[[Draft. Incomplete]]

Cluster Emergence on a Global Continuum

by

Danny Quah *
LSE Economics Department
October 1999

*I thank the British Academy, the ESRC, the MacArthur Foundation, and the Andrew Mellon Foundation for financial support; and Larry Blume, Yannis Ioannides, Peter Neary, and Ben Zissimos for useful discussions. Participants at a Warwick globalization workshop, a MacArthur Foundation "Social Interactions and Economic Inequality" group meeting, and an LSE workshop have also provided helpful comments.

Cluster Emergence on a Global Continuum

by

Danny Quah

LSE Economics Department

October 1999

ABSTRACT

This paper models spatial economic development, making explicit both time and space. Locations are not just points—which would leave unanswered the question, What happens in between?—but instead a continuum. Equilibrium is a law of motion in spatial distribution dynamics, or a transition kernel in measures on geographical space. The paper provides a model of economic geography without transportation costs; it is used to study the evolution across Earth of financial, Internet, telecommunications, and computer activity.

Keywords: distribution dynamics, Fourier, Internet, location, space, time, Toeplitz, weightless economy

JEL Classification: C33, D30, 018, 033

Communications to: D. Quah, LSE, Houghton Street, London WC2A 2AE.

Tel: +44/0 207 955-7535, Email: `dquah@econ.lse.ac.uk`
(URL) `http://econ.lse.ac.uk/~dquah/`

Cluster Emergence on a Global Continuum
by
Danny Quah *
LSE Economics Department
October 1999

1 Pockets of poverty and clusters of excellence

That economic activity is not uniformly distributed across either space or time is the central issue of study in economic geography and economic growth. Indeed, one can go further. When the underlying units over which we measure economic activity are taken to be, in turn, people, firms, and then countries, one gets statements, respectively, about income distribution and inequality, industrial performance and structure, and patterns of cross-country convergence.

Distributions across space and time, however, allow an analysis not available across the discrete units in people, firms, and countries. That analysis is the *in-between*: An economic theory that restricts the behavior of locations A and B should also imply predictions on all locations between A and B. Put differently, one might wish to say more than simply that location A has more economic activity than location B: Why take the locations to be given in the first place? Does the analysis also explain why yet another location, C lying between A and B, does not display the same clustering of economic activity observed in either A or B? Even if empirically, we only observe A and B, it is useful to understand, Why not C?

This reasoning recalls earlier econometric reasoning for analyzing dynamic economic models in continuous time, then working out the

*I thank the British Academy, the ESRC, the MacArthur Foundation, and the Andrew Mellon Foundation for financial support; and Larry Blume, Yannis Ioannides, Peter Neary, and Ben Zissimos for useful discussions. Participants at a Warwick globalization workshop, a MacArthur Foundation "Social Interactions and Economic Inequality" group meeting, and an LSE workshop have also provided helpful comments.

implications for observed behavior measured only at discrete time intervals (Hansen and Sargent (1991), Sims (1971)). The motivation is, if anything, even stronger for space than for time. Spatial locations can develop and evolve in ways that are more intricate than with time points. Entrepreneurs and policy-makers routinely contemplate starting up production centers in previously economically-barren regions. The counterpart in time when one contemplates, say, round-the-clock economic activity occurs less often.

Yet another motivation is given by Krugman and Venables (1997): they seek an international trade theory without the artifact of countries as distinct trading entities. Just as one can move away from the idea of analyzing space as isolated points, one can be interested also in investigating the patterns of international specialization that arise from an initially seamless, homogeneous world.

This paper shares some of the tools and ideas in Krugman and Venables (1997) and Turing (1952). As do those analyses, this paper attempts to model how distinct patterns of clustering can arise, dynamically, from an otherwise uniform, initially seamless underlying space. In the dynamics, such *emergence*—the appearance of patterns where initially there were none—can be related to ideas about the economics of superstars (MacDonald (1988), Rosen (1981)) and to ideas about self-organization in evolutionary biology (Kauffman (1993), Turing (1952)). Unlike Krugman and Venables (1997), however, this paper studies global clustering patterns that arise, not from transportation costs, but instead in a seamless world where timeliness is important rather than geographical distance. The paper thus provides a motivation for economic geography without transportation costs.

The basic ideas in the model are twofold. First, in a world where transportation costs are trivially low, location can still be important through, on a global scale, differing timeliness in production and consumption. Timezones matter even if physical distances do not. Second, forces for agglomeration can arise not just from increasing returns—as in Krugman and Venables (1997)—but from ordinary complementarity, even in regular constant returns to scale technologies. In the model below, dynamically optimizing producers having rational expectations balance adjustment costs against complementarities: That tension suffices to produce, in transition to steady state, emergent clusters.

The conjecture is that in a stochastic stationary equilibrium, those clusters will remain a prominent feature of the equilibrium probability law of motion and hence will be observable. This will be even when the *average* (or unconditional expectations of) behavior—corresponding to the steady state in the deterministic model below—shows no such clustering. The stochastic version is not worked out in this paper, however, as we will see the deterministic model already presents quite intricate technical difficulties.

A second concern in this paper (the first being the problem of the "in-between") is to provide an appropriate framework for the empirical analysis of spatial dynamics, tuned in ways focussing on economic questions. Thus, parts of the paper are devoted to clarifying why a number of conventional empirical designs can be misleading or uninformative for the issues of interest.

Third, the model in the paper is used to organize facts about the dynamic pattern of financial, Internet, telecommunications, and computing activity across the globe. As elements of the so-called weightless economy, the resulting economic commodities are disrespectful of geography in the traditional sense.[1] They do not incur transportation costs. Why should distinct clusters emerge? The model below provides an explanation for this, based on timeliness in production.

The remainder of the paper is organized as follows. Section 2 provides examples—some stylized hypothetical ones; others real-world—illustrating the forces of interest in this study. These examples highlight issues that are relevant in general. Section 3, by contrast, turns to a base analytical model that is specialized: It describes a world where transportation costs are irrelevant but timeliness matters, and characterizes the resulting dynamic-distributional equilibrium. Section 4 applies the model to study empirical distribution dynamics of Internet and telecommunications activity across the world. Section 5

[1] See Cairncross (1997), Coyle (1997), and Quah (1996a, 1997b, 1999).

concludes.

2 Examples

This section provides a number of examples—some stylized, some real-world—to motivate the analysis below. (It can be omitted by those who feel the motivation in the Introduction is already sufficient.) The examples illustrate that the theoretical concerns expressed above are general, and moreover have direct implications for empirical analysis. They suggest why regression-based methods—for instance, panel-data analyses that construct and investigate the conditional representative—need not be usefully revealing for studying spatial and geographical economies, even if, spuriously, those methods exploit information from both the cross section and times series dimensions in a dataset. Thus, this section can also be read as one listing pitfalls in empirical analysis of spatial behavior. Arbia (1997) and Arbia and Espa (1999) consider related issues; some of the discussion below overlaps with theirs.

2.1 Emerging twin peaks; spatial conditioning

The first example shows how emergence can be observed in the distribution of incomes across countries, and why spatial dependence matters.

Fig. 1 reproduces a picture from Quah (1997a). It depicts the evolution through time of the cross-section income distribution across over 120 countries. The earlier distribution, at time t, is drawn here to be unimodal, but is intended only to be nondescript. By contrast, the distribution at time $t + s$ shows the emergence of two modes, one a cluster of rich, the other a cluster of poor. A pattern has arisen where previously none was observable.

Fig. 2 shows stochastic kernels for the cross-country income distributions over time (panel a.) and conditioning on spatial neighbors (panel b.).[2] Applying the stochastic kernel in panel a. repeatedly

[2] Quah (1997a) describes the construction of these stochastic ker-

to arbitrary initial income distributions produces an emerging twin-peakedness in the sequence of image distributions. Panel b. shows that conditioning on space "explains" that emerging twin-peakedness feature in the cross-country distribution dynamics.

Fig. 3 describes the mechanics of the spatial conditioning performed in Fig. 2.b. Countries or regions lie in a physical geography that is the two-dimensional plane; their per capita incomes form a surface above that geography. Conditioning on physical neighbors means taking into account that region [or country] j is contiguous with certain regions, but not with others. Such information is used in Fig. 2.b, but not in routine econometric analysis (with panel data methods, say).

2.2 Individual effects in panel data regression

The data in Figs. 1–3 display simultaneously extensive cross-section and time-series variation. Confronted with this, researchers have traditionally turned to panel-data methods for conducting empirical analysis. Panel-data techniques certainly do apply to data that have extensive cross-section and time-series variation. However, this does not immediately imply that they are always revealing and appropriate for such data.

Fig. 4 shows how panel-data techniques can, in general, be misleading and uninformative. The plot shows two dynamic cross sections evolving through time. One converges to a degenerate point mass; the other, to two distinct clusters. But because a panel-data regression constructs a conditional representative or average of the cross section, all the dynamics that the researcher can detect in either case is that the growth rate is negatively related to initial conditions.[3] To emphasize the point, notice that if the researcher, as is

nels.

[3] This example is one where Galton's fallacy (Quah, 1993) does not, in fact, apply: the negative correlation correctly indicates a diminishing of the cross-section dispersion. That, however, is not what matters. Nor is the shortcoming repaired by using, say, estimated

standard in panel-data analysis, conditions out individual effects—either with fixed- or random-effects models or using a variety of other possibilities—then the two dynamic cross-sections in Fig. 4 are not even distinguishable, despite their implications for clustering being so different.

The point made here about panel-data analysis, of course, goes beyond just spatial or geographic issues. It would apply even in their absence. But the potential for misinterpretation is compounded when one is interested in clusterings across geography.

It is, clearly, not just regression-based or averaging methods that can mislead. When studying spatial clusterings empirically, a researcher is naturally led to calculate measures of activity within distinct regions and then to describe concentration by Gini coefficients, say, or some other functional of the distribution across regions.

The next two examples show how this can distort empirical regularities.

2.3 Arbitrary discrete boundaries. MAUP

The first of these derives from the possibility that when decisions are made on where to locate economic activity, political or statistical accounting regional boundaries might be irrelevant.

Those data are then available on the resulting discrete accounting regional units. Fig. 5 summarizes the resulting difficulty for empirical analysis. Suppose that geography is homogeneous and that exogenous government policies are identical across regions. Turning to Fig. 5, in either panel, suppose that four plants locate in the same physical locations at the four vertices of a square. In the first panel, the regional boundaries happen to be drawn resulting in each plant recorded to belong to a different region. By contrast, in the second panel, all four plants are recorded to be in the same region, with all other regions seen to have no economic activity. The distribution of activity is uniform in the first case, but highly concentrated and unequal—clustered—in the second.

standard errors that correct for spatial or cross-section dependence.

This pattern of observations might be interpreted to call for an explanation: Why, when economic circumstances are otherwise homogeneous, does clustering occur in one case but not in another? The true reason is not deep; it is just a unintended consequence of an arbitrary discretization in regional units. Arbia (1997) calls this the *modifiable unit areal problem* or *MAUP*.

I have described the difficulty in this example as one having nothing to do with regression-based analysis. If, however, MAUP holds in a dataset, any regression study analyzing the differing regional concentrations in those data will also be, immediately, problematic.[4]

2.4 Aspatial regional analysis. Wide-sense dependence

The second example, abstracting from regression-based analysis, illustrates what happens when an analysis does not take space explicitly into consideration when analyzing regional empirics.

Suppose a researcher calculates the empirical distribution of economic activity across regions, and assesses the inequality characteristics in that distribution. The underlying idea (e.g., Krugman, 1991)) is a natural one: the more concentrated the distribution—the higher the inequality—the more evidence for spatial agglomeration.

Fig. 6 shows how two different patterns of Y regional economic activity—the lower and upper panels on the left—can be aliased into the same cross-region income distribution. In the upper left panel, rich and poor regions alternate in space. The right side of the tracks intermingle seamlessly with the wrong side. By contrast, in the lower left panel, the origin on the geography axis forms a firm dividing line. All the poor are to the left of it; all the rich, to the right. However, despite the markedly different economic behavior across the two scenarios, both induce the identical income distribution, on the right, one-half rich and one-half poor.[5]

[4] Ellison and Glaeser (1997) have studied geographic concentration in US manufacturing.

[5] Some empirical researchers *do* attempt to take into account such spatial properties. Examples include Hanson (1997), Moreno and

Implicit in Fig. 6 is a message more general than simply that income distributions hide important information. Crucial elements of economic geography are hidden unless one takes into account explicitly how one cluster of economic activity is located relative to another. The point is not just that clusters occur; where they occur relative to others matters as well. A researcher can well analyze the first question taking potential locations as given—the question then is how much richer is one location rather than another. Analysis of the latter, however, cannot proceed if one takes locations to be prior to the analysis. Instead, what is needed as primitive in the modelling is a seamless continuous geography, without locations on that geography already identified as potential cluster points.

3 Economic geography on a continuum without transportation costs

The model developed here is one of an economic geography where the underlying space of putative locations is a featureless continuum. No point in that space is, therefore, identified a priori as a potential cluster of economic activity.

The mathematical modelling tools I use build on analyses in Krugman and Venables (1997) and Turing (1952). Krugman and Venables (1997) studied how concentrations of manufacturing and agriculture production would evolve across different locations. Turing studied how simply physical laws could account for the emergence of pronounced features from an otherwise homogeneous embryo.

While the structure is general and applies to a wide range of possibilities, for concreteness I study the location of economic activity on the three-dimensional globe given in Fig. 7. Again, for concreteness, the model should be viewed to describe activity in the form of news content provision, financial trading, or Internet-based production.

Trehan (1997), and Quah (1996b).

3.1 Forty-eight hour workdays

The canonical activity I model might be described as follows.

Suppose a journalist works for MSNBC and is located in New Jersey, on the east coast of the US. What she produces in fresh news content can be used—for production or consumption—24 hours a day by economic agents around her.

However, at the end of her 24-hour day, even as agents around her are shutting down, economic agents in US timezones to the west of New Jersey continue to remain active. These people, logging on to the MSNBC website, can still use fresh news content. Thus, this journalist might be described to have the possibility of working 27-hour days. Indeed, on a seamless globe, every economic agent can work 48 hours a day on each fixed calendar date.

What matters in such a world is not geographical distance but timezone location. In Fig. 7 all locations on a fixed longitude are equivalent, from the perspective just described above.

3.2 Isomorphic to the unit circle

Abstracting from singularities at the North and South poles, the three-dimensional globe in Fig. 7 is, for economic purposes, isomorphic to the equator. The two-dimensional location problem on the globe thus reduces to a one-dimensional one on a circle.[6]

Call the resulting continuum geography $\mathbb{G}$, and denote representative points in it by $z, z', z'' \in \mathbb{G}$ (Fig. 8). Without loss, identify $\mathbb{G}$ with the unit circle in the complex plane. Then, viewing $\mathbb{G}$ from the North pole, each $z \in \mathbb{G}$ corresponds to a unique angle $\omega \in (-\pi, \pi]$, measured in radians, such that

$$z = e^{i\omega} \in \mathbb{G}.$$

[6] The equilibrium below will be described by Fourier analysis on the unit circle. This suggests that the more complete two-dimensional location model can be analogously treated using higher-dimensional Fourier analysis, but the problem remains to be rigorously analyzed. A related special case is studied in (Turing, 1952, Sec. 12).

From Fig. 8, positive ω's indicate an easterly direction, so that higher positive ω's describe locations earlier in time. Denote the angle corresponding to any $z \in \mathbb{G}$ by $\omega(z) \in (-\pi, \pi]$ such that $\omega(z) = i^{-1} \log(z)$ (unique mod π).

3.3 Timely joint production

Assume that at location z output $Y(z)$ depends on factor inputs located not just at z but at z' all over the globe. Output Y is homogeneous, and normalized to have price 1.

To continue the news production example, assume that in this joint production, the efforts of factor inputs that occur earlier in time, provided those efforts are fresh, are more useful in making $Y(z)$ than those produced later.

Describe this by the following:

Definition *A* ***timeliness*** *function is a mapping*

$$\mathsf{T} : \mathbb{G} \times \mathbb{G} \to [0, 1],$$

satisfying three conditions:

1. *For each $z \in \mathbb{G}$, the maximum in $\mathsf{T}(\cdot, z)$ is attained at z, i.e.,*

   $$\mathsf{T}(z, z) = 1 \geq \mathsf{T}(z', z) \; \forall z' \in \mathbb{G}.$$

   Local production is always the most timely.

2. *For all $z, z' \in \mathbb{G}$, timeliness $\mathsf{T}(z', z)$ depends only $\omega(z) - \omega(z')$ mod π, i.e.,*

   $$\forall \omega \qquad \mathsf{T}(z', z) = \mathsf{T}(e^{\omega(z')i}, e^{\omega(z)i}) = \mathsf{T}(e^{(\omega(z')+\omega)i}, e^{(\omega(z)+\omega)i}).$$

   This can be described as timeliness being invariant under equatorial rotation. Alternatively, we can say that timeliness is radially homogeneous: Timeliness depends only on timezone separation, not on where production currently occurs.

3. *For all $z \in \mathbb{G}$, timeliness $\mathsf{T}(z', z)$ is monotone decreasing in $|\omega(z') - \omega(z)|$, and moreover*

$$\omega(z') - \omega(z) > 0 \text{ and } z'' \stackrel{\text{def}}{=} e^{(\omega(z)-(\omega(z')-\omega(z)))i} \implies \mathsf{T}(z', z) \geq \mathsf{T}(z'', z).$$

Timeliness never increases as one proceeds further away from where production is occurring; production earlier at a given timezone separation is always more timely than production later at an equal timezone separation.

Fig. 9 shows a representative T timeliness function.

Condition 3 specifies that timeliness is, in general, asymmetric, and therefore is unlike physical distance.

3.4 The distribution of economic activity

Production occurs with only one factor input, which I will call capital.

Capital is homogeneous. For simplicity, fix its total quantity at 1. Capital can, however, be distributed in different ways around the globe. Denote the spatial density of capital across geography by $f : \mathbb{G} \to \mathbb{R}_+$, with

$$\int_{\mathbb{G}} f(z) = 1, \qquad f(z) \geq 0. \tag{1}$$

Density f will evolve through time. Together with timeliness T, density f implies output at each time point as the following assumed CES form:

$$\forall z \in \mathbb{G} : \quad Y(z) = \left[\int_{\mathbb{G}} [\mathsf{T}(z', z) f(z')]^{\gamma} \, dz' \right]^{1/\gamma}, \qquad \gamma > 0. \tag{2}$$

(Thus, f denotes both the factor input and its density, but this usage is without ambiguity.)

The factor input is most productive for $Y(z)$ when it is located at z; however, as its timezone separation from z increases, its marginal productivity for output at z declines. Indeed, holding the spatial density f constant, the marginal productivity of factor inputs at z' will have a profile given exactly by the timeliness function $T(z', \cdot)$.

From (2) the distribution of output need not be any obvious transformation of that in the factor input. However, equation (2) readily implies that if f is uniform, i.e., is constant at $1/2\pi$, then the spatial distribution of output is similarly uniform:

$$\begin{aligned} Y(z) &= (2\pi)^{-1} \left[\int_{\mathbb{G}} \mathsf{T}(z', z)^{\gamma} \, dz' \right]^{1/\gamma} \\ &= (2\pi)^{-1} \|\mathsf{T}\|_{\gamma} \quad \text{constant independent of } z, \end{aligned}$$

where the γ-norm

$$\begin{aligned} \|\mathsf{T}\|_{\gamma} &\stackrel{\text{def}}{=} \left[\int_{\mathbb{G}} \mathsf{T}(z', 1)^{\gamma} \, dz' \right]^{1/\gamma} \\ &= \left[\int_{\mathbb{G}} \mathsf{T}(z', z)^{\gamma} \, dz' \right]^{1/\gamma} \quad \text{by radial homogeneity.} \end{aligned}$$

3.5 Individual maximization and dynamic adjustment

At location z the owner of the factor input behaves competitively, taking as given the factor price path $\{W_t(z) : t \in [0, \infty]\}$. He compares this with

$$\overline{W}_t \stackrel{\text{def}}{=} \int_{z' \in \mathbb{G}} W_t(z') \, dz', \tag{3}$$

so that the instant-t profit rate at z is

$$\left(W_t(z) - \overline{W}_t \right) f_t(z).$$

Equation (3) specifies W comparison with the *average* of W's over geography. In certain circumstances, it might be more natural to let $\overline{W}$ be instead the maximum $\sup_{z' \in \mathbb{G}} W_t(z')$ (when selling the factor input) or the minimum $\inf_{z' \in \mathbb{G}} W_t(z')$ (when buying). The assumption in (3) is that when the producer attempts to increase or decrease the current extant $f_t(z)$ there is no control over its source or destination. Thus, the relevant comparison payment is an average over all possible locations.

Assume the producer also faces adjustment costs

$$C(\dot{f}_t(z)) \geq 0, \quad \text{where } C' > 0 \text{ and } C'' \geq 0,$$

where $\dot{f}_t(z)$ denotes the time derivative of $f(z)$ at t. At any fixed location the quantity of factor input can be changed only at a cost; that marginal adjustment cost, moreover, is rising, the faster is the change $\dot{f}(z)$. For concreteness, assume adjustment costs are quadratic:

$$C(\dot{f}_t(z)) = \frac{1}{2}\zeta \times \dot{f}_t(z)^2, \qquad \zeta > 0. \tag{4}$$

Having assumed the integral $\int_{\mathbb{G}} f$ is fixed at 1, investments $\dot{f}(z)$ are really only reallocations across space, not changes in the aggregate stock of the factor input. It is easy, if notationally cumbersome, to assume aggregate growth does occur.

At each instant t the producer at z behaves competitively, taking the time paths of other producers' actions $\{f_s(z') : z' \neq z, \ s \geq t\}$ and factor compensations $\{W_s(z') : z' \in \mathbb{G}, \ s \geq t\}$, as given. The producer is forward-looking, has rational expectations, and selects the time-path of $f(z)$ to maximize the present discounted value of profits at a constant discount rate $\rho > 0$. Thus, at instant t the producer at z solves:

$$\forall \ t \geq 0 : \qquad \sup_{\{f_s(z):s\geq t\}} \int_{s\geq t} e^{-\rho s} \left[\left(W_s(z) - \overline{W}_s \right) f_s(z) - C(\dot{f}_s(z)) \right] \, ds, \tag{5}$$

subject to $\forall \ s$:

$$\overline{W}_s = \int_{z'\in\mathbb{G}} W_s(z') \, dz',$$
$$C(\dot{f}_s(z)) = \frac{1}{2}\zeta \dot{f}_s(z)^2,$$
$$\left\{ f_t(z'), \ \text{all } z' \in \mathbb{G} \right\},$$
$$\left\{ f_s(z'), \ z' \neq z, \ z' \in \mathbb{G}, \ s \geq t \right\}.$$

The first two constraints simply repeat (3) and (4). The remainder state that the producer at z takes as given the initial distribution at t and the actions of all the other producers.

Assume the factor input is rewarded for its contribution to global production. At z that is the marginal product of $f(z)$ in global production. To calculate this, first notice that from (2) the marginal product of $f(z)$ at z' is:

$$\frac{\partial Y(z')}{\partial f(z)} = \mathsf{T}(z, z')^{\gamma} \left(Y_t(z')/f_t(z)\right)^{1-\gamma}.$$

The return to f at location z is therefore:

$$\begin{aligned} W_t(z) &= \int_{\mathbb{G}} \frac{\partial Y(z')}{\partial f(z)}\, dz' \\ &= \int_{\mathbb{G}} \mathsf{T}(z, z')^{\gamma} \left(Y_t(z')/f_t(z)\right)^{1-\gamma}\, dz'. \quad (\text{since } \mathsf{T}(z, z) = 1). \end{aligned} \tag{6}$$

An *equilibrium* is a collection of time paths

$$\{\, f_t(z),\ z \in \mathbb{G},\ t \geq 0 \,\}$$

or, equivalently, a profile of densities on $\mathbb{G}$ satisfying producers' dynamic profit-maximization (5); at factor compensation paths given by (6) and (2); and with all markets clearing (1).

3.6 Equilibrium distribution dynamics

To characterize equilibrium it is convenient to introduce a few further definitions relating to distribution dynamics on a geographical space $\mathbb{G}$.

Let $\mathbb{P}_{\mathbb{G}}$ denote the collection of probability densities on $\mathbb{G}$. An equilibrium is a *process* on $\mathbb{P}_{\mathbb{G}}$, i.e., a mapping

$$X : [0, \infty] \to \mathbb{P}_{\mathbb{G}},$$

satisfying equations (1), (2), (5), and (6) . We will confine attention to processes that are differentiable in time. Let $\mathbb{Q}_{\mathbb{G}}$ denote the collection of time-derivatives of such processes. Because elements of $\mathbb{P}_{\mathbb{G}}$

obey the integral condition (1), by construction, the elements of $\mathbb{Q}_{\mathbb{G}}$ satisfy

$$q \in \mathbb{Q}_{\mathbb{G}} \implies \int_{z\in\mathbb{G}} q_t(z)\, dz = 0 \quad \forall t \geq 0.$$

An equilibrium X is *invariant* when

$$\dot{X}(t) = 0 \quad \forall\ t.$$

The value that X takes when invariant is an *invariant density*.

While, in general, equilibria might be sought in the space of all possible processes, for the current study it suffices to consider *Markov equilibria*, i.e., equilibria X for which there exist operators $T_{t,X(t)}$ indexed by t and X mapping $\mathbb{P}_{\mathbb{G}}$ to $\mathbb{Q}_{\mathbb{G}}$ such that:

$$\dot{X}(t) = T_{t,X(t)}X(t)$$

and with operator adjoints representable by a stochastic kernel so that for $f_t = X(t)$, we have:[7]

$$\dot{f}_t(e^{i\omega}) = \int_{-\pi}^{\pi} \left[M_{t,f_t}(e^{i\omega'}, e^{i\omega})\right] f_t(e^{i\omega'})\, d\omega' \quad \forall\ \omega. \tag{7}$$

Despite its integral form, the process described by (7) need not be "linear", as the operator T and thus the stochastic kernel M depend on both time t and state f_t.

Inspecting the equilibrium (1), (2), (5), and (6) , we see that Markov equilibria will arise when the optimization program (5) is recursive. This, in turn, obtains when in equilibrium all agents employ *Markov strategies*, selecting $f_t(z)$ as a function only of $f_t(z')$, all $z' \neq z$.

To see that a Markov equilibrium is possible, notice that the dynamic optimization problem (5) has first-order condition

$$W_t(z) - \overline{W}_t + \zeta\frac{d\dot{f}_t(z)}{dt} - \rho\zeta\dot{f}_t(z) = 0.$$

[7] Such a stochastic kernel construction is given in Quah (1997a). Rigorous descriptions are in, e.g., Futia (1982), and Stokey and Lucas (1989).

Solving stable roots backwards and unstable roots forwards (e.g., Sargent, 1987). gives the forward-looking adjustment rule

$$\dot{f}_t(z) = \zeta^{-1} \int_0^\infty e^{-s\rho}[W_{t+s}(z) - \overline{W}_{t+s}]\, ds. \tag{8}$$

The producer at z increases $f(z)$ if current-location returns are, dynamically, better than elsewhere. That response is stronger, the larger is the appropriate present discounted value of this location premium. On the other hand, the larger are adjustment costs ζ, the more sluggish is the response.

By the rational expectations assumption, the producer at z realizes that W and $\overline{W}$ are determined by (6) and (3) respectively. Thus, if producers at all $z' \neq z$ employ Markov strategies, then the forward-looking adjustment process in (8) is also Markov.

Equation (8) contains within it forces for agglomeration and opposing ones for stagnation or sluggishness.[8] Recall the concentration of T about $z' = z$ from Fig. 9. When $1/\gamma$ exceeds 1, the marginal product (6) is higher whenever factor inputs are concentrated in the immediate neighborhood. Other things equal, then, there would be a force for any initial spatial agglomeration to grow, feeding on itself. Opposing that, however, is the expense of moving factor input in from elsewhere, due to adjustment costs ζ. The larger is ζ, the smaller is $\dot{f}$, for any path of returns W. This tension between the shape of T and the magnitude of γ will determine equilibrium simultaneously across space and time.

Ignoring initial conditions, an invariant uniform equilibrium (and thus a trivial Markov equilibrium) always exists. If

$$f_t(z) = \frac{1}{2\pi} \quad \forall\ z \in \mathbb{G},\ \text{all } t \in [0, \infty],$$

[8] Such tendencies are well known from models with increasing-returns technologies and transportation costs as in, e.g., Krugman and Venables (1997). While formally similar, the mechanism here is, obviously, quite different in its economic description.

then, as before,

$$Y_t(z) = \frac{1}{2\pi}\|\mathsf{T}\|_\gamma$$

so that

$$W_t(z) = \|\mathsf{T}\|_\gamma^{1-\gamma} \quad \forall\ z \in \mathbb{G},\ \text{all } t \in [0, \infty].$$

This implies $\dot{f}_t(z) = 0$, thereby supporting the uniform distribution to be invariant.

Combining the discussion above, when agents use Markov strategies, the adjustment rule (8) gives the stochastic kernel representation:

$$\begin{aligned} \dot{f}_t(z) &= \zeta^{-1} \int_0^\infty e^{-s\rho}[W_{t+s}(z) - \overline{W}_{t+s}]\, ds \\ &= \int_{z' \in \mathbb{G}} M_{t,f_t}(z', z) f_t(z') dz'. \end{aligned} \tag{9}$$

Let $\overline{f}$ denote an invariant density in (9), and consider a first-order Taylor's series approximation in f_t around $\overline{f}$:

$$\dot{f}_t(z) = \int_{z' \in \mathbb{G}} \theta_M(z, z') \left(f_t(z') - \overline{f}(z')\right)\, dz' - \lambda_M \times \left(f_t(z) - \overline{f}(z)\right), \tag{10}$$

where $\theta_M - \lambda_M I$ is the Frechet derivative of $(M_{t,f_t} f_t)$ with respect to the density f_t; I is the identity operator; and λ_M is positive, real, and increasing in ζ.

Radial homogenity in T and thus in the equilibrium (1), (2), (5), and (6) implies that for all stochastic kernels M in (10), the Frechet derivative θ_M is Toeplitz:[9]

$$\theta_M(e^{i\omega}, e^{i\omega'}) = \theta_M(e^{(\omega+\omega'')i}, e^{(\omega'+\omega'')i}) \mod 2\pi \quad \forall\ M \text{ and } \omega, \omega', \omega''. \tag{11}$$

[9] A key reference for Toeplitz operators is Grenander and Szegö (1958). Useful expositions of the Fourier tools I use below are Sargent (1987) and Titchmarsh (1962).

This Toeplitz property allows explicit characterization of the dynamic equilibria.

Theorem *The matrix form θ_M has as its right eigenfunctions the complete orthonormal set of complex exponentials*

$$\left\{(2\pi)^{-1/2}e^{i\omega j} : \omega \in (-\pi,\pi],\ j = -\infty,\ldots,+\infty\right\} \tag{12}$$

independent of M. Its spectrum is the (discrete) Fourier transform of any one of the sections $\theta_M(z,\cdot)$, independent of z but varying with M. Finally, the family (12) also comprises the left eigenfunctions of θ_M, so that if

$$\phi_j(\omega) \stackrel{\text{def}}{=} (2\pi)^{-1/2}e^{i\omega j}$$

is a right eigenfunction at eigenvalue s_j, then ϕ_{-j} is a left eigenfunction at the same eigenvalue.

Proof *Use (11) first to calculate*

$$\begin{aligned}\int_{-\pi}^{+\pi}\theta_M(z,e^{i\omega'})e^{i\omega' j}d\omega' &= \int_{-\pi}^{+\pi}\theta_M(1,e^{(\omega'-\omega)i})e^{i\omega' j}d\omega' \\ &= e^{i\omega j}\int_{-\pi}^{+\pi}\theta_M(1,e^{i\omega'})e^{i\omega' j}d\omega'. \end{aligned}\tag{13}$$

Thus, notice that for any integer j,

$$\phi_j(\omega) \stackrel{\text{def}}{=} (2\pi)^{-1/2}e^{i\omega j} \qquad \omega \in (-\pi,\pi]$$

is an eigenfunction, for

$$\begin{aligned}(\theta_M\phi_j)(\omega) &= (2\pi)^{-1/2}\int_{-\pi}^{+\pi}\theta_M(e^{i\omega},e^{i\omega'})e^{i\omega' j}d\omega' \\ &= \left(\int_{-\pi}^{+\pi}\theta_M(1,e^{i\omega'})e^{i\omega' j}d\omega'\right)(2\pi)^{-1/2}e^{i\omega j} \qquad \text{by (13)} \\ &= s_j \times \phi_j(\omega),\end{aligned}$$

where the corresponding eigenvalue has been defined as:

$$s_j \stackrel{\text{def}}{=} \left(\int_{-\pi}^{+\pi} \theta_M(1, e^{i\omega'}) e^{i\omega' j} d\omega' \right). \tag{14}$$

The spectrum (the entire set of eigenvalues)

$$\{ s_j : j = 0, \ \pm 1, \ \pm 2, \dots \}$$

is, from (14)*, the Fourier transform* $\int_{-\pi}^{+\pi} \theta_f(1, e^{i\omega}) e^{i\omega j} d\omega$ *(all integer* j*). By radial homogeneity, each Fourier coefficient* s_j *is the same as* $\int_{-\pi}^{+\pi} \theta_f(z', e^{i\omega}) e^{i\omega j} d\omega$*, for all* z'*. Finally, calculate*

$$\begin{aligned}
(\phi_{-j}\theta_M)(\omega') &= \int_{-\pi}^{+\pi} \phi_{-j}(\omega)\theta_M(e^{i\omega}, e^{i\omega'})\, d\omega \\
&= (2\pi)^{-1/2} \int_{-\pi}^{+\pi} e^{-i\omega j}\theta_M(1, e^{(\omega'-\omega)i})\, d\omega \\
&= (2\pi)^{-1/2} \int_{-\pi}^{+\pi} e^{(\omega''-\omega)ij}\theta_M(1, e^{\omega'' i})\, d\omega'' \\
&= (2\pi)^{-1/2} e^{i\omega' j} \int_{-\pi}^{+\pi} \theta_M(1, e^{\omega'' i}) e^{\omega'' ij}\, d\omega'' \\
&= \phi_{-j}(\omega') s_j
\end{aligned}$$

(where in the second equation, the change in sign of the differential element has been absorbed into the change in the integral limits).

Q.E.D.

The distribution dynamics in (10) can then be analyzed by taking Fourier transforms over $(-\pi, \pi]$ on both sides, so that for all integer

j, we have:

$$
\begin{aligned}
&(2\pi)^{-1/2} \int_{-\pi}^{+\pi} \dot{f}_t(e^{i\omega}) e^{i\omega j}\, d\omega \\
&\quad = (2\pi)^{-1/2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \theta_M(e^{i\omega}, e^{i\omega'}) e^{i\omega j} \left(f_t(e^{i\omega'}) - \overline{f}_t(e^{i\omega'}) \right) d\omega' \, d\omega \\
&\qquad - (2\pi)^{-1/2} \lambda_M \times \int_{-\pi}^{+\pi} \left(f_t(e^{i\omega}) - \overline{f}_t(e^{i\omega}) \right) e^{i\omega j}\, d\omega \\
&\quad = (s_{-j} - \lambda_M) \times (2\pi)^{-1/2} \int_{-\pi}^{+\pi} \left(f_t(e^{i\omega}) - \overline{f}_t(e^{i\omega}) \right) e^{i\omega j}\, d\omega
\end{aligned}
$$

Denoting the Fourier transforms (across space, not time):

$$
\forall \text{ integer } j : \qquad
\begin{aligned}
\tilde{\dot{f}}_{t,j} &= (2\pi)^{-1/2} \int_{-\pi}^{+\pi} \dot{f}_t(e^{i\omega}) e^{i\omega j}\, d\omega, \\
\tilde{f}_{t,j} &= (2\pi)^{-1/2} \int_{-\pi}^{+\pi} f_t(e^{i\omega}) e^{i\omega j}\, d\omega, \\
\tilde{\overline{f}}_j &= (2\pi)^{-1/2} \int_{-\pi}^{+\pi} \overline{f}(e^{i\omega}) e^{i\omega j}\, d\omega,
\end{aligned}
$$

the dynamic equation above becomes simply

$$
\tilde{\dot{f}}_{t,j} = (s_{-j} - \lambda_M) \left(\tilde{f}_{t,j} - \tilde{\overline{f}}_j \right), \qquad j = 0, \ \pm 1, \ \pm 2, \ldots . \tag{15}
$$

Equation (9), through its first-order Taylor's series expansion (10), has become in (15) a (doubly) infinite collection of equations, indexed by the set of integers.

The orthogonalization achieved by Fourier transforms in (15) allows convenient characterization of the distribution dynamics in (9). As time unfolds and the distribution f evolves, the spectrum

$$
\{ s_j : j = 0, \ \pm 1, \ \pm 2, \ldots \}
$$

changes. Representation (15), however, does not. Thus, whether f in (9) is dynamically stable around a putative steady state $\overline{f}$ is easily verified. If the spectrum, when perturbed downwards by λ_M, always has real part positive, then the steady state $\overline{f}$ is unstable in all directions. If, on the other hand, the perturbed spectrum is always negative, then the distribution dynamics are locally stable in all directions.

Finally, if, as in Fig. 10, some parts of the spectrum have real part dominated by λ_M but other parts otherwise, then the equilibrium is dynamically stable for deviations in f from $\overline{f}$ only in specific directions. This corresponds to, say, the saddlepoint stable path in Cass-Koopmans-type growth models. In particular, from (15), the deviation $f - \overline{f}$ must have a Fourier transform that vanishes for all indexes j where the real part $\Re s_{-j}$ exceeds λ_M. All such distributions f—lying in a restricted subspace of $\mathbb{P}_{\mathbb{G}}$—converge to $\overline{f}$ at a rate equal to $\sup_{j:\Re s_{-j}<\lambda_M} |\Re s_{-j} - \lambda_M|$.

Restricting attention to the uniform invariant steady state for $\overline{f}$, the convergent subspace can be characterized further: Except for degenerate cases, that convergent subspace comprises distributions whose densities are (periodic) waveforms on $\mathbb{G}$. To see this, notice that on the convergent subspace,

$$\Re s_{-j} > \lambda_M \implies \tilde{f}_{t,j} = \tilde{\overline{f}}_j.$$

Call $\mathbb{J}$ the set of j's above, and denote the complement by $\mathbb{J}'$. Then, taking the inverse Fourier transform of $\tilde{f}_{t,j} - \tilde{\overline{f}}_j$ gives

$$f_t(e^{i\omega}) - \overline{f}(e^{i\omega}) = (2\pi)^{-1/2} \sum_{j\in\mathbb{J}'} \tilde{q}_{t,j} e^{-i\omega j}, \tag{16}$$

where $\tilde{q}$ is the Fourier transform of some q with $\int_{-\pi}^{+\pi} q(e^{i\omega})\, d\omega = 0$. In other words, f differs from the uniform invariant $\overline{f}$ by a linear combination of sinusoidal components. Since $\mathbb{J}'$ is in general restricted—cannot be the entire set of integers—the linear combination on the right in (16) cannot be constant in ω. Equation (16) thus produces (periodic) waveforms—linear combinations of sinusoidal fluctuations—on $\mathbb{G}$, as depicted in Fig. 11.

It is important to realize that the cycles in spatial geography in (16) are related to but not inherited directly from cyclicality that might have been built into the model. Indeed, the closest that the underlying structure comes to waveforms is the timeliness specification T or Fig. 9. But there the map T is single-peaked, and certainly exhibits no obvious waves in $\omega' - \omega$. Instead, the cycles in space in (16) are due to interaction between timeliness T, costly adjustment ζ, and convergent distribution dynamics.

Finally, these spatial cycles (16) bear what Sargent calls the "hallmark cross-equation restrictions of rational expectations models": Cycles in space arise from the same economic parameters as cycles in time. Dynamically optimizing producers, with rational expectations, trade off adjustment costs against complementarity gains. This tension exists both in time and across space, and suffices to generate, in transition to deterministic steady state, emergent clusters. Fluctuations in time and space are related, therefore, through the same economic parameters. However, while it is the spectrum (or eigenvalues) of the Markov transition operator that determines convergence in time, it is the set of eigenfunctions that determines the cyclicality in space.

A conjecture not worked out here is that in a fully stochastic, stationary steady-state equilibrium, those clusters will remain prominent in the equilibrium probability law of motion—and hence will be observable—even if *average* (or unconditional expectations of) behavior, in the form of the uniform invariant steady state, displays no such clustering.

4 Empirical example: Spatial distribution dynamics on Earth

This section studies the spatial distribution of financial and Internet activity across Earth.

This is interesting for at least two reasons: First, as elements of the weightless economy, these activities bear no transportation costs and are thus disrespectful of geography in the traditional sense. Do they nevertheless show clustering in more than one distinct locations, and if so why?

Second, as suggested in Quah (1997b, 1999), they are activities that show high growth and potentially might account for significant fractions of total GDP. Understanding their dynamics is therefore a question of some macroeconomic interest.

One economic mechanism that has been discussed, informally, is where, say, Bangalore—now an important software center—is approximately half a day from Silicon Valley. Thus, when one workshift ends, the other begins. Similarly, financial centers such as Tokyo, London, or New York are sometimes rationalized.

These, however, are economic stories that are, in essence, about economic agents having to sleep. The model in Section 3, while it does produce such periodicities in space across the globe, is a quite different mechanism. The spatial cycles in the model's equilibrium arise from trading off timely agglomeration and adjustment costs, not from agents sleeping at different segments of global [not local] time. Indeed, the latter would manifest, in the language of Section 3, as a T function having waveforms built in. This would then strengthen the cyclicality in (16); and it would make the emergence expressed in that equation less surprising.

[[TO BE DONE STILL]]

5 Conclusion

This paper has studied spatial distribution dynamics—dynamic economic geography—in a model where transportation costs are irrelevant.

In the model, the underlying physical geography is taken to be initially a seamless, featureless continuum. The paper described a number of reasons for starting thus, rather than with, say, a fixed finite set of points (and nowhere else) where economic activity can potentially occur.

The principal of these reasons is the following. In studying economic growth, the researcher asks for the circumstances under which country A might become as developed and wealthy as another country

B. Similarly, in studying firms, the researcher might seek to understand conditions under which firm A performs better than firm B. In parallel, a researcher in economic geography can study when some region A is more prosperous than another region B.

However, geography affords an analysis that is unavailable when one studies the growth of countries. This is the analysis that asks when a third location C might emerge to be economically active when previously it was not. A model that takes the set of locations to be, a priori, fixed at only A and B cannot address such a question.

The solution adopted in this paper is to consider the initial space of locations to be unrestricted: all points on the globe (in the analysis here) are possibilities for economic activity.

There are other, empirical reasons for preferring an analysis that does not restrict locations to be simply a discrete set of points. Section 2 discussed this in some detail; Arbia (1997) and Arbia and Espa (1999) provide yet another analysis. I do not repeat the lessons here.

Analyzing distribution dynamics on a continuum presents, in general, substantial technical difficulties. The approach taken here exploited Fourier and Toeplitz analyses, so that the final result, expressed compactly as (15) in Section 3, is quite easily manipulated.

While motivations and economic mechanisms differ, many of the ideas I use here appear also in Krugman and Venables (1997) and Turing (1952). The former studied manufacturing and agriculture production across a discrete set of locations; the latter studied emergence properties in biology. Turing (1952) mentions a continuum discussion, together with a reliance on Fourier methods—much was, however, left implicit. In a number of ways, the technical discussion in this paper did no more than connect the dots in these two excellent earlier articles.

The empirical application in Section 4 to elements of the so-called weightless economy showed . . .

[[**TO DO STILL**]]

References

Arbia, Giuseppe (1997) "Modelling the geographical concentration of economic activities," Discussion Paper 2, DMQTE, University G. D'Annunzio, Pescara, Italy, January

Arbia, Giuseppe, and Giuseppe Espa (1999) "The empirical analysis of spatial concentration," Working Paper, DMQTE, University G. D'Annunzio, Pescara, Italy, February

Cairncross, Frances (1997) *The Death of Distance* (London: Orion Business Books)

Coyle, Diane (1997) *The Weightless World* (London: Capstone)

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

Futia, Carl (1982) "Invariant distributions and the limiting behavior of Markovian economic models," *Econometrica* 50(1), 377–408, January

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

Hansen, Lars Peter, and Thomas J. Sargent (1991) "Two difficulties in interpreting vector autoregressions," In *Rational Expectations Econometrics,* ed. Lars Peter Hansen and Thomas J. Sargent (Boulder CO: Westview Press) chapter 4, pp. 77–119

Hanson, Gordon H. (1997) "Market access and economic geography," Working Paper, Department of Economics, University of Texas, October

Kauffman, Stuart A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press)

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

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

MacDonald, Glenn M. (1988) "The economics of rising stars," *American Economic Review* 78(1), 155–166, March

Moreno, Ramon, and Bharat Trehan (1997) "Location and the growth of nations," *Journal of Economic Growth* 2(4), 399–418, December

Quah, Danny (1993) "Galton's fallacy and tests of the convergence hypothesis," *The Scandinavian Journal of Economics* 95(4), 427–443, December

_ (1996a) "The invisible hand and the weightless economy," Occasional Paper 12, Centre for Economic Performance, LSE, London, May

_ (1996b) "Regional convergence clusters across Europe," *European Economic Review* 40(3–5), 951–958, April

_ (1997a) "Empirics for growth and distribution: Polarization, stratification, and convergence clubs," *Journal of Economic Growth* 2(1), 27–59, March

_ (1997b) "Increasingly weightless economies," *Bank of England Quarterly Bulletin* 37(1), 49–56, February

_ (1999) "The weightless economy in economic development," Working Paper 155, United Nations University World Institute for Development Economics Research, Helsinki, January

Rosen, Sherwin (1981) "The economics of superstars," *American Economic Review* 71(5), 845–858, December

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

Stokey, Nancy L., and Robert E. Lucas (1989) *Recursive Methods in Economic Dynamics* (Cambridge MA: Harvard University Press)

Titchmarsh, E. C. (1962) *Introduction to the Theory of Fourier Integrals,* 3rd ed. (Oxford: Oxford University Press)

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

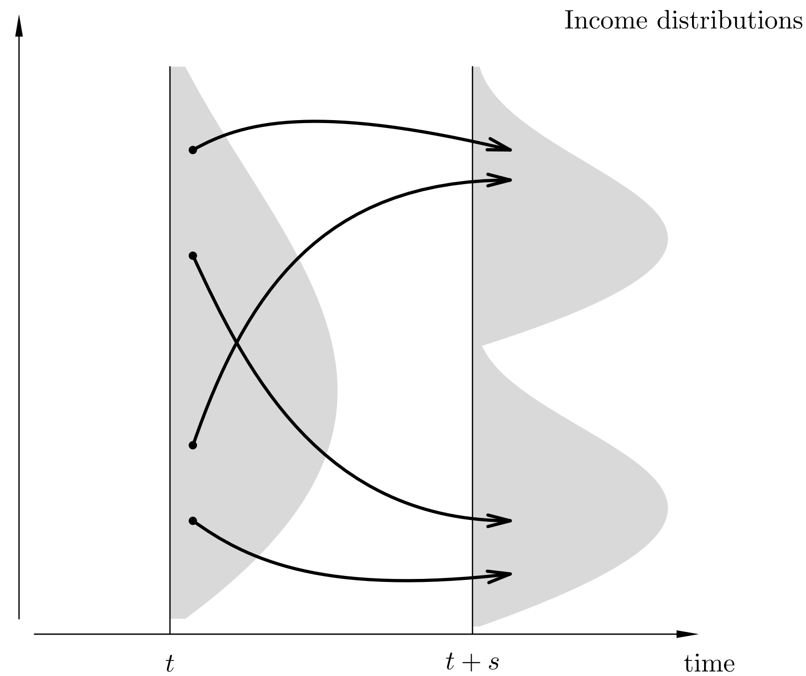
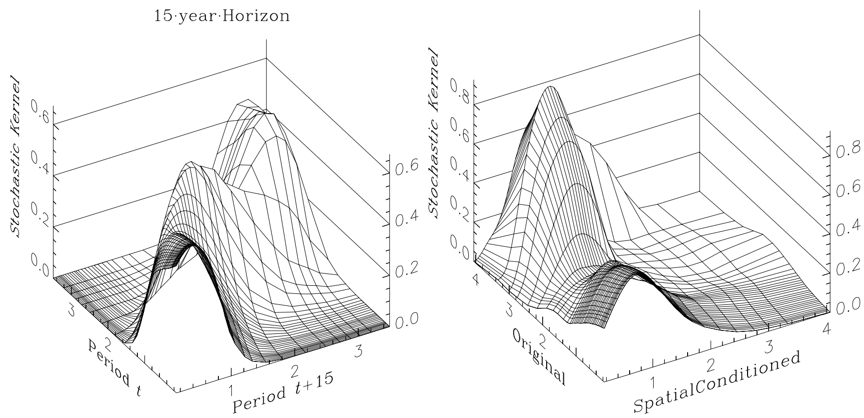


Fig. 1: Emerging twin peaks The earlier cross-section income distribution is nondescript. Over time, two modes appear and (potentially) persist. Some elements of the cross section succeed; some fail; yet others either languish or preserve their relative success.



a. 15-year transitions b. Space

Fig. 2: Spatial conditioning Panel a. shows the stochastic kernel for transitions across 15 years. Applied repeatedly to arbitrary initial income distributions, this kernel produces an emerging twinpeakedness. Panel b. shows the stochastic kernel conditioning on a country's neighbors in space. Conditioning on space "explains" the emerging twinpeakedness of income distributions across countries.

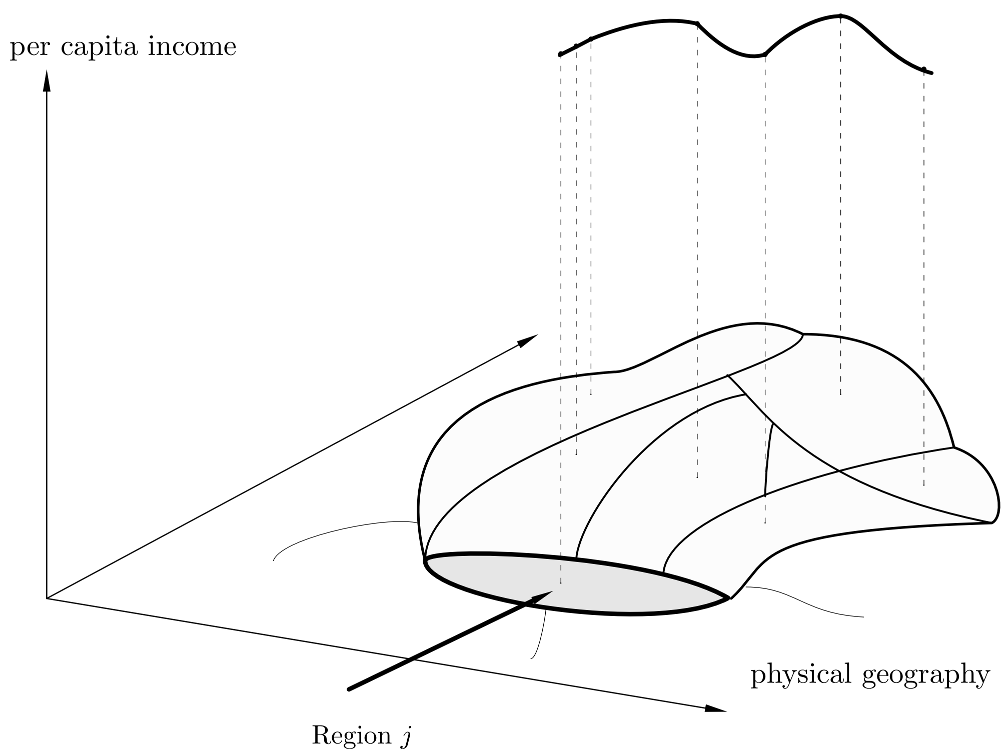


Fig. 3: Regional income distribution dynamics The distinguished region j is spatially contiguous with other regions, and might evolve jointly with them. Income dynamics form a fluctuating blanket on top of the physical geography.

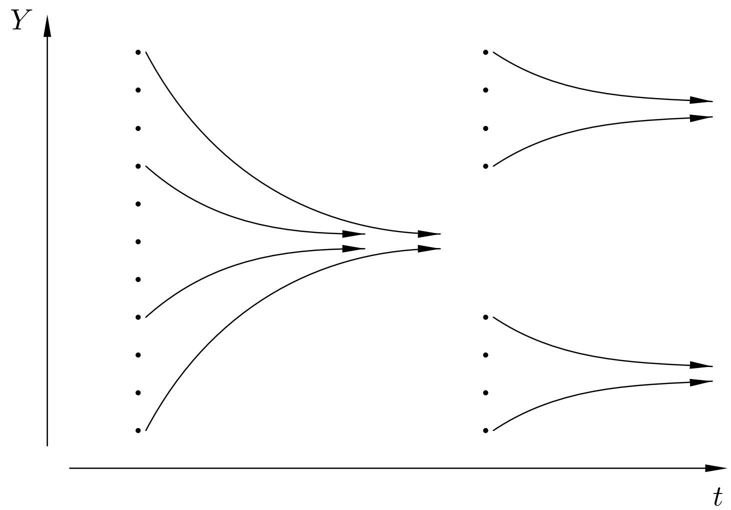


Fig. 4: The effects of (panel-data) conditioning regression
In constructing a conditional representative, one loses information on cross-section clusterings.

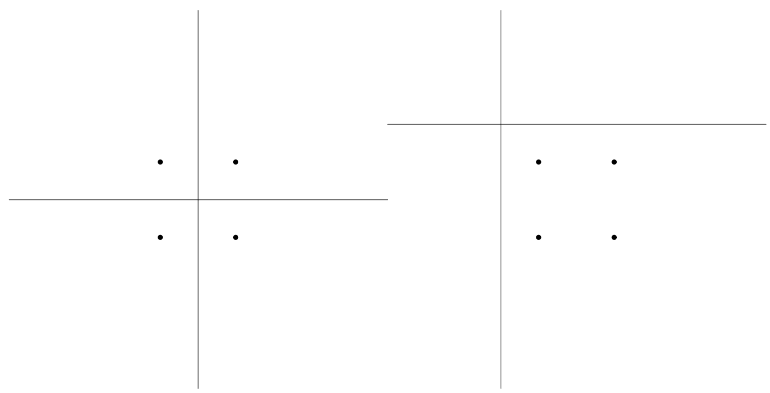


Fig. 5: Arbitrary discrete boundaries. Identical location configurations on continuous space leading to different concentration values

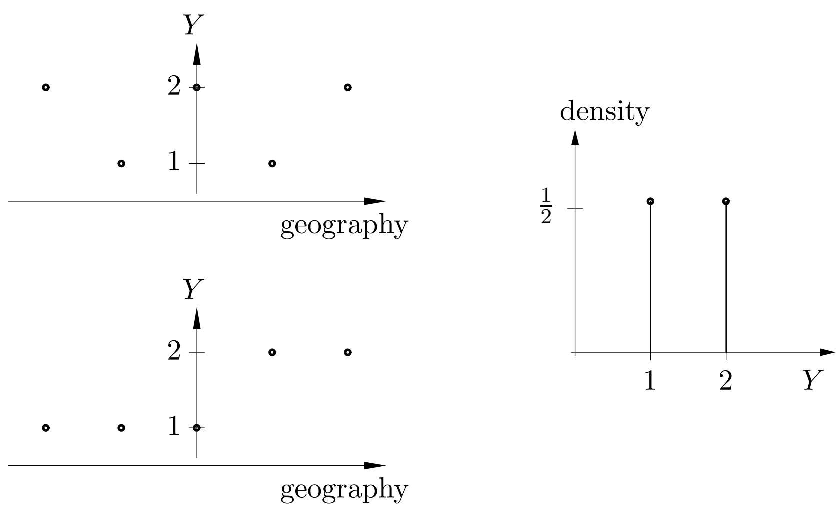


Fig. 6: Aspatial regional distributions Different configurations of spatial location aliased to the same density in economic activity

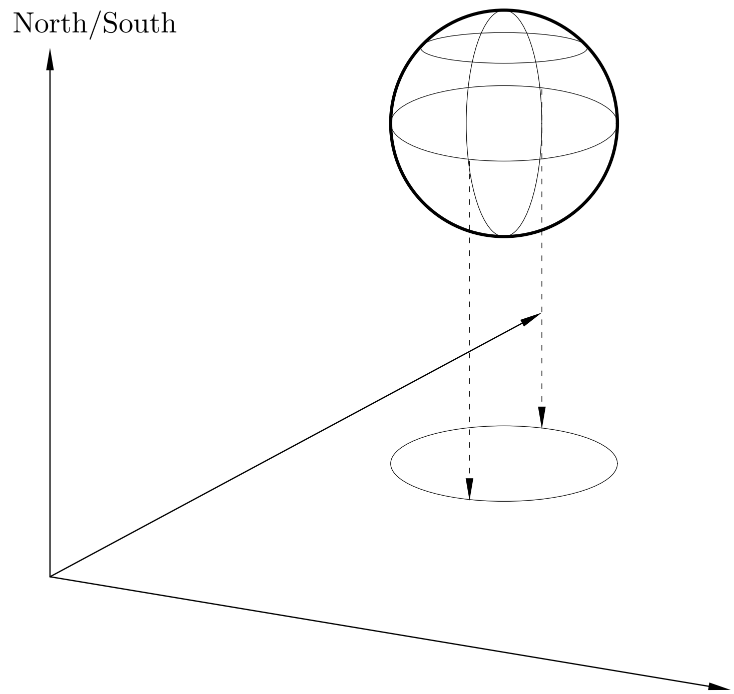


Fig. 7: Globalization in timezones Ignoring degeneracies at the North and South Poles, when time matters but not geographical distance, the homogeneous globe is isomorphic with a circle

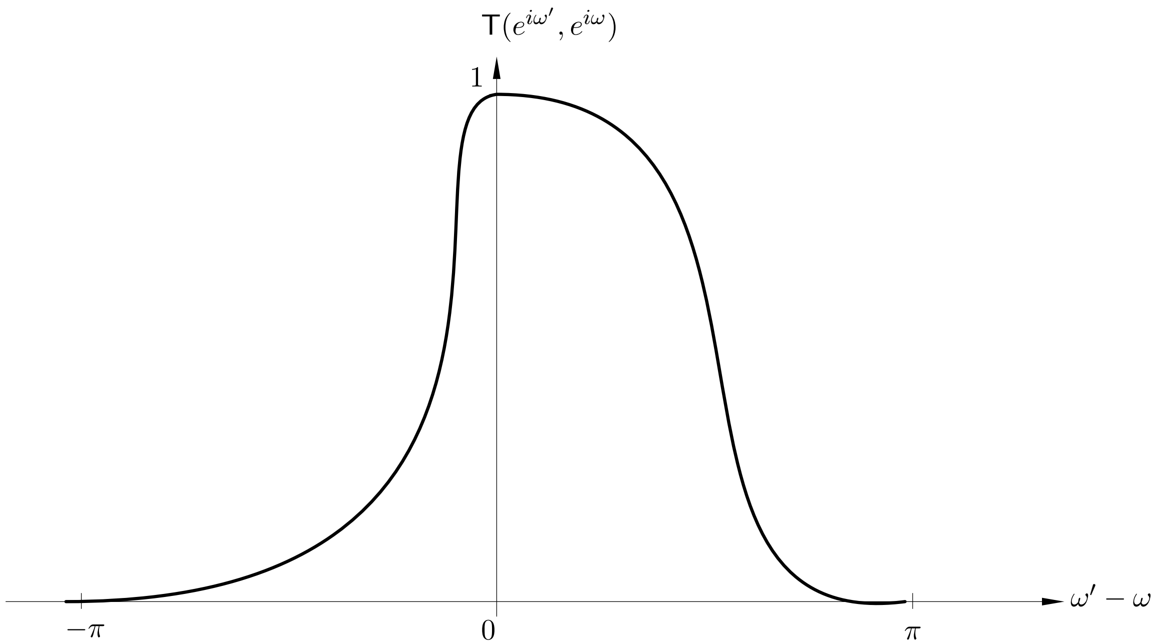


Fig. 9: Timeliness map Subject to (possible) translation costs, knowledge arriving in time is valued; knowledge afterwards, less so.

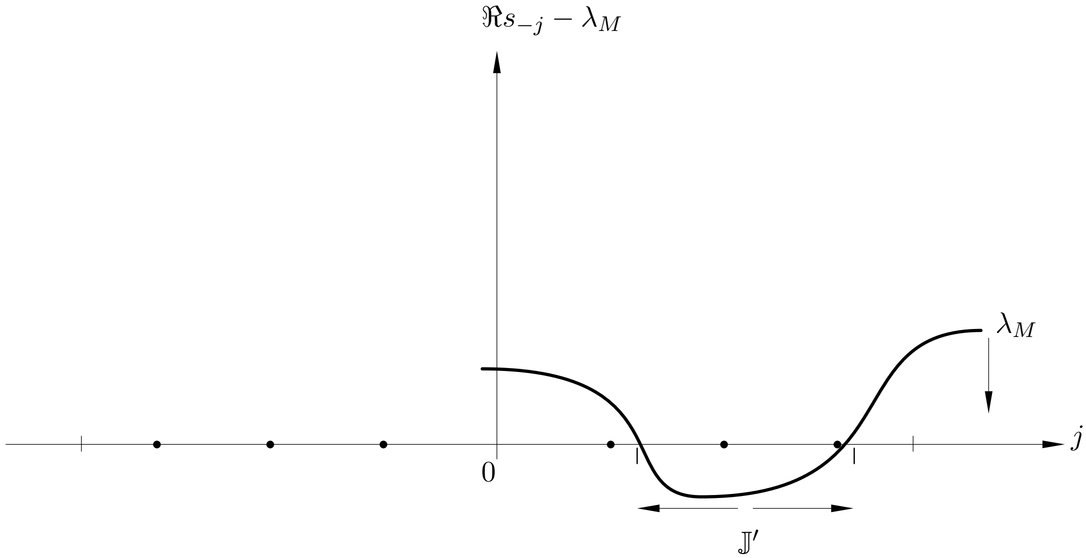


Fig. 10: Real part of the spectrum of θ_M Displaced downwards by λ_M, the components that remain activate the associated complex exponentials in the family of eigenfunctions. The graph is defined only for integer j, but is depicted continuous to ease visualization. In the example shown here, at least two eigenvalues are stable.

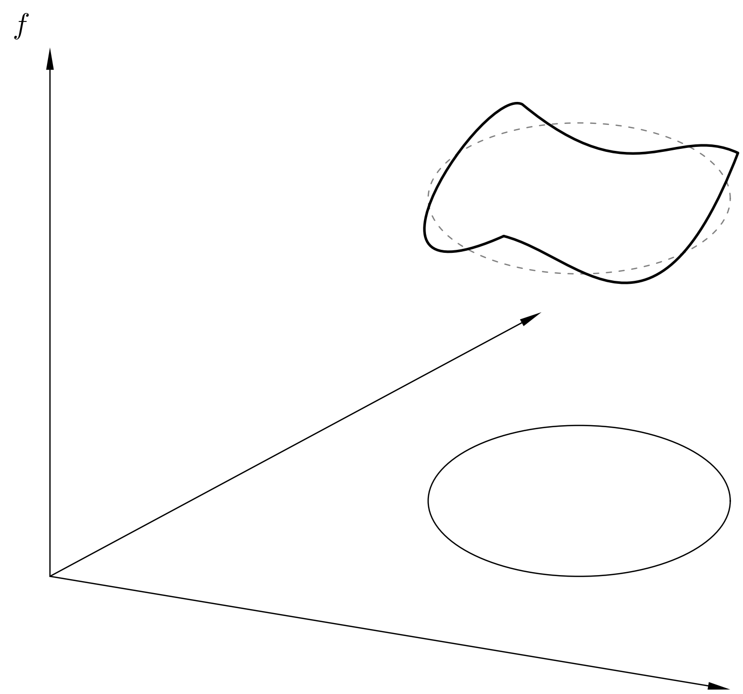


Fig. 11: Cycles in space Local perturbations converge back to steady state only when they display cyclicalities in space