \mathbb{R}^{\dim(\theta)}$, and generate randomness in the driver intensity Λ by endowing on θ a pdf $\mathsf{g}(\theta; \vartheta)$ with hyperparameters $\vartheta \in \mathbb{R}^{\dim(\vartheta)}$. Given observed point data $z_1, z_2, \ldots, z_n \in \mathcal{S}$, the likelihood function then is

$$\mathcal{L}(\vartheta) = \int_{\mathbb{R}^{\dim(\theta)}} \widetilde{\mathcal{L}}(\theta) \mathsf{g}(\theta; \vartheta) \, d\theta$$

where the conditional likelihod

$$\widetilde{\mathcal{L}}(\theta) = \left(\prod_{j=1}^{n} \lambda_\theta(z_j) \right) \exp \left\{ - \int_{\mathcal{S}} \lambda_\theta(z') \, d\nu(z') \right\}.$$

A Gibbs sampler will compute these as well as more elaborate related likelihood functions (e.g., Wolpert and Ickstadt, 1998).

However, rather than attempt maximum-likelihood estimation for Cox and Poisson cluster processes in this paper, we employ instead in Section 5 an ad hoc weighted nonlinear least squares procedure working off Ripley's K-function. If, for the Poisson cluster process in Defn. 4.5, we specify the primary process ζ to have unknown constant intensity $\mu = \theta_1$; the pdf Pr to be Poisson with mean θ_2; and the clustering probability density f to be Gaussian on $\mathbb{R}^2$ with mean vector 0 and variance-covariance matrix $\theta_3 \cdot I$, then Ripley's K-function can be shown, following Cressie (1993, 8.5.3), to be

$$K(r, \theta) = \left[\pi r^2 + \theta_1^{-1} \cdot \left\{ 1 - e^{\frac{-r^2}{4\theta_3}} \right\} \right] \times \theta_1 \theta_2.$$

For the estimates for θ in Section 5, we solved:

$$\min_{\theta \in \mathbb{R}^3_+} \mathsf{D}(\theta) = \int_0^{\bar{r}} \left\{ \widehat{K}(r)^c - K(r, \theta)^c \right\}^2 \mathsf{W}(r) \, dr,$$

where $\widehat{K}$ is estimated empirically and where we experimented with the tuning coefficient c over $[0.25, 0.50]$, allowed bandwidth $\bar{r}$ to vary, and selected $\mathsf{W} : \mathbb{R}_+ \to \mathbb{R}_+$ as a non-negative weighting kernel.

We used both tsrf and the computer package `Gauss` to produce the numerical results in Section 5.

7.4 Further references

Daley and Vere-Jones (1988) and Cressie (1993) describe mathematical foundations for the spatial point process theory in this paper. Other useful references include Cox and Isham (1980), Diggle (1983), and Ripley (1977).

By recognizing that in equation (13) the intensity λ is the Radon-Nikodym derivative of count measure N with respect to Lebesgue measure, we can generalize the class of spatial processes considered to take into account yet richer specifications of spatial dependence (e.g., Wolpert and Ickstadt, 1998).

References

Ciccone, Antonio, and Robert E. Hall (1996) "Productivity and the density of economic activity," *American Economic Review* 86(1), 54–70, March

Cox, David R., and Valerie Isham (1980) *Point Processes* (London: Chapman and Hall)

Cressie, Noel (1993) *Statistics for Spatial Data* (New York: John Wiley)

Daley, Daryl J., and D. Vere-Jones (1988) *An Introduction to the Theory of Point Processes* Springer Series in Statistics (New York: Springer-Verlag)

David, Paul A., Dominique Foray, and Jean-Michel Dalle (1998) "Marshallian externalities and the emergence and spatial stability of technological enclaves," *Economic Innovation and New Technology* 6, 147–182

Devereux, Michael P., Rachel Griffith, and Helen Simpson (1999) "The geographic distribution of production activity in the UK," Working Paper 26/99, IFS, London, September

Diggle, Peter J. (1983) *Statistical Analysis of Spatial Point Patterns* (New York: Academic Press)

Duranton, Gilles, and Henry G. Overman (2002) "Testing for localisation using micro-geographic data," Working Paper, LSE, London, December

Ellison, Glenn, and Edward L. Glaeser (1997) "Geographic concentration in U.S. manufacturing industries: A dartboard approach," *Journal of Political Economy* 105(5), 889–927, October

Fujita, Masahisa, Paul Krugman, and Anthony Venables (1999) *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge: MIT Press)

Grenander, Ulf, and Gabor Szegö (1958) *Toeplitz Forms and Their Applications,* 2nd ed. (New York: Chelsea Publishing Company)

Henderson, J. Vernon (2003) "Marshall's scale economies," *Journal of Urban Economics* 53(1), 1–28, January

Ioannides, Yannis M., and Henry G. Overman (2000) "Spatial evolution of the US urban system," Working Paper, Tufts University and LSE, London, May

Jaffee, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993) "Geographic localization of knowledge spillovers as evidenced in patent citations," *Quarterly Journal of Economics* 108(3), 577–598, August

Kolko, Jed (2002) "Silicon mountains, silicon molehills: Geographic concentration and convergence of Internet industries in the US," *Information Economics and Policy* 14(2), 211–232, June

Krugman, Paul (1991) *Geography and Trade* (Cambridge: MIT Press)

Krugman, Paul, and Anthony J. Venables (1997) "The seamless world: A spatial model of international specialization," Working Paper, LSE, London, April

Overman, Henry G., Stephen Redding, and Anthony J. Venables (2001) "The economic geography of trade, production, and income: A survey of empirics," Discussion Paper 2978, CEPR, London, September

Quah, Danny (2000) "Internet cluster emergence," *European Economic Review* 44(4–6), 1032–1044, May

——— (2001) "Demand-driven knowledge clusters in a weightless economy," Working Paper, Economics Dept., LSE, London, April

——— (2002) "Spatial agglomeration dynamics," *American Economic Review (Papers and Proceedings)* 92(2), 247–252, May

Rauch, James E. (1993) "Productivity gains from geographic concentration of human capital: Evidence from the cities," *Journal of Urban Economics* 43(3), 380–400, November

Ripley, Brian D. (1977) "Modelling spatial patterns (with discussion)," *Journal of the Royal Statistical Society, Series B* 39(2), 172–212

Sargent, Thomas J. (1987) *Macroeconomic Theory,* Second ed. (New York NY: Academic Press)

Sims, Christopher A. (1971) "Discrete approximations to continuous time lag distributions in econometrics," *Econometrica* 39(3), 545–564, July

Turing, Alan M. (1952) "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London Series B* 237, 37–72, August

Wolpert, Robert L., and Katja Ickstadt (1998) "Poisson/gamma random field models for spatial statistics," *Biometrika* 85, 251–267