Endogenous Agglomeration in a Many-region World*

Takashi Akamatsu†
Tohoku University

Tomoya Mori‡
Kyoto University and RIETI

Minoru Osawa§
Kyoto University

Yuki Takayama¶
Kanazawa University

September 17, 2020

Abstract: We study a general family of economic geography models that features endogenous agglomeration. In many-region settings, the spatial scale—"global" or "local"—of the dispersion force(s) in a model plays a key role in determining the resulting endogenous spatial patterns and comparative statics. A global dispersion force accrues from competition *between* different locations and leads to the formation of multiple economic clusters (or cities). A local dispersion force is caused by crowding effects *within* each location and induces the flattening of each city. By distinguishing local and global dispersion forces, we can define three prototypical classes, namely, models that have only global, only local, and both local and global dispersion forces. The three model classes engender qualitatively different spatial patterns. Multiple cities are formed only when a global dispersion force is at work; otherwise, only a unimodal distribution can be formed. A city can have its spatial extent only when a local dispersion force is at work. A wide variety of extant models are reduced into the three prototypical classes. Our framework adds consistent interpretations to the empirical literature and also provides general insights into treatment effects in structural economic geography models.

Keywords: agglomeration; dispersion; economic geography; many regions; spatial scale.

JEL Codes: C62, R12, R13

*We are indebted to Michael Pflüger, Esteban Rossi-Hansberg, and Jacques-François Thisse for their constructive comments on the previous version of this paper, entitled "Spatial scale of agglomeration and dispersion: Theoretical foundations and empirical implications" (Akamatsu et al., 2017b). We are grateful for all seminar and conference participants at various institutions. The constructive comments from the reviewers and the editors are appreciated. This research has been supported by JSPS KAKENHI 17H00987, 18H01556, and 19K15108. This research was conducted as part of the project, "Agglomeration-based Framework for Empirical and Policy Analyses of Regional Economies", undertaken at the Research Institute of Economy, Trade and Industry.

†Email: akamatsu@plan.civil.tohoku.ac.jp

‡Email: mori@kier.kyoto-u.ac.jp

§Corresponding author. Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, Kyoto 606-8501 Japan. Tel: +81-75-753-7123. Email: osawa.minoru.4z@kyoto-u.ac.jp / osawa.minoru@gmail.com

¶Email: ytakayama@se.kanazawa-u.ac.jp

1 Introduction

Population and economic activities are astoundingly localized in space. For any spatial resolution—within countries, regions, or cities—disproportional concentrations of population, firms, or shops are present. Currently, just five cities (MSAs) of the United States, which make up about 5% of its cultivated land area, produces over 20% of the country's nominal GDP. In Japan, its three major prefectures account for over 30% of the nominal GDP and 20% of the total population of the country, while taking up less than 5% of the total inhabitable area in the country.

Over the past four decades, spatial economics has developed numerous theoretical and quantitative models to account for the uneven distribution of economic activities across cities and regions. The rich vein of theoretical modeling for endogenous spatial agglomeration (e.g., Fujita et al., 1999b; Baldwin et al., 2003; Duranton and Puga, 2004; Fujita and Thisse, 2013) has been an important source of intuition-building devices. In simplified geographical environments such as two-region models, the peaks and troughs in the space economy are explained as the endogenous outcomes of the various trade-offs between positive and negative incentives for spatial concentration. The accumulated knowledge for the general equilibrium modeling of spatial phenomena, together with the increased availability of fine spatial economic data, has allowed economists to construct quantitative models in a progressively detailed manner (see, e.g., Redding and Rossi-Hansberg, 2017 and Proost and Thisse, 2019, Section 5.2, for surveys). The exponential increase in the number of quantitative studies motivates us to ask the following question: is there any general and systematic means to classify and interpret the various models that are proposed in different contexts?

To provide an answer to this question, we shed new light on *endogenous economic mechanisms* in models of spatial agglomeration. As a natural successor of two-region theories of spatial agglomeration, we study endogenous agglomeration of mobile agents from an ex-ante uniform state in symmetric many-region geography. Our framework covers an important family of models that encompasses a wide range of extant models (from intraurban to interregional models).[1] We show that the *spatial scale* of negative externalities (i.e., *dispersion forces*) in a given model, rather than particular microfoundations of the model, determines the spatial patterns the model engenders endogenously. The spatial scale of a dispersion force is *local* when the force arises from the congestion effects *inside* each region (e.g., urban costs due to higher land rent in cities); it is *global* when the force depends on the proximity structure *between* regions (e.g., competition between firms in different regions due to interregional trade). We show that the spatial pattern can become *multimodal* only when there is a global dispersion force; it becomes *unimodal* when there is only a local dispersion force.

A simple thought experiment can illustrate the qualitative difference between local and global dispersion forces and why the difference leads to different spatial patterns. Consider a many-region model of spatial agglomeration, which incorporates some agglomeration forces. When there are no dispersion forces, all the mobile agents in the model will locate in one region. If some dispersion

[1] For example, we cover all static models—to the best of our knowledge—with a continuum of homogeneous agents with constant-elasticity-of-substitution preferences and a single type of iceberg interregional transport costs (e.g., Krugman, 1991b; Helpman, 1998; Allen and Arkolakis, 2014). We will provide more examples in Section 2.

		Global	
		Absent	Present
Local	Absent	–	**Class I**
	Present	**Class II**	**Class III**

Table 1: Spatial scale(s) of dispersion force(s) and model classes

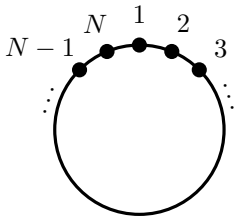


Figure 1: N-region racetrack economy.

force is added to the model, then some mobile agents may deviate from the region. In a many-region economy with heterogeneous interregional distances, the spatial scale of the model's dispersion force(s) affects the resulting spatial patterns. If only local dispersion force is present, mobile agents will *disperse locally*; they move to nearby locations to mitigate congestion, leading to the formation of a unimodal pattern. If only global dispersion force is present, mobile agents will *disperse globally*; they head to a remote region from the populated region to avoid competition, leading to a multimodal pattern. The contrast between unimodal and multimodal patterns cannot be studied in two-region settings where interregional distance is summarized by a single transport cost parameter, as there is only one alternative region to head for.

The classification of dispersion forces sets the basis for a unified taxonomy of theoretical or structural spatial economic models. With the dichotomy between local and global dispersion forces, we can define three prototypical model classes. A model is of Class I (II) if it has only a global (local) dispersion force and it is of Class III if it has both (Table 1).

To illustrate how these model classes differ qualitatively in their implications, we use a symmetric geography as the common testbed. We assume a many-region *racetrack economy*, as considered by Papageorgiou and Smith (1983) and Krugman (1993), in which regions with the same local characteristics are symmetrically located over a circle (Figure 1). The setup allows us to abstract from the exogenous geographical advantages induced by the shape of the transport network or idiosyncratic regional characteristics. Despite the symmetry, this setup preserves the heterogeneity of interregional distances, one of the most important features of the real-world geography. On this geography, we study how endogenous agglomeration occurs due to instability of the uniform distribution, in the spirit of Papageorgiou and Smith (1983).

Proposition 1 shows that the endogenous spatial patterns that can emerge from an ex-ante uniform distribution in the circular economy substantially differ across model classes, i.e., depending on the spatial scale(s) of dispersion force(s) in a model. If a model is of Class I, a multimodal pattern emerges

(a) Multimodal distribution (b) Unimodal distribution

Figure 2: Unimodal and multimodal spatial distributions

from symmetry (Figure 2a). If a model is of Class II, only a unimodal pattern emerges (Figure 2b). If a model is of Class III, then both possibilities arise, depending on the transport cost level. These results accord with the basic workings of the local and global dispersion forces. A multimodal distribution is a manifestation of global repulsion effects *between* economic concentrations and thus emerges from Classes I and III; a unimodal distribution emerges only when local congestion effects *within* economic clusters are effective, and thus it is observed in Classes II and III.

Our numerical simulations on the eight-region racetrack economy (Section 5.2) complement the theoretical predictions of **Proposition 1**. The difference between model classes appears in their responses to interregional transport costs. For Class I models, a multimodal distribution (i.e., many small cities) as shown in Figure 2a endogenously emerge when the transport cost is high. A decrease in the transport cost induces a decrease in the number of cities, an increase in the spacing between them, and an increase in the size of each city. By contrast, in Class II models, when the transport cost is high, there is a unimodal distribution (i.e., a single city) as shown in Figure 2b. When the transport cost decreases, it causes "suburbanization" by reducing the peak population density of the city. Class III is the most general among the three classes. It is a synthesis of Classes I and II. When the transport cost is high, a Class III model behaves as a Class I model and many small cities emerge. When the transport cost is low, a single dispersed city exists, similar to a Class II model. At moderate levels of transport cost, multiple dispersed cities are generated (see Figure 12b). A decrease in transport costs simultaneously causes a decrease in the number of cities (as in Class I models) and the flattening of each city (as in Class II models). Section 6.2 provides numerical simulations on asymmetric settings to illustrate that these qualitative contrasts between model classes in terms of the number, size, and spatial extent of cities robustly generalize beyond the symmetric racetrack economy.

The behavior of Class III models provides a consistent interpretation of several characteristic tendencies observed in the evolution of the population distribution in Japan during 1970–2015 (see Appendix C for details). This period witnessed an almost from-scratch development of highways and high-speed railway networks triggered by the Tokyo Olympics of 1964. The total highway (high-speed railway) length increased from 879 km (515 km) to 14,146 km (5,350 km), which is more than a 16 (10) times increase. Let a "city" be defined as the set of contiguous 1km$\times$1km cells with a population density of at least 1,000/km^2 and a total population of at least 10,000. Then, 302 cities survived throughout the

45 years, experiencing an average 21% increase in population size (controlling for national population growth). That is, there was a selective concentration towards a subset of cities, analogous to the implications of Class I and III models. The process was also associated with the flattening of each city: there was a 94% increase in area size and a 22% decrease in population density for an average individual city, analogous to the predictions of Class II and III models. Given the presence of exogenous asymmetries in the real world, we cannot conclude to what extent endogenous factors, on which we focus in this paper, are responsible for this phenomenon; however, the relative rise and fall of local and global dispersion forces seem to provide consistent interpretations for the real economy.

We also offer an additional result (**Proposition 2**) that reveals the effects of exogenous regional advantages (e.g., differences in amenities or productivity), which play a key role in counterfactual analyses based on calibrated quantitative economic geography models (see Redding and Rossi-Hansberg, 2017). Naturally, for a given transport cost level, an exogenously advantageous region attracts more population than the average. We show that, when interregional access improves from the initial transport cost level, the role of exogenous regional advantages is strengthened and weakened in Class I and II models, respectively. If exogenous heterogeneity causes one region to attract more population, then such asymmetry will be magnified and reduced in Class I and II models, respectively. This again indicates that the spatial scale of the dispersion force in a given model, rather than their particular microfoundations, crucially governs the comparative static results of the model.

Our theoretical results reduce numerous economic geography models to a few classes, according to the spatial scale of their dispersion forces. Our approach is philosophically related to those of Arkolakis et al. (2012) or Allen et al. (2019), who consider and analyze general model classes that encompass a wide range of trade models with unique equilibrium. Our approach is complementary to theirs, in that we focus on the cases where the models feature the multiplicity of equilibria and endogenous agglomeration. Recent evidence suggests that the multiplicity of equilibria and path dependence matter in the space economy in the long run (Bleakley and Lin, 2012; Michaels and Rauch, 2018). Consequently, the models that feature endogenous regional asymmetry can be useful for long-term counterfactual analyses. However, a well-known drawback of such models is that they may exhibit complex behaviors and cause technical and computational difficulties. In this regard, our results on the relationship between the spatial scale of dispersion forces and the resulting spatial patterns are useful. For instance, our classification can be employed for choosing models to quantify under the presence of the possible multiplicity of equilibria. Class III is the most general and may thus replicate the reality best among the three classes, as seen from the Japan example discussed above. Class I would suffice if we are interested in the global patterns of economic agglomeration such as the number and population size of cities. If a major city region with a monocentric structure is the scope of the analysis, then Class II may be a reasonable choice.

Section 2 introduces a class of economic geography models this paper focuses on. As a primer for our approach, we explore the simplest geographical setup, a two-region economy. Section 3 introduces the formal definitions of the spatial scale of dispersion forces and model classes. Section 4 presents **Proposition 1**. Section 5 provides illustrative examples for **Proposition 1** and relate our results to the

empirical literature. Section 6 explores the effects of asymmetries to show the robustness of implications from formal discussions in the symmetric racetrack economy; **Proposition 2** is provided in this section. Section 7 concludes the paper. All proofs are in Appendix A.

2 Basic framework

We introduce a generic format for many-region economic geography models. The symmetric racetrack economy is explained. Definition 1 introduces *canonical models*, the fundamental model class we focus on. Several two-region examples are discussed for illustration.

2.1 Economic geography models: A general format

Consider an economy comprised of N regions. Let $\mathcal{N} \equiv \{1, 2, \ldots, N\}$ be the set of regions. There is a continuum of agents. Each agent chooses a region to locate in. Let $x_i \geq 0$ be the mass of agents in region i and $\boldsymbol{x} \equiv (x_i)_{i \in \mathcal{N}}$ be the spatial distribution of agents. We assume $\sum_{i \in \mathcal{N}} x_i = 1$. The set of all possible spatial distributions is $\mathcal{X} \equiv \{\boldsymbol{x} \geq \boldsymbol{0} \mid \sum_{i \in \mathcal{N}} x_i = 1\}$. The payoff for the agents in region i is given by $v_i(\boldsymbol{x})$. We assume that $\boldsymbol{v}(\boldsymbol{x}) \equiv (v_i(\boldsymbol{x}))_{i \in \mathcal{N}}$ is one-time differentiable if $x_i > 0$ for all $i \in \mathcal{N}$.

Throughout our analyses, the term *region* simply indicates a discrete spatial unit in which a mobile agent can locate. Whether the model is interpreted to be intraurban, interregional, or international is not essential for our formal results. A "region" may alternatively be called an urban zone, a municipality, a country, and so forth.

Agents can freely relocate across N regions to improve their payoffs. Then, $\boldsymbol{x} \in \mathcal{X}$ is a *spatial equilibrium* if the following Nash equilibrium condition is met:

$$\begin{cases} v^* = v_i(\boldsymbol{x}) \text{ for any region } i \in \mathcal{N} \text{ with } x_i > 0, \\ v^* \geq v_i(\boldsymbol{x}) \text{ for any region } i \in \mathcal{N} \text{ with } x_i = 0, \end{cases} \tag{1}$$

where $v^* = \max_{i \in \mathcal{N}} \{v_i(\boldsymbol{x})\}$ is the associated equilibrium payoff level.

An indispensable feature of an economic geography model is the presence of spatial frictions, or distance-decay effects. Examples include those for the shipment of goods or for communication among agents. Thus, $\boldsymbol{v}$ depends on a *proximity matrix* $\mathbf{D} = [\phi_{ij}]$ that summarizes the interregional transport costs. Each entry $\phi_{ij} \in (0, 1]$ is the freeness of interactions between regions i and j (e.g., freeness of trade, freeness of transportation, freeness of social communication).

Payoff function $\boldsymbol{v}$ can include positive and negative externalities of spatial concentration, which may depend on interregional transport costs. Owing to the positive externalities, economic geography models often face multiple spatial equilibria. It is customary to introduce equilibrium refinement based on *local stability* under myopic dynamics. We follow this strategy. All the formal claims on the stability of equilibria in this paper hold true for the various standard dynamics employed in the literature.

2.2 Symmetric racetrack economy

To focus on the roles of endogenous economic interactions in many-region settings, this paper considers an N-region economy in which regions are symmetrically placed over a circle and interactions are possible only through the circular network (see Figure 1 in Section 1).

Assumption RE (Racetrack economy)**.** Proximity matrix $\mathbf{D} = [\phi_{ij}]$ is given by $\phi_{ij} = \phi^{\ell_{ij}}$, where $\phi \in (0,1)$ is the freeness of interactions between two consecutive regions and $\ell_{ij} \equiv \min\{|i-j|, N-|i-j|\}$ is the shortest-path distance over the circumference. Further, N is a multiple of four.

This assumption abstracts from the location-fixed advantages induced by the *shape* of the underlying transport network. For example, in a long narrow economy, the peripheral regions have fewer opportunities to access the other regions; the central portion is advantageous because of the shape of space. In a racetrack economy, by contrast, every region has the same level of accessibility to the other regions.[2] We assume that N is a multiple of four only for expositional simplicity. See Appendix A.1, in particular Remark A.3 for the cases in which N is an odd or just an even. Also, we will use two-region examples where we do not need a many-region economy for discussion.

Further, we assume that payoff function v does not introduce any ex-ante asymmetries across regions. We abstract away all asymmetries in regional characteristics to focus on endogenous factors. Technically, this can be formalized as follows.

Assumption S (Symmetry of payoff function)**.** For all $\boldsymbol{x} \in \mathcal{X}$, payoff function v satisfies $v(\mathbf{P}\boldsymbol{x}) = \mathbf{P}v(\boldsymbol{x})$ for all permutation matrices $\mathbf{P}$ that satisfy $\mathbf{PD} = \mathbf{DP}$.

Assumption S is called *equivariance* (see Golubitsky and Stewart, 2003).

Example 2.1. Suppose $N = 4$. Then, Assumption RE is that

$$\mathbf{D} = \begin{bmatrix} 1 & \phi & \phi^2 & \phi \\ & 1 & \phi & \phi^2 \\ & & 1 & \phi \\ \textit{Sym.} & & & 1 \end{bmatrix}. \tag{2}$$

The shape of the circular economy is invariant even if we swap the indices of regions 1 and 3. The following permutation matrix represents the effect of this re-indexing:

$$\mathbf{P} = \begin{bmatrix} & & 1 & \\ & 1 & & \\ 1 & & & \\ & & & 1 \end{bmatrix}. \tag{3}$$

We see $\mathbf{PD} = \mathbf{DP}$ and $\mathbf{P}$ satisfies the hypothesis of Assumption S. In this way, $\mathbf{PD} = \mathbf{DP}$ ensures that the adjacency relationships between regions remain invariant under the permutation of the indices represented by $\mathbf{P}$. The re-indexed spatial distribution is $\boldsymbol{x}' = \mathbf{P}\boldsymbol{x}$, where $\boldsymbol{x}$ is the original one. If v

[2]Our analysis is thus complementarity to the many-region analysis by Matsuyama (2017) who, by using a specific trade model as as the basic lens, focused on the role of the shape of transport network.

does not include any region-fixed exogenous advantages, we must have $v_1(\boldsymbol{x}') = v_3(\boldsymbol{x})$, $v_2(\boldsymbol{x}') = v_2(\boldsymbol{x})$, $v_3(\boldsymbol{x}') = v_1(\boldsymbol{x})$, and $v_4(\boldsymbol{x}') = v_4(\boldsymbol{x})$, that is, $\boldsymbol{v}(\boldsymbol{x}') = \boldsymbol{v}(\mathbf{P}\boldsymbol{x}) = \mathbf{P}\boldsymbol{v}(\boldsymbol{x})$ as in Assumption S. ■

Under Assumption RE and S, the uniform distribution of agents $\bar{\boldsymbol{x}} \equiv (\bar{x}, \bar{x}, \ldots, \bar{x})$ with $\bar{x} = \frac{1}{N}$ is always a spatial equilibrium. This paper considers what spatial patterns can emerge from $\bar{\boldsymbol{x}}$.

2.3 A first view of endogenous agglomeration

Before studying endogenous spatial patterns in the N-region economy, this section revisits a simple two-region example for illustration. Suppose there are two regions that have identical characteristics. The proximity matrix for this setup is expressed as follows:

$$\mathbf{D} = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}, \tag{4}$$

where $\phi \in (0, 1)$ is the freeness of interactions between the two regions.

The uniform distribution of agents, $\bar{\boldsymbol{x}} \equiv (\bar{x}, \bar{x})$ with $\bar{x} = \frac{1}{2}$, is always a spatial equilibrium. How such a symmetric pattern becomes unstable and an endogenous regional asymmetry of the form $\boldsymbol{x} = (x', x'')$ with $x' > x''$ is generated? As demonstrated by Papageorgiou and Smith (1983) as well as Krugman (1991b), local instability of $\bar{\boldsymbol{x}}$ leads to the formation of spatial agglomeration.

There is a general model-independent characterization for the stability of $\bar{\boldsymbol{x}}$: it is stable (unstable) if the payoff gain of an agent relocating from one region to the other is negative (positive). The gain for a deviant can be evaluated by the elasticity of the payoff difference:

$$\omega = \frac{\bar{x}}{\bar{v}} \frac{\partial (v_1(\bar{\boldsymbol{x}}) - v_2(\bar{\boldsymbol{x}}))}{\partial x_1} = \frac{\bar{x}}{\bar{v}} \left(\frac{\partial v_1(\bar{\boldsymbol{x}})}{\partial x_1} - \frac{\partial v_2(\bar{\boldsymbol{x}})}{\partial x_1} \right), \tag{5}$$

where $\bar{v}$ is the uniform payoff level at $\bar{\boldsymbol{x}}$, so that $\boldsymbol{v}(\bar{\boldsymbol{x}}) = (\bar{v}, \bar{v})$.

If $\omega < 0$, then a marginal increase in the mass of agents in a region induces a *relative payoff decrease* therein compared to the other region; there are no incentives for agents to migrate and thus $\bar{\boldsymbol{x}}$ is stable. In fact, $\omega < 0$ implies that $\bar{\boldsymbol{x}}$ is an evolutionary stable states (Sandholm, 2010, Observation 8.3.11). If otherwise $\omega > 0$, the migration of a small fraction of agents from region 2 to 1 induces a *relative payoff increase* in region 1 compared to region 2; it means that agents are strictly better off when they migrate to the other region, thereby $\bar{\boldsymbol{x}}$ is unstable.

Suppose that $\bar{\boldsymbol{x}}$ is initially stable, i.e., $\omega < 0$. Then, migration become profitable for mobile agents (i.e., endogenous regional asymmetry emerges) when ω turns to positive. By its construction, ω is a function of exogenous parameters of $\boldsymbol{v}$. Monotonic change of the freeness of interregional interactions ϕ can switch the sign of ω to destabilize $\bar{\boldsymbol{x}}$.

We can show that ω is an eigenvalue of $\mathbf{V} \equiv \frac{\bar{x}}{\bar{v}} \nabla \boldsymbol{v}(\bar{\boldsymbol{x}}) = \frac{\bar{x}}{\bar{v}} [\frac{\partial v_i}{\partial x_j}(\bar{\boldsymbol{x}})]$, the matrix of payoff elasticity with respect to $\boldsymbol{x}$ evaluated at $\bar{\boldsymbol{x}}$, and its associated eigenvector is $\boldsymbol{z} \equiv (1, -1)$; that is, $\mathbf{V}\boldsymbol{z} = \omega \boldsymbol{z}$. This is because (5) implies $\mathbf{V}\boldsymbol{z} = \omega \boldsymbol{z}$, as we note that $\frac{\partial v_1}{\partial x_2}(\bar{\boldsymbol{x}}) = -\frac{\partial v_2}{\partial x_1}(\bar{\boldsymbol{x}})$ and $\frac{\partial v_1}{\partial x_1}(\bar{\boldsymbol{x}}) = \frac{\partial v_2}{\partial x_2}(\bar{\boldsymbol{x}})$ because of the ex-ante symmetry of the regions. Since $\boldsymbol{z}$ represents a population increase in one region and a decrease in the other, it is *the* migration pattern in the two-region economy. Obviously, $\boldsymbol{z}$ is model independent.

2.4 Canonical models

As the two-region example in Section 2.3 illustrates, endogenous economic forces in a model at the symmetric state $\bar{x}$ is summarized by the eigenvalue(s) of $\mathbf{V}$. In our many-region analysis, we focus on a specific family of models, which we call *canonical models*, by imposing an assumption on the structure of $\mathbf{V}$ so that we can derive its eigenvalues analytically. As we will illustrate with examples, canonical models encompass a wide range of extant economic geography models.

Definition 1 (Canonical models)**.** Consider an economic geography model with payoff function v parametrized by proximity matrix $\mathbf{D} = [\phi_{ij}]$. Suppose Assumptions RE and S and let $\bar{\mathbf{D}}$ be the row-normalized version of $\mathbf{D}$, whose (i,j)th element is given by $\frac{\phi_{ij}}{\sum_{k\in\mathcal{N}}\phi_{ik}}$. The model is *canonical* if there exists a rational function G that is continuous over $[0,1]$ and satisfies

$$\mathbf{V} = G(\bar{\mathbf{D}}), \tag{6}$$

where $\mathbf{V} = \frac{\bar{x}}{\bar{v}}\nabla v(\bar{x})$ is the payoff elasticity matrix at $\bar{x}$. We call G the *gain function* of the model.

In Definition 1, for a rational function (i.e., the ratio of two polynomials) of form $G(t) = \frac{G^\sharp(t)}{G^\flat(t)}$ with polynomials $G^\sharp(t)$ and $G^\flat(t) \neq 0$, we define $G(\bar{\mathbf{D}}) = G^\flat(\bar{\mathbf{D}})^{-1}G^\sharp(\bar{\mathbf{D}})$, where, for a polynomial $G^\sharp(t) = c_0 + c_1 t + c_2 t^2 + \cdots$, we let $G^\sharp(\bar{\mathbf{D}}) = c_0\mathbf{I} + c_1\bar{\mathbf{D}} + c_2\bar{\mathbf{D}}^2 + \cdots$, with $\mathbf{I}$ being the identity matrix. The assumption that G is rational is not restrictive because any continuous function defined on a closed interval can be approximated as closely as desired by a polynomial according to the Weierstrass approximation theorem.

Consider a canonical model, so that we have $\mathbf{V} = G(\bar{\mathbf{D}})$ with a rational function G. Suppose that $(\chi, \boldsymbol{z})$ is an eigenvalue–eigenvector pair of $\bar{\mathbf{D}}$, i.e., $\bar{\mathbf{D}}\boldsymbol{z} = \chi\boldsymbol{z}$. Then, it is known that $(\omega, \boldsymbol{z})$, where

$$\omega = G(\chi), \tag{7}$$

is the associated eigenvalue–eigenvector pair of $\mathbf{V}$, i.e., $\mathbf{V}\boldsymbol{z} = \omega\boldsymbol{z}$ (Fact B.1 in Appendix B). Every eigenvalue χ of $\bar{\mathbf{D}}$ is a monotonically decreasing in ϕ under Assumption RE (Lemma B.2 in Appendix B).

For a wide range of extant models, there are two matrix polynomials, $G^\sharp(\bar{\mathbf{D}})$ and $G^\flat(\bar{\mathbf{D}})$ that satisfy $\mathbf{V} = G^\flat(\bar{\mathbf{D}})^{-1}G^\sharp(\bar{\mathbf{D}})$; thereby, there is a rational function G that satisfies the hypotheses in Definition 1. Definition 1 covers, for example, models of endogenous city center formation (e.g., Beckmann, 1976), economic geography models with a single monopolistically competitive industry (e.g., Krugman, 1991b; Helpman, 1998), and economic geography variants of the "universal gravity" framework (Allen et al., 2019), which in turn encompasses perfectly competitive Armington models with labor mobility (Allen and Arkolakis, 2014). Section 4 provides more examples. In particular, canonical models include models that assume (i) a single type of homogeneous mobile agents with constant-elasticity-of-substitution preferences and (ii) a single sector that is subject to iceberg interregional transport costs.[3]

[3] As noted by Allen et al. (2019) and Arkolakis et al. (2012), this class of models includes various important models in the literature. The iceberg assumption is widely employed in both theoretical and quantitative studies for tractability. However, we should also note that it is not an innocuous assumption for modeling a spatial economy (see, e.g., Hummels and Skiba, 2004; Irarrazabal et al., 2015; Proost and Thisse, 2019, Section 3.5.2).

Gain function G summarizes the endogenous effects in the model. In fact, we can formally define agglomeration and dispersion forces.

Definition 2. A *dispersion (agglomeration) force* in a canonical model is a negative (positive) term in its gain function G.

Example 2.2 (The Beckmann model)**.** Beckmann (1976) proposed a seminal model for the formation of an urban center within a city. Numerous variants of the model have been proposed since. Although the original formulation assumes a continuous space, we consider a discrete-space version to fit the model to our context. Consider the following multiplicative specification, where $\gamma > 0$:

$$v_i(\boldsymbol{x}) = x_i^{-\gamma} E_i(\boldsymbol{x}). \tag{8}$$

The first component, $x_i^{-\gamma}$, reflects *negative externalities* due to congestion. The second, $E_i(\boldsymbol{x})$, represents *positive externalities*; agents prefer proximity to others. A typical specification for $E_i(\boldsymbol{x})$ is

$$E_i(\boldsymbol{x}) = \sum_{j \in \mathcal{N}} \mathrm{e}^{-\tau \ell_{ij}} x_j, \tag{9}$$

where $\tau > 0$ is the distance-decay parameter and $\ell_{ij} > 0$ is the distance between i and j (Tabuchi, 1986; Fujita and Ogawa, 1982; de Palma et al., 2019). The proximity matrix for the model is $\mathbf{D} = [\mathrm{e}^{-\tau \ell_{ij}}]$.

We can show that $\mathbf{V} = -\gamma \mathbf{I} + \bar{\mathbf{D}}$. By letting $G(t) = -\gamma + t$, we have $\mathbf{V} = G(\bar{\mathbf{D}})$. Thus, the model is a canonical model whose gain function is $G(t) = -\gamma + t$. Negative term $-\gamma$ in $G(t)$ corresponds to the congestion effect through $x_i^{-\gamma}$ and positive term t to positive externalities $E_i(\boldsymbol{x})$. The former is the loss from congestion, whereas the latter is the gains from the additional proximity to be induced by migration. The two effects are summarized by G.

If $N = 2$, $\phi = \mathrm{e}^{-\tau \ell_{12}} = \mathrm{e}^{-\tau \ell_{21}} \in (0, 1)$ is the level of externalities that spill over from one location to the other. We have $\bar{\mathbf{D}} = \frac{1}{1+\phi} \mathbf{D}$. The only possible migration direction in this case, $\boldsymbol{z} = (1, -1)$, is an eigenvector of $\bar{\mathbf{D}}$ because it satisfies $\bar{\mathbf{D}} \boldsymbol{z} = \chi \boldsymbol{z}$ with

$$\chi = \frac{1 - \phi}{1 + \phi} \in (0, 1). \tag{10}$$

We observe that χ is monotonically decreasing in $\phi \in (0, 1)$; if ϕ is small (large), χ is large (small). We have $\omega = G(\chi) = -\gamma + \chi$ with $\chi = \frac{1-\phi}{1+\phi}$. We see χ disappears when $\phi \approx 1$, reflecting that the relative location in the economy becomes irrelevant for agents when ϕ is large; only congestion effect $-\gamma$ is left in that case. If $\gamma < 1$, then $\bar{\boldsymbol{x}}$ is stable for $\phi \in (\phi^*, 1)$ and unstable for $\phi \in (0, \phi^*)$, where $\phi^* \equiv \frac{1-\gamma}{1+\gamma}$ is the solution for $\omega(\phi) = G(\chi(\phi)) = 0$; the uniform distribution is unstable when $\phi \in (0, \phi^*)$, i.e., there must be endogenous asymmetry when $\phi \in (0, \phi^*)$. If $\gamma \geq 1$, then $\bar{\boldsymbol{x}}$ is stable for *all* $\phi \in (0, 1)$. That is, strong congestion effects suppress endogenous agglomeration. ■

3 Spatial scale of dispersion forces and the three model classes

This section considers the models by Krugman (1991b) and Helpman (1998) to illustrate how the *spatial scale* of dispersion force affects the mechanics of spatial agglomeration. Definition 4 introduces the spatial scale of a dispersion force, which is the main concept used in this paper. Definition 5 introduces three prototypical model classes we can find in the literature.

3.1 The reversed scenarios of Krugman and Helpman

In the two-region economy, the Krugman and Helpman models are known to exhibit a sharp contrast regarding *when* endogenous regional asymmetry emerges. In the Krugman (Helpman) model, uniform distribution $\bar{x}$ is stable when ϕ is low (high) and spatial agglomeration occurs when ϕ is high (low). Their predictions are thus "opposites" of each other and "Krugman's scenario is reversed" in the Helpman model (Fujita and Thisse, 2013, Chapter 8).

Brief definitions of the many-region versions of the models follow. See Appendix D for details.

Example 3.1 (The Krugman model)**.** The payoff function of the Krugman model is given by

$$v_i(\boldsymbol{x}) = w_i(\boldsymbol{x}) P_i(\boldsymbol{x})^{-\mu} \qquad \forall i \in \mathcal{N}, \tag{11}$$

where $w_i(\boldsymbol{x})$ is the nominal wage of mobile workers for a given spatial distribution of mobile workers $\boldsymbol{x}$ and $P_i(\boldsymbol{x})$ is the Dixit–Stiglitz price index in region i:

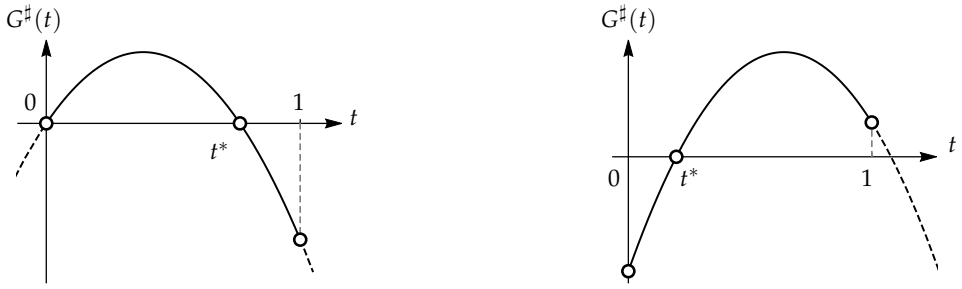
$$P_i(\boldsymbol{x}) \equiv \left(\sum_{j \in \mathcal{N}} x_j \left(w_j(\boldsymbol{x}) \tau_{ji} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \qquad \forall i \in \mathcal{N}, \tag{12}$$

where $\mu \in (0,1)$ is the expenditure share of manufactured goods, $\sigma > 1$ the elasticity of substitution between horizontally differentiated varieties, and $\tau_{ij} \geq 1$ the iceberg transport cost; τ_{ij} units must be shipped from origin i for one unit to arrive at destination j. Nominal wage $\boldsymbol{w}(\boldsymbol{x}) = (w_i(\boldsymbol{x}))_{i \in \mathcal{N}}$ is the unique solution for a system of nonlinear equations that summarizes the market equilibrium conditions under a fixed $\boldsymbol{x}$ (i.e., the gravity flows of interregional trade, goods and labor market clearing, and the zero-profit condition of monopolistically competitive firms):

$$w_i x_i = \sum_{j \in \mathcal{N}} \frac{x_i \left(w_i \tau_{ij} \right)^{1-\sigma}}{\sum_{k \in \mathcal{N}} x_k \left(w_k \tau_{kj} \right)^{1-\sigma}} e_j \qquad \forall i \in \mathcal{N}, \tag{13}$$

where $e_i \equiv \mu \left(w_i x_i + l_i \right)$ is region i's expenditure on differentiated goods and $l_i > 0$ the immobile demand in region i. The proximity matrix is $\mathbf{D} = [\phi_{ij}] = [\tau_{ij}^{1-\sigma}]$. ■

Example 3.2 (The Helpman model)**.** Using the same notation as in the Krugman model, the payoff



(a) The Krugman model (b) The Helpman model

Figure 3: $G^\sharp(t)$ for the Krugman and Helpman models

function of mobile agents in the Helpman model is given by:

$$v_i(\boldsymbol{x}) = \left(\frac{x_i}{a_i}\right)^{-\gamma} (w_i(\boldsymbol{x}) + \bar{r})^{\mu} P_i(\boldsymbol{x})^{-\mu} \qquad \forall i \in \mathcal{N}, \tag{14}$$

where a_i is the endowment of housing stock in region i, $\gamma \equiv 1 - \mu \in (0,1)$ the expenditure share of housing goods, and $\bar{r} = \gamma \sum_{i \in \mathcal{N}} w_i x_i$ is an equal dividend from the total rental revenue from housing in the economy. The market equilibrium conditions under a given $\boldsymbol{x}$ are summarized by (13) with $e_i = \mu(w_i + \bar{r})x_i$. The proximity matrix for the model is the same as that of the Krugman model. ■

We revisit the "reversed scenario" in the symmetric two-region economy with our notations. Suppose $N = 2$. Let $\tau_{12} = \tau_{21} = \tau > 1$ and other regional characteristics, (l_i, a_i), are uniform. Then, $\mathbf{D}$ is given by (4) with $\phi \equiv \tau^{1-\sigma}$. Appendix D shows that for both the models, $\mathbf{V} = \frac{\bar{x}}{\bar{v}} \nabla \boldsymbol{v}(\bar{\boldsymbol{x}})$ is given by

$$\mathbf{V} = G^\flat(\bar{\mathbf{D}})^{-1} G^\sharp(\bar{\mathbf{D}}), \tag{15}$$

where $G^\flat(t) \equiv 1 - \frac{\mu}{\sigma} t - \frac{\sigma - 1}{\sigma} t^2$ and

$$G^\sharp(t) = \quad C_1 t - \quad C_2 t^2 \qquad \text{(the Krugman model)}, \tag{16}$$

$$G^\sharp(t) = -\gamma + C_1 t - (C_2 - \gamma) t^2 \qquad \text{(the Helpman model)}, \tag{17}$$

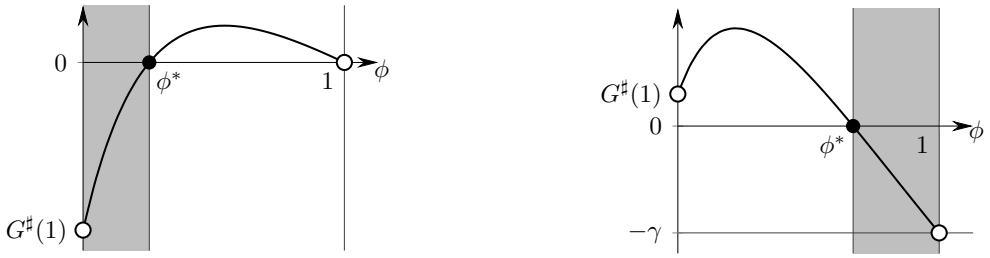
with $C_1 \equiv \frac{\mu}{\sigma - 1} + \frac{\mu}{\sigma}$ and $C_2 \equiv \frac{\mu^2}{\sigma - 1} + \frac{1}{\sigma}$. Thus, both the two models are canonical. The gain function for these models is given by

$$G(t) = \frac{G^\sharp(t)}{G^\flat(t)}. \tag{18}$$

Since we assume $N = 2$, we have $\omega = G(\chi)$ where $\chi = \frac{1-\phi}{1+\phi}$ as in Example 2.2, with $\phi = \tau^{1-\sigma}$.[4]

For each model, $G^\sharp$ summarizes the *net* relative magnitudes of the agglomeration and dispersion forces. As we have $G^\flat(\chi(\phi)) > 0$ for all ϕ, $\bar{\boldsymbol{x}}$ is stable for the ranges of ϕ that satisfies $\omega^\sharp \equiv G^\sharp(\chi(\phi)) < 0$.

[4] In the context of two-region "new economic geography" models, Fujita et al. (1999b) calls χ "a sort of index of trade cost" (page 57), whereas Baldwin et al. (2003) calls it "a convenient measure of closed-ness" (page 46).



(a) The Krugman model

(b) The Helpman model

Figure 4: $\omega^\sharp \equiv G^\sharp(\chi(\phi))$ for the Krugman and Helpman models

Figure 3 shows $G^\sharp(t)$ for the Krugman and Helpman models. We assume that $G^\sharp(t)$ has one root t^* in $(0,1)$. This requirement will be formalized as Assumption E in Section 3.3. If no such t^* exists, there is no switch in the stability of $\bar{\boldsymbol{x}}$ for $\phi \in (0,1)$.

The "reversed scenario" is illustrated in Figure 4. It shows composite function $\omega^\sharp(\phi) \equiv G^\sharp(\chi(\phi))$ for each model; $\bar{\boldsymbol{x}}$ is stable if $\omega^\sharp < 0$. The shaded region in each panel shows the range of ϕ where $\bar{\boldsymbol{x}}$ is stable; concretely, $\bar{\boldsymbol{x}}$ in the Krugman model is stable if $\phi \in (0, \phi^*)$, whereas $\bar{\boldsymbol{x}}$ in the Helpman model is stable if $\phi \in (\phi^*, 1)$. As expected, $\bar{\boldsymbol{x}}$ is stable for low (high) values of ϕ in the Krugman (Helpman) model and unstable otherwise.

3.2 Spatial scale of dispersion forces

The "reversed scenario" in the two-region economy stems from differences in the nature of their *dispersion forces*. In (16), the agglomeration (dispersion) force in the Krugman model is captured by the first (second) term. In the Helpman model, the second term in (17) reflects the agglomeration force, whereas the first and third terms the dispersion forces.

In fact, the agglomeration forces in the two models are equivalent; $C_1\chi$ in $\omega = G^\sharp(\chi)$ arises from the price index of the differentiated varieties (12). A region with a larger set of suppliers (firms) in the market has a lower price index. Mobile workers prefer such a region if the nominal wage is the same. The agglomeration force declines as ϕ increases; reflecting this, $C_1\chi = C_1\frac{1-\phi}{1+\phi}$ is smaller when ϕ is larger.

The dispersion force in the Krugman model is the so-called *market-crowding effect* between firms (Baldwin et al., 2003, Chapter 2). If a firm is geographically close to others, the firm can only pay a low nominal wage because of competition. Therefore, mobile workers are discouraged to enter a region in which firms face fierce market competition with other firms in that location as well as *nearby regions* thereof. The dispersion force thus depends on proximity structure $\mathbf{D}$ and appears as a negative second-order term, $-C_2\chi^2$. This force is stronger when χ is large, that is, when ϕ is small.

The main dispersion force in the Helpman model, on the other hand, represents a *local congestion effect* that stems from competition in the housing market of each region.[5] The housing market does not

[5]The market-crowding effect also exists in the Helpman model: $-(C_2 - \gamma)\chi^2$ in $G^\sharp(\chi)$. However, under any choice of μ and σ, it cannot stabilize $\bar{\boldsymbol{x}}$ when ϕ is small, as we will discuss in Section 3.3.

depend on interregional proximity structure **D** but only on the mass of agents *within* each region. The force thus appears in $G^\sharp$ as negative constant term, $-\gamma$. Since the agglomeration force ($C_1\chi$) declines as ϕ increases, the relative strength of the dispersion force rises with ϕ.

The comparison between the Krugman and Helpman models highlights that the key difference is whether the dispersion force acts *within* or *between* locations. To denote this distinction, we introduce the formal notion of *spatial scale* of dispersion forces.

We first introduce *net gain functions* that ignores the denominator of G, $G^\flat$, which is positive and thus irrelevant for the stability of $\bar{\boldsymbol{x}}$.

Definition 3. A *net gain function* $G^\sharp$ for a canonical model with gain function G is a polynomial that satisfies $\text{sgn}[G(t)] = \text{sgn}[G^\sharp(t)]$ for all $t \in (0,1)$.

Example 3.3. The net gain functions for the Krugman and Helpman models are, respectively, given by (16) and (17), as $G^\flat(t) > 0$ for all $t \in (0,1)$. For the Beckmann model, $G^\sharp(t) = G(t) = -\gamma + t$. ■

With this preparation, we can introduce the spatial scale of dispersion forces, which refines the definition of dispersion forces in Definition 2.

Definition 4 (Spatial scale of dispersion forces)**.** A negative constant term in net gain function $G^\sharp$ is called a *local dispersion force*. The other negative terms in $G^\sharp$ are called *global dispersion forces*.

3.3 Three prototypical model classes

By identified local and global dispersion forces, we can introduce a formal categorization of canonical models based on three prototypical shapes of G. As shown in Table 1, the three-type categorization corresponds to the spatial scale(s) of the dispersion force(s).

Definition 5. A canonical model with gain function G is said to be:

(a) *Class I*, if there can be one and at most one root $t^* \in (0,1)$ for G such that $G(t) > 0$ for $t \in (0, t^*)$, $G(t^*) = 0$, and $G(t) < 0$ for $t \in (t^*, 1)$.

(b) *Class II*, if there can be one and at most one $t^{**} \in (0,1)$ such that $G(t) < 0$ for $t \in (0, t^{**})$, $G(t^{**}) = 0$, and $G(t) > 0$ for $t \in (t^{**}, 1)$.

(c) *Class III*, if there can be two and at most two $t \in (0,1)$ such that $G(t) = 0$, denoted by $t^{**} < t^*$, with $G(t) < 0$ for $t \in (0, t^{**}) \cup (t^*, 1)$ and $G(t) > 0$ for $t \in (t^{**}, t^*)$.

The Krugman and Helpman models are, respectively, Class I and II. The first two model classes in Definition 5 can be called, respectively, "Krugman-type" and "Helpman-type." Class III features the combined characteristics of Class I and II. Basically, the definition classifies models by the spatial scale(s) of dispersion force(s) in the models as shown in Table 1.

Example 3.4. Class I models include Krugman (1991b), Puga (1999), Forslid and Ottaviano (2003), Pflüger (2004), and Harris and Wilson (1978). ■

Example 3.5. Class II models include Helpman (1998), Murata and Thisse (2005), Redding and Sturm (2008), Allen and Arkolakis (2014), Redding and Rossi-Hansberg (2017) (§3), and Beckmann (1976). ■

Example 3.6. Class III models include Tabuchi (1998), Pflüger and Südekum (2008), as well as Takayama and Akamatsu (2011). ■

Definition 5 can be equivalently formulated in terms of net gain function $G^\sharp$, i.e., we can replace G with $G^\sharp$. Although Definition 5 does not place any restriction on the concrete functional form of G, net gain functions $G^\sharp$ for models in the literature, including those in the above examples, are usually simple quadratic functions with model-dependent coefficients $\{c_0, c_1, c_2\}$:

$$G^\sharp(t) = c_0 + c_1 t + c_2 t^2. \tag{19}$$

The net gain functions for the Krugman and Helpman models are shown in (16) and (17) in Section 3.1. Appendix D provides the derivations of $\{c_0, c_1, c_2\}$ for the models by Krugman (1991a), Helpman (1998), Pflüger and Südekum (2008), and Allen and Arkolakis (2014). See Akamatsu et al. (2019) for the derivations for the other models in the examples.

The coefficients $\{c_0, c_1, c_2\}$ in the quadratic representation (19) are some functions of the model parameters. A particular economic force in a model can impact several coefficients simultaneously.

Example 3.7. We recall that the Helpman model has the following net gain function:

$$G^\sharp(t) = \underbrace{-\gamma}_{c_0} + \underbrace{\left(\frac{\mu}{\sigma-1} + \frac{\mu}{\sigma}\right)}_{c_1} t \underbrace{- \left(\frac{\mu^2}{\sigma-1} + \frac{1}{\sigma} - \gamma\right)}_{c_2} t^2 = c_0 + c_1 t + c_2 t^2.$$

There are two model parameters, $\mu \in (0,1)$ and $\sigma > 1$, as $\gamma = 1 - \mu$. We observe that μ affects all the coefficients simultaneously, whereas σ affects both c_1 and c_2. Since μ is the expenditure share of manufactured goods, increasing μ strengthens agglomeration force in $c_1 > 0$, while it reduces the intensity of the local dispersion force $c_0 = -\gamma = -(1-\mu)$ at the same time. ■

Remark 3.1. If $G^\sharp$ takes the quadratic form (19), then we can translate Definition 5 to the following conditions on $\{c_0, c_1, c_2\}$. A model is Class I if and only if its parameters (e.g., μ and σ in the Krugman and Helpman models) can be chosen to satisfy $G^\sharp(1) = c_0 + c_1 + c_2 < 0$ and either $G^\sharp(0) = c_0 > 0$ or $G^\sharp(0) = c_0 = 0$ and $G^{\sharp\prime}(0) = c_1 > 0$. A model is Class II if and only if its parameters can be chosen to satisfy $G^\sharp(0) = c_0 < 0$ and either $G^\sharp(1) = c_0 + c_1 + c_2 > 0$ or $G^\sharp(1) = c_0 + c_1 + c_2 = 0$ and $G^{\sharp\prime}(1) = c_1 + 2c_2 < 0$. A model is Class III if and only if its parameters can be chosen to satisfy $G^\sharp(0) = c_0 < 0$, $G^{\sharp\prime}(0) = c_1 > 0$, $G^\sharp(1) = c_1 + c_2 + c_3 < 0$, $G^{\sharp\prime}(1) = c_1 + 2c_2 < 0$, and $c_1^2 - 4c_0c_2 > 0$; the last condition is the discriminant condition so that $G^\sharp$ has two roots in $(0,1)$.

For example, continue with Example 3.7. When μ and σ vary, $G^\sharp(t)$ has at most one root $t^* \in (0,1)$, and $G^\sharp(t) < 0$ if $t < t^*$ and $G^\sharp(t) > 0$ if $t > t^*$. To see this, observe that $G^\sharp(0) = c_0 = -\gamma < 0$ and $G^\sharp(1) = c_0 + c_1 + c_2 = \frac{(1-\mu)(1-(1-\mu)\sigma)}{\sigma(\sigma-1)}$. If $1 \leq (1-\mu)\sigma$, then no root for $G^\sharp$ exists in $(0,1)$, thereby $G^\sharp(t) < 0$ for all $t \in (0,1)$; that is, endogenous agglomeration cannot occur for this parameter range. If $1 > (1-\mu)\sigma$, then $G^\sharp(1) > 0$ and $G^\sharp$ has one and only one root in $(0,1)$. Thus, the model is Class II. ■

The coefficients $\{c_0, c_1, c_2\}$ are obtained by collecting the terms of $G^\sharp$ according to the order of t. By this rearrangement, we can focus on the "order" of endogenous economic forces in the model. In (19), c_0 summarizes all the forces that work *within* each location (whether negative or positive), which may be called the "zero-th order" spatial effects; c_1 summarizes the direct effects *between* regions (from one region to another), which can be interpreted as the "first order" spatial effects; c_2 summarizes further effects induced by the first-order effects between regions, which are the "second order" spatial effects (from region k to region j, and then from region j to region i).

A subtle point about Definition 5 is that it takes parameter dependence into account. The number of root of G in $(0,1)$ depends on the parameter value.

Example 3.8. The Beckmann model in Example 2.2 satisfies $G(t) = -\gamma + t$. If $\gamma \in (0,1)$, then G has one and only one root in $(0,1)$ when γ varies; if $\gamma \geq 1$ ($\gamma \leq 1$), then G has no root in $(0,1)$ and $\bar{x}$ is always stable (unstable). The Beckmann model is a Class II model according to Definition 5. However, the model does not produce endogenous spatial patterns if $\gamma \geq 1$. ■

To study endogenous agglomeration due to changes in the level of transport costs, we must exclude trivial cases where $\bar{x}$ is (un)stable for all levels of transport cost. We impose the following condition.

Assumption E (Endogenous agglomeration occurs)**.** Consider a Class I, II, or III canonical model with gain function G. Then, the model parameters are such that G has exactly one root (two roots) in $(0,1)$ if the model is Class I or II (Class III).

Example 3.9. For the Helpman model, Assumption E requires $1 > (1-\mu)\sigma$. For the Beckmann model, it requires $\gamma \in (0,1)$. For the Krugman model, it is the so-called no-black-hole condition (Fujita et al., 1999b), $1 < (1-\mu)\sigma$. ■

Also, Class III models may have some parametric ranges where G has only one root in $(0,1)$; Assumption E excludes such situations.

A more subtle point about Definition 5 is that it classifies canonical models on the basis of the spatial scale of the "effective" dispersion force(s). This is illustrated by Example 3.7. According to Definition 4, there are global dispersion forces in the Helpman model because $c_2 < 0$ under Assumption E. However, unlike the Krugman model (Class I), the global dispersion forces in the Helpman model does not stabilize the uniform distribution when transport cost is high. In the net sense, the model's global dispersion force is not effective. Consequently, the only effective dispersion force in the Helpman model is the local dispersion force represented by $c_0 < 0$. Table 1 classifies models in terms of the spatial scale(s) of their effective dispersion force(s).

Remark 3.2. Definition 5 identifies three prototypical model classes found in the literature. Models available in the literature usually have natural restrictions on the values of their parameters that unambiguously determine the model class each of them belongs to. Technically, however, there can be flexible models that can behave as both Classes I and II. Consider the following artificial example: $G^\sharp(t) = \alpha(-\gamma + t)$, where α and γ are the basic parameters. If such a model exists, the model is Class II if $\alpha > 0$, Class I if $\alpha < 0$, and indeterminate if there is no restriction on α. ■

4 How model class matters in the many-region economy

In the two-region economy, the differences in the spatial scales of the dispersion forces in the Krugman and Helpman models induce the "reversed scenario" regarding the timing of endogenous asymmetry. This is because a global (local) dispersion force is triggered when ϕ is low (high). In many-region economy, there is a contrast not only in timing but also in *endogenous spatial patterns*. A major watershed between "Krugman-like" and "Helpman-like" models exists in terms of endogenous spatial patterns. By considering endogenous agglomeration under Assumptions RE and S, this section provides **Proposition 1**, which categorizes endogenous spatial distributions that can emerge from the spatially uniform distribution in canonical models.

4.1 Endogenous agglomeration in a racetrack economy

Before studying how each model class behaves, we first explain endogenous agglomeration under Assumptions RE and S. The uniform pattern $\bar{\boldsymbol{x}} \equiv (\bar{x}, \bar{x}, \ldots, \bar{x})$ ($\bar{x} \equiv \frac{1}{N}$) is a spatial equilibrium. The question is what are the spatial patterns that emerge due to the destabilization of uniform distribution $\bar{\boldsymbol{x}}$ through economic forces represented by payoff function $\boldsymbol{v}$.

Consider an infinitesimal migration of agents $\boldsymbol{z} = (z_i)_{i\in\mathcal{N}}$ from $\bar{\boldsymbol{x}}$ so that the new spatial distribution becomes $\boldsymbol{x}' \equiv \bar{\boldsymbol{x}} + \boldsymbol{z}$. We have $\sum_{i\in\mathcal{N}} z_i = 0$, as the total mass of agents should not change. Analogous to the two-region case, the marginal gain for agents due to such a deviation can be evaluated by

$$\bar{\omega} \equiv \boldsymbol{z}^\top \mathbf{V} \boldsymbol{z}, \tag{20}$$

where $\mathbf{V} = \frac{\bar{x}}{\bar{v}} \nabla \boldsymbol{v}(\bar{\boldsymbol{x}})$ is the payoff elasticity matrix. Note that $\bar{\omega}$ is the first-order approximation of the elasticity of the average payoff because

$$\frac{\bar{x}}{\bar{v}} \left(\sum_{i\in\mathcal{N}} v_i(\boldsymbol{x}') x_i' - \sum_{i\in\mathcal{N}} v_i(\bar{\boldsymbol{x}}) \bar{x} \right) \approx \boldsymbol{z}^\top \mathbf{V} \boldsymbol{z}, \tag{21}$$

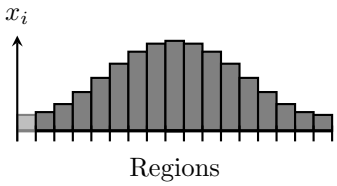
where $\bar{v}$ is the payoff level at $\bar{\boldsymbol{x}}$. If $\bar{\omega} < 0$ for *any* $\boldsymbol{z}$, then $\bar{\boldsymbol{x}}$ is stable.

Under Assumptions RE and S, there is a *model-independent* way to conveniently represent all possible migration patterns:

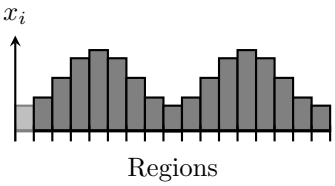
$$\boldsymbol{z} = \sum_k \zeta_k \boldsymbol{z}_k, \tag{22}$$

where $\{\boldsymbol{z}_k\}$ are the eigenvectors of $\mathbf{V}$ and $\{\zeta_k\}$ are their coefficients. We normalize $\|\boldsymbol{z}_k\|^2 = \boldsymbol{z}_k^\top \boldsymbol{z}_k = 1$ for all k. By (22), we interpret $\boldsymbol{z}$ as the weighted sum of the "basic" migration patterns $\{\boldsymbol{z}_k\}$. The basic migration patterns are model-independent in the sense that they are the eigenvectors of $\mathbf{V}$ irrespective of the properties of $\boldsymbol{v}(\boldsymbol{x})$ other than Assumption S.

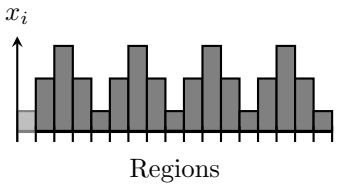
In the two-region economy, $\boldsymbol{z} = (1, -1)$ is the only possible migration pattern. In contrast, the many-region economy allows multiple possibilities. Each $\boldsymbol{z}_k$ is in itself a migration pattern and is a



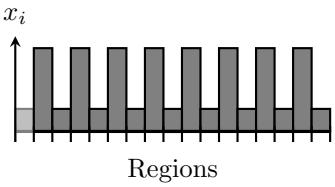
(a) A megacity ($\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_1$)



(b) Two major cities ($\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_2$)



(c) Four small cities ($\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_4$)



(d) Eight small cities ($\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_8$)

Figure 5: Schematic illustrations of spatial patterns induced by each migration pattern ($N = 16$). For expositional simplicity, we present the cases when k is a power of 2.

cosine curve with k equally spaced peaks. There are essentially $\frac{N}{2}$ basic migration patterns and the number of peaks is the largest when $k = \frac{N}{2}$ (see Appendix A.1). This is because the concentration of agents in every other region achieves the maximum number of symmetric peaks.[6]

Example 4.1. Figure 5 shows spatial patterns $\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_k$ ($k = 1, 2, 4, 8$) for $N = 16$ with a small $\zeta > 0$. If we regard each peak as the location of a city by analogy to the continuous-space mode by de Palma et al. (2019), then we may interpret that basic migration patterns $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, $\boldsymbol{z}_4$, and $\boldsymbol{z}_8$ express, respectively, the formation of a megacity (Figure 5a), two major cities (Figure 5b), four small cities (Figure 5c), and eight small cities (Figure 5d). ■

Let ω_k be the eigenvalue of $\mathbf{V}$ associated with $\boldsymbol{z}_k$, i.e., $\omega_k \boldsymbol{z}_k = \mathbf{V}\boldsymbol{z}_k$. Then, because $\boldsymbol{z}_k^\top \boldsymbol{z}_l = 0$ for $k \neq l$,

$$\bar{\omega} = \boldsymbol{z}^\top \mathbf{V} \boldsymbol{z} = \Big(\sum_k \zeta_k \boldsymbol{z}_k\Big)^\top \mathbf{V} \Big(\sum_k \zeta_k \boldsymbol{z}_k\Big) = \sum_k \zeta_k^2 \boldsymbol{z}_k^\top \mathbf{V} \boldsymbol{z}_k = \sum_k \zeta_k^2 \omega_k. \tag{23}$$

Thus, $\bar{\omega}$ is maximized by choosing the basic migration pattern that has the largest eigenvalue as $\boldsymbol{z}$:

$$\max_{\boldsymbol{z}} \frac{\bar{\omega}}{\|\boldsymbol{z}\|^2} = \omega_{\max} \quad \text{where} \quad \omega_{\max} \equiv \max_k \{\omega_k\}. \tag{24}$$

If $\omega_{\max} < 0$, then $\bar{\boldsymbol{x}}$ is stable; when $\omega_{\max}$ turns to positive, $\bar{\boldsymbol{x}}$ becomes unstable. Each ω_k is the gain from migration towards $\boldsymbol{z}_k$'s direction, and $\bar{\boldsymbol{x}}$ becomes unstable when some migration direction becomes profitable. The spatial pattern that emerge at the onset of instability is $\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_{\max}$, where $\boldsymbol{z}_{\max}$ is the basic migration pattern associated with $\omega_{\max}$. This extends the discussion based on ω and $\boldsymbol{z} = (1, -1)$ in the two-region case.

[6] Concretely, we can choose $\{\boldsymbol{z}_k\}$ as follows: $\boldsymbol{z}_k \propto (\cos(\frac{2\pi}{N} ki))_{i=1}^N$ for $k = 1, 2, \ldots, \frac{N}{2}$, and $\boldsymbol{z}_k \propto (\sin(\frac{2\pi}{N}(N-k)i))_{i=1}^N$ for $k = \frac{N}{2} + 1, \ldots, N$. In particular, we have $\boldsymbol{z}_{\frac{N}{2}} \propto (1, -1, 1, -1, \ldots, 1, -1)$.

We need the concrete formulae for eigenvalues $\{\omega_k\}$ of $\mathbf{V}$. For canonical models, we have $\mathbf{V} = G(\bar{\mathbf{D}})$ where $\bar{\mathbf{D}}$ is row-normalized proximity matrix and G some rational function. The two-region formula $\omega = G(\chi)$ generalizes as follows:

$$\omega_k = G(\chi_k) \qquad \forall k, \tag{25}$$

where χ_k is the eigenvalue of $\bar{\mathbf{D}}$ associated with $\boldsymbol{z}_k$.

Each χ_k is an index of the marginal increase of the average proximity among agents when the k-centric pattern $\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_k$ emerges. Further, χ_k decreases in number of peaks k. This is because the average proximity from one agent to other agents is the largest in a single-peaked (i.e., unimodal) pattern as in Figure 5a, while it decreases as the number of peaks in the spatial distribution increases. In fact,

$$\max_k\{\chi_k\} = \chi_1 \quad \text{and} \quad \min_k\{\chi_k\} = \chi_{\frac{N}{2}} \tag{26}$$

for any given value of $\phi \in (0,1)$ (Akamatsu et al., 2012), where we recall that the maximum possible number of symmetric peaks is $\frac{N}{2}$. Every χ_k takes value on $(0,1)$ and is a decreasing function of ϕ, reflecting that, irrespective of k, the proximity of an agent to others increases monotonically when the freeness of interregional interactions increases monotonically.

Note that $\{\boldsymbol{z}_k\}$ and $\{\chi_k\}$ are model independent. They summarize the properties of the symmetric racetrack geography but not those of the payoff function. Thus, the model-dependent properties are encapsulated by gain function G. That said, $\omega_{\max} = \max_k \omega_k = \max_k G(\chi_k)$ at the onset of instability depends crucially on the properties of $G(\chi)$. In other words, the model class mattes.

4.2 Qualitative difference in endogenous spatial patterns

The following proposition characterizes the endogenous spatial patterns that Class I, II, or III models engender when $\bar{\boldsymbol{x}}$ becomes unstable.

Proposition 1. *Consider a canonical model of either Class I, II, or III. Suppose Assumptions RE, S, and E.*

(a) If the model is of Class I, then there exists $\phi^* \in (0,1)$ such that $\bar{\boldsymbol{x}}$ is stable for all $\phi \in (0,\phi^*)$ and unstable for all $\phi \in (\phi^*,1)$; the instability of $\bar{\boldsymbol{x}}$ in ϕ^* leads to the formation of a multimodal pattern with $\frac{N}{2}$ peaks.

(b) If the model is of Class II, there exists $\phi^{**} \in (0,1)$ so that $\bar{\boldsymbol{x}}$ is stable for all $\phi \in (\phi^{**},1)$ and unstable for all $\phi \in (0,\phi^{**})$; the instability of $\bar{\boldsymbol{x}}$ in ϕ^{**} leads to the formation of a unimodal pattern.

(c) If the model is of Class III, there exist $\phi^*, \phi^{**} \in (0,1)$ with $\phi^* < \phi^{**}$ so that $\bar{\boldsymbol{x}}$ is stable for all $\phi \in (0,\phi^*) \cup (\phi^{**},1)$; the instabilities of $\bar{\boldsymbol{x}}$ at ϕ^* and ϕ^{**} lead to the formation of a multimodal pattern with $\frac{N}{2}$ peaks and a unimodal pattern, respectively.

Note that (a) and (b) generalizes the "reversed scenarios" of Krugman and Helpman in that $\bar{\boldsymbol{x}}$ is stable for the low (high) values of ϕ in Class I (II) models. The new phenomenon is that model classes qualitatively differ in the *number of peaks (modes)* they endogenously produce. Class I models engender

a multimodal distribution with $\frac{N}{2}$ peaks, which may interpreted as the formation of $\frac{N}{2}$ small cities; Class II models entail a unimodal distribution, which may be interpreted as the formation of a single city center or a monocentric city system. Class III is a synthesis of Classes I and II. If $N = 16$, the instability at ϕ^* in Classes I and III leads to the formation of eight-centric pattern (Figure 5d), whereas that at ϕ^{**} in Classes II and III leads to the formation of monocentric pattern (Figure 5a).

Proposition 1 is obtained by identifying $\omega_{\max}$ at the onset of instability and builds on the maximality of χ_1 and the minimality of $\chi_{\frac{N}{2}}$. The simplest example is the Beckmann model.

Example 4.2. The Beckmann model (Example 2.2) is Class II because $G(t) = G^\sharp(t) = -\gamma + t$ satisfies Definition 5 (b). Under Assumptions RE and S, at any level of ϕ, we have

$$\omega_{\max} = \max_k\{\omega_k\} = \max_k\{G(\chi_k)\} = \max_k\{-\gamma + \chi_k\} = -\gamma + \max_k\{\chi_k\} = -\gamma + \chi_1.$$

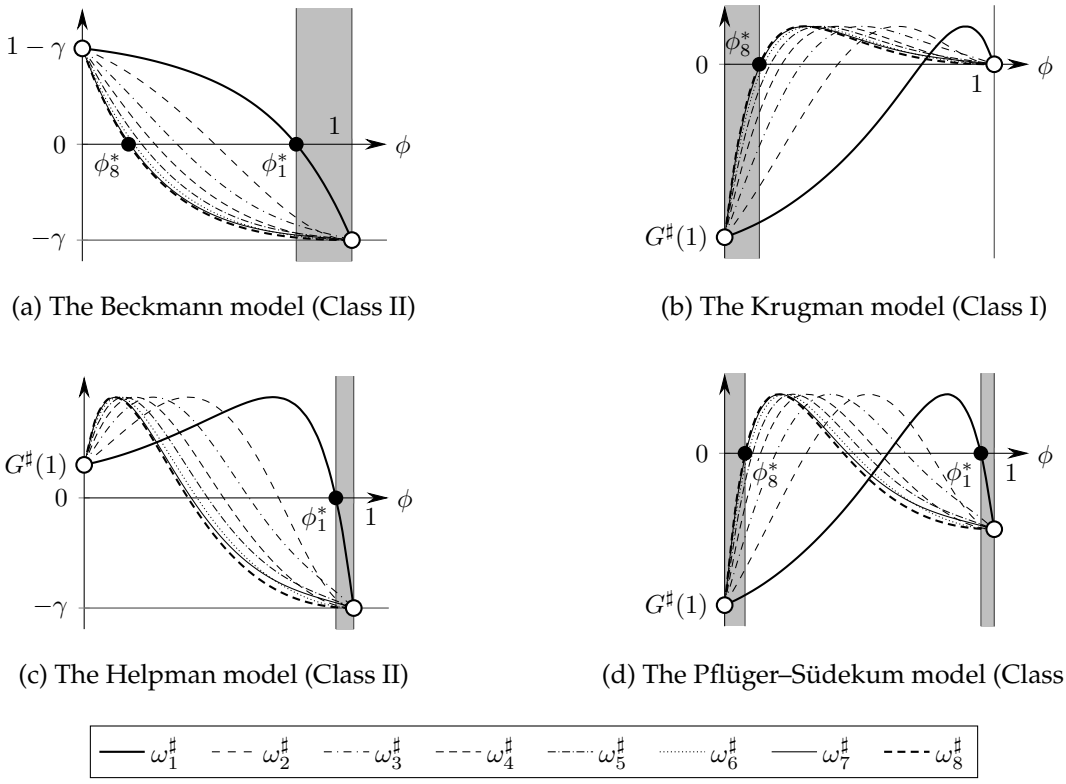
Thus, z_1 is the most profitable migration pattern for mobile agents at any given ϕ, and an instability of $\bar{x}$ leads to the formation of a unimodal distribution, i.e., a single city center (e.g., Figure 5a). The maximality of ω_1 is intuitive. In the model, the formation of a single large city center is the most beneficial outcome for every agent because agents prefer proximity to others, albeit agents must disperse around the city center to avoid local congestion effects. Because $\chi_1(\phi) \in (0,1)$ is a monotonically decreasing function of ϕ with $\chi_1(0) = 1$ and $\chi_1(1) = 1$, $\omega_{\max} = -\gamma + \chi_1 < 0$ for $\phi \in (\phi^{**}, 1)$ where $\phi^{**} = \phi_1^*$ is the unique solution for $\chi_1(\phi) = \gamma$, which exists under Assumption E.

These arguments can be simply understood by drawing the curves of $\{\omega_k^\sharp\}$ on the ϕ axis, where $\omega_k^\sharp(\phi) = G^\sharp(\chi_k(\phi))$. Figure 6a shows $\{\omega_k^\sharp\}$ for the Beckmann model under $N = 16$. When all the curves stay below ϕ axis, $\bar{x}$ is stable (the gray region); when one and only one $\omega_k^\sharp$ cuts the axis, then $\bar{x}$ deviates in z_k direction. As seen from the figure, $\omega_1^\sharp$ is the first to cut the axis. ■

By assuming $N = 16$, representative examples of general equilibrium models from the three classes are also shown in Figure 6.

Example 4.3. Figure 6b and Figure 6c respectively depict $\omega_k^\sharp$ for the Krugman and Helpman models as the leading examples of Classes I and II. We see from the figures that $\max\{\omega_k^\sharp\} = \omega_8^\sharp$ for the Krugman model, whereas $\max\{\omega_k^\sharp\} = \omega_1^\sharp$ for the Helpman model for all ϕ of interest, i.e., in the ranges of ϕ where $\bar{x}$ is stable (the gray regions). When $\bar{x}$ becomes unstable, $8 = \frac{16}{2}$ cities emerge for the former model, whereas a single city emerges for the latter. Figure 6d shows $\omega_k^\sharp$ for an instance of Class III, the Pflüger and Südekum (2008) model. For this model, there are two ranges of ϕ under which $\bar{x}$ is stable. With both local and global dispersion forces, the model behaves as a Class I (II) model at the lower (higher) extreme of ϕ. ■

There are two remarks on **Proposition 1**. First, it builds on *local analysis* around uniform distribution $\bar{x}$ and characterizes the first bifurcation from $\bar{x}$. It may be of interest whether we can formally investigate what happens thereafter. However, the proposition cannot be generalized without imposing more assumptions on the payoff function v. To obtain stronger results beyond **Proposition 1**, we have to either introduce intricate classifications for the properties of the higher-order differentials of v or



(a) The Beckmann model (Class II)

(b) The Krugman model (Class I)

(c) The Helpman model (Class II)

(d) The Pflüger–Südekum model (Class III)

$\omega_1^\sharp$ $\omega_2^\sharp$ $\omega_3^\sharp$ $\omega_4^\sharp$ $\omega_5^\sharp$ $\omega_6^\sharp$ $\omega_7^\sharp$ $\omega_8^\sharp$

Figure 6: Examples of $\omega_k^\sharp \equiv G^\sharp(\chi_k(\phi))$ when $N = 16$.

simply focus on a specific model.[7] In practice, however, **Proposition 1** provides sufficient insights for understanding the evolution of the spatial pattern afterwards. Section 5.2 will illustrate this point through a series of numerical examples.

Second, as we emphasized in Section 2.2, **Proposition 1** assumes a *complete geographical symmetry* to focus on purely endogenous interactions in the payoff function v. By Assumptions RE and S, we abstract away regional heterogeneities and geographical advantages. It is thus of interest to what extent or in what sense the implications of **Proposition 1** generalize to asymmetric cases, given that the latest quantitative spatial models incorporate flexible structures regarding interregional transport costs and differences in local characteristics. To address this issue, Section 6.1 will provide formal analyses of the effects of heterogeneous local characteristics. Also, Section 6.2 numerically studies the effects of exogenous geographical advantages caused by the existence of boundaries.

5 Illustrations

As a concrete illustration of **Proposition 1**, Section 5.1 revisits the Krugman and Helpman models, the leading instances of Classes I and II, in the $N = 4$ racetrack economy. To illustrate how **Proposition 1**

[7]For example, by focusing on specific Class I models, Akamatsu et al. (2012) and Osawa et al. (2017) provided detailed analysis of the successive bifurcations after the first bifurcation. For models of Classes II and III in the literature, feasibility of such analysis depends on the specification of v.

(a) A monocentric ecoomy ($\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_1$)

(b) A duocentric economy ($\bar{\boldsymbol{x}} + \zeta \boldsymbol{z}_2$)

Figure 7: Schematic illustrations of spatial patterns

allows us to understand the evolution of the spatial pattern in each model class, Section 5.2 numerically explores a specific example from each model class in the $N = 8$ racetrack economy.

5.1 Beyond the reversed scenarios

Let $N = 4$ under Assumptions RE and S. This is the simplest case in which different regions can have different neighbors. Example 2.1 shows $\mathbf{D}$ in this case.

Uniform distribution $\bar{\boldsymbol{x}} = (\bar{x}, \bar{x}, \bar{x}, \bar{x})$ with $\bar{x} \equiv \frac{1}{4}$ is a spatial equilibrium. As discussed, $\bar{\boldsymbol{x}}$ is stable if all the eigenvalues of $\mathbf{V} = \frac{\bar{x}}{\bar{v}} \nabla \boldsymbol{v}(\bar{\boldsymbol{x}})$ are negative. There are two ($= \frac{N}{2} = \frac{4}{2}$) eigenvalues of interest, which we denote by ω_1 and ω_2. Associated with them, there are two "basic" migration patterns:

$$\boldsymbol{z}_1 = \frac{1}{\sqrt{2}}(1, 0, -1, 0) \quad \text{and} \quad \boldsymbol{z}_2 = \frac{1}{2}(1, -1, 1, -1). \tag{27}$$

We see that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ do not include any model parameters (i.e., model independent).

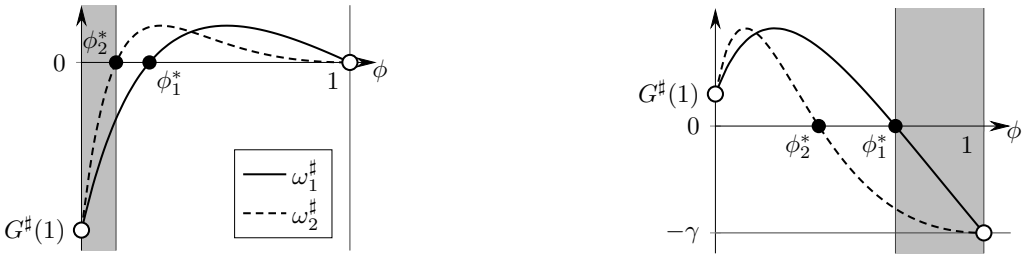
Figure 7 shows the schematics of the two outcomes. One represents the formation of a unimodal distribution, which may be interpreted the formation of a monocentric economy (Figure 7a). The other represents the formation of a multimodal distribution that features two small cities vying with each other (Figure 7b).

The question is which of the two patterns emerges endogenously in the Krugman and Helpman models. **Proposition 1** shows that the Krugman (Helpman) model produces a multimodal (unimodal) distribution. We ask whether $\omega_{\max} = \max\{\omega_1, \omega_2\} = \omega_1$ or $\omega_{\max} = \omega_2$ when $\bar{\boldsymbol{x}}$ becomes unstable, as we recall each ω_k represents the net agglomerative force towards $\boldsymbol{z}_k$'s direction.

Figure 8 provides the answer to the question. For each model, it depicts $\omega_1^\sharp \equiv G^\sharp(\chi_1(\phi))$ and $\omega_2^\sharp \equiv G^\sharp(\chi_2(\phi))$ on the ϕ axis, where $G^\sharp(\chi)$ is the same as in the two-region case for each model in (16) and (17) (Figure 3). As $\text{sgn}[\omega_k] = \text{sgn}[\omega_k^\sharp]$, $\bar{\boldsymbol{x}}$ is stable if the two curves stay below the horizontal axis (the shaded areas):

$$\bar{\boldsymbol{x}} \text{ is stable when } \begin{cases} \phi \in (0, \phi_2^*) & \text{(the Krugman model)}, \\ \phi \in (\phi_1^*, 1) & \text{(the Helpman model)}. \end{cases}$$

There is an analogy with the "reversed scenario" regarding *when* $\bar{\boldsymbol{x}}$ is stable. A sharp contrast is present



(a) The Krugman model

(b) The Helpman model

Figure 8: $\omega_1^\sharp$ and $\omega_2^\sharp$ for the Krugman and Helpman models.

in the *spatial patterns* that emerges from $\bar{\boldsymbol{x}}$ immediately after the bifurcation, which are, respectively,

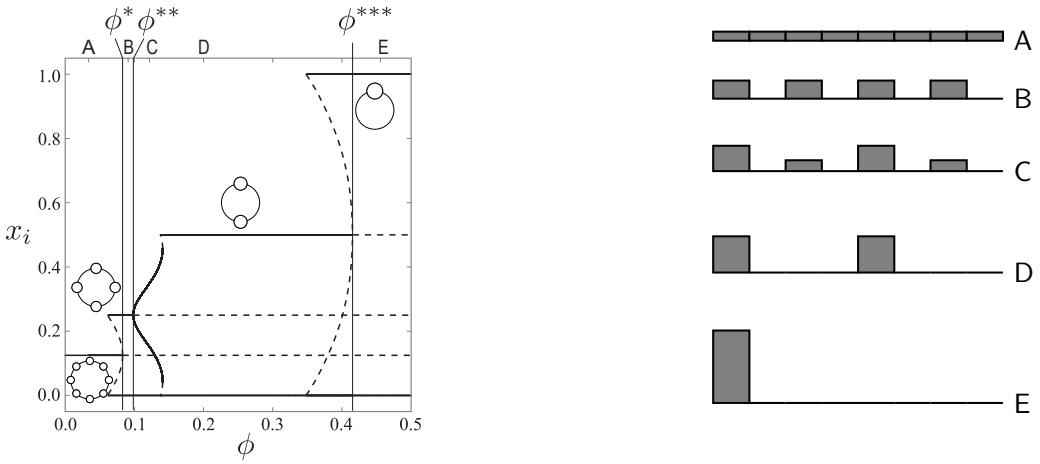
$$\begin{cases} \text{the multimodal pattern (Figure 7b)} & \text{(the Krugman model),} \\ \text{the unimodal pattern (Figure 7a)} & \text{(the Helpman model).} \end{cases}$$

These are examples for **Proposition 1** (a) and (b). For instance, we see $\omega_2 > \omega_1$ at ϕ_2^* for the Krugman model, because $\omega_2(\phi_2^*) = 0$ and $\omega_1(\phi_2^*) < 0$.[8]

We can understand the contrast as follows. Consider the Helpman model. As discussed in Section 3.2, the dispersion force in the Helpman model is local and triggered when ϕ is high. When ϕ is at its lower extreme ($\phi \approx 0$), agents concentrate in a single region. This is because the local dispersion force is less important than the benefits of agglomeration when interregional transportation is prohibitively costly. Mobile agents prefer concentrating towards a smaller number of regions because of the agglomeration force. In fact, in both the models, agents should result in a "black-hole" concentration in a single region if there is no effective dispersion force. The spatial pattern is close to a completely monopolar pattern, for example, $\boldsymbol{x} \approx (0, 1, 0, 0)$. As ϕ increases, the relative rise in the local dispersion force induces a crowding-out from the populated region to the nearby regions. The spatial pattern become $\boldsymbol{x} = (x', x, x', x'')$ with $x > x' > x''$, which can also be seen as a monocentric pattern. As ϕ increases, the spatial pattern gradually flattens and it connects to $\bar{\boldsymbol{x}}$ at ϕ_1^*. Thus, by starting from $\bar{\boldsymbol{x}}$, **Proposition 1** considers gradual *decrease* of ϕ to determine the dispersion process in a reverse-reproduced way. At ϕ_1^*, the spatial pattern must deviate in the direction of the "formation" of a unimodal pattern (Figure 7a).

By contrast, the dispersion force in the Krugman model is global and triggered when ϕ is low. The global dispersion force in the model stems from competition between firms over consumers' demand. When ϕ is low, there are few incentives for firms to concentrate in a small number of regions because the shipment of goods incurs large transport costs. As ϕ increases, the spatial extent to which a firm can supply goods increases. For a firm, this brings more opportunities to access a wider range of consumers but also leads to tougher competition with other firms that are geographically close. At some point, firms are better off forming small agglomerations, so that each agglomeration has its

[8]One should not predict stable spatial patterns *after* the bifurcation by the relative magnitudes of $\omega_1^\sharp$ and $\omega_2^\sharp$ because they are obtained by local analysis at $\bar{\boldsymbol{x}}$. Section 5.2 studies spatial patterns after the first bifurcation by numerical simulations.



(a) Bifurcation diagram

(b) Spatial patterns

Figure 9: Class I model (Krugman, 1991b)

dominant market area but is relatively remote from other major agglomerations of firms, as in the two-city pattern (Figure 7b).

The global dispersion force represents the repulsive effects across different locations and supports the formation of multiple cities, whereas the local dispersion force represents the crowding effects that induce the flattening of each city. In a many-region economy, these forces lead to the formation of qualitatively different spatial patterns. The contrast in spatial patterns is hidden in the two-region setup because the only possible migration pattern is $z = (1, -1)$.

5.2 Number, size, and spatial extent of cities

Proposition 1 builds on *local stability analysis* around uniform distribution $\bar{x}$. However, it provides essential insights for understanding the global evolutionary behavior of the spatial pattern in a circular economy after the first bifurcation. To highlight this point, we numerically simulate an overall evolutionary path of the spatial structure for selected models from Classes I, II, and III in the $N = 8$ racetrack economy. See Appendix D for the details of the models and the parameter settings.

5.2.1 Class I model

Figure 9 reports the evolutionary path of stable equilibrium patterns in the course of increasing ϕ for the Krugman model. In Figure 9a, the black solid (dashed) curves depict the stable (unstable) equilibrium values of x_i at each ϕ. Figure 9b shows the schematic illustration of the stable spatial pattern on the path. The letters in Figure 9b correspond to those in Figure 9a.

Consider a gradual increase in ϕ from $\phi \approx 0$. Uniform distribution $\bar{x}$ is initially stable until ϕ reaches the so-called "break point" ϕ^* where a *bifurcation* from $\bar{x}$ occurs. At ϕ^*, the spatial pattern is pushed towards the formation of $\frac{8}{2} = 4$ cities. This confirms **Proposition 1** (a). The spatial pattern immediately converges towards a four-cities pattern after ϕ^* is passed.

A further increase in ϕ triggers the second and third bifurcations at ϕ^{**} and ϕ^{***}, respectively. These bifurcations sequentially double the spacing between cities, each time halving their number, $4 \to 2 \to 1$, in a close analogy to the first bifurcation at ϕ^*. This is the "spatial period-doubling cascade" behavior discussed by Ikeda et al. (2012a); Akamatsu et al. (2012); Osawa et al. (2017); Ikeda et al. (2018a) using specific Class I models.[9] At the higher extreme of ϕ, a complete monopolar pattern emerges. This behavior can be understood as a gradual increase in the effective market area of each city due to a decline in transport costs. The spatial extent of each city is one regional unit at any level of ϕ because there exists no local dispersion force.

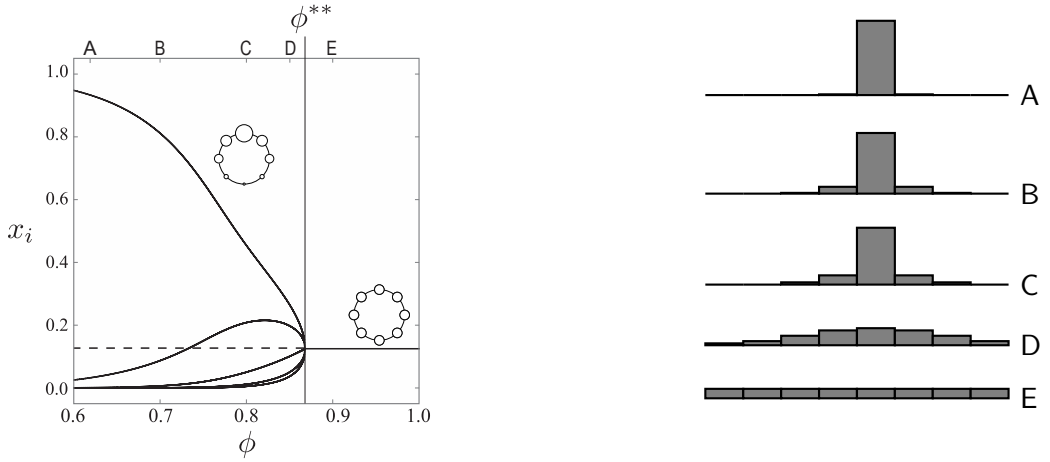
In the model, cities become larger when interregional access improves. However, such an effect is limited to the "selected" regions. Furthermore, as ϕ increases, once selected regions can decline to form an *agglomeration shadow* (Arthur, 1994; Fujita and Krugman, 1995) of other regions. For example, consider the fifth region from the left in Figure 9b. This region is selected at the transitions at ϕ^* and ϕ^{**}, that is, the impact of an increase in ϕ is positive. However, after ϕ^{***} is encountered, it immediately loses its population. For the region, a monotonic increase in ϕ implies a winning situation followed by a losing one. The global dispersion force in Class I models is thus related to the *rise and fall of major cities*. Class I models do not provide robust predictions for each city, but they do for the *overall* spatial distribution of cities: the number of cities and spacing between them monotonically decreases and increases, respectively, with the monotonic reduction in interregional transport costs.

Remark 5.1. The empirical evidence on regional agglomeration presented by Duranton and Turner (2012) and Faber (2014) is related to the predictions of Class I models. The former study focused on the growth of large metropolitan areas in the United States, while the latter analyzed the growth of peripheral counties in China. The former (latter) study revealed a positive (negative) correlation between the magnitude of growth and the interregional transportation infrastructure level of a given region. For Class I models, these opposite responses may simply reflect different sides of the same coin. That is, both results may indicate a tendency of selective concentration towards larger regions for an improvement in interregional transportation access (as discussed in the Introduction for Japan). ■

5.2.2 Class II model

Next, Figure 10 shows the results for a Class II model, namely, the model by Allen and Arkolakis (2014) (see Section D.2.4). This model incorporates a local dispersion force but no global dispersion force; $\bar{x}$ is stable for higher values of ϕ. As in Section 5.1, we see the evolutionary process in a reverse-reproduced way, that is, in the course of a monotonic *decrease* in ϕ. The bifurcation at ϕ^{**} leads to the "emergence" of a unimodal pattern. This is *the* bifurcation in the model: when ϕ decreases further, the spatial pattern monotonically and smoothly converges to a complete concentration in a single region. We interpret a region that locally maximizes population size (region i such that $x_i > x_{i-1}$ and $x_i > x_{i+1}$ where mod N for indices) as the location of a city. Then, this model endogenously produces at most one city. Class II

[9] Akamatsu et al. (2012) and Osawa et al. (2017) study, respectively, the models by Pflüger (2004) and Harris and Wilson (1978) using the racetrack assumption. As these papers demonstrate, for Class I models in the literature, we can formally study the bifurcations at ϕ^{**} and ϕ^{***} if we assume a specific functional form for v.



(a) Bifurcation diagram

(b) Spatial patterns

Figure 10: Class II model (Allen and Arkolakis, 2014)

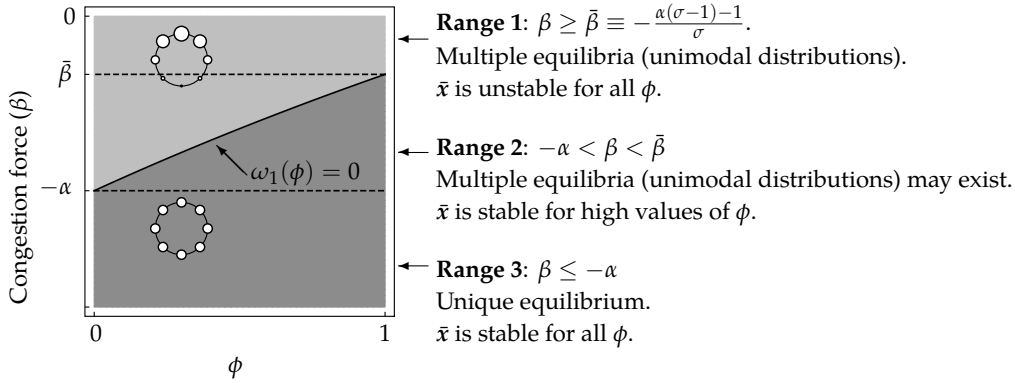
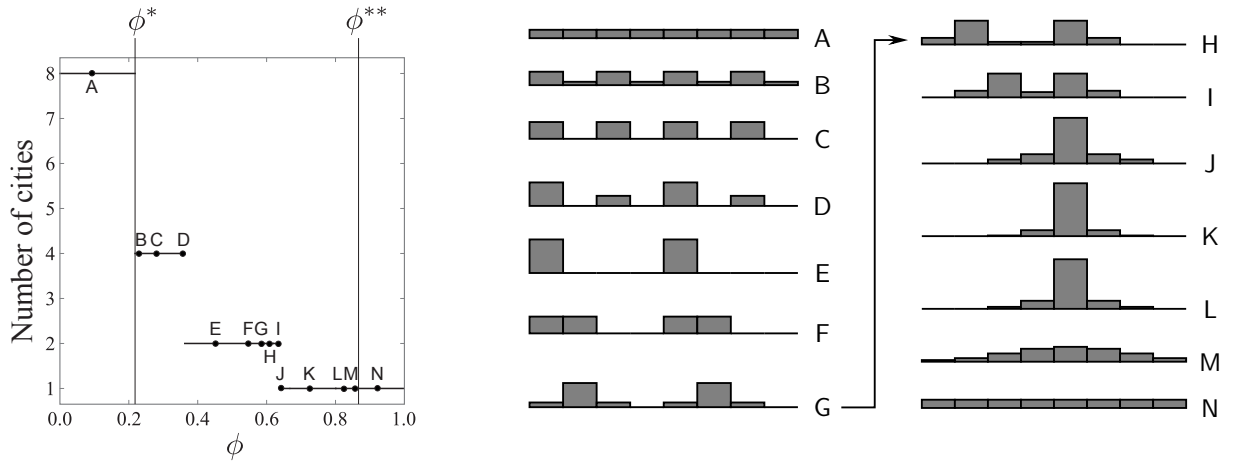


Figure 11: Uniqueness and stability of equilibria in the Allen–Arkolakis model

models would be interpreted as expressing the flattening of a single big city during improvement in interregional transportation access.

Remark 5.2. In Class II models, we can ensure the uniqueness of the spatial equilibrium *regardless of the level of interregional transport costs* by imposing a strong local dispersion force (e.g., Redding and Sturm, 2008; Allen and Arkolakis, 2014). Example 2.2 provides a prototypical situation in which a strong congestion force suppress the possibility of endogenous asymmetry (i.e., the $\gamma \geq 1$ case). The uniqueness of equilibrium implies the stability of $\bar{x}$ for all ϕ in a racetrack economy; no endogenous asymmetry can emerge, since $\bar{x}$ is always an equilibrium. Figure 11 provides our classification of possible spatial patterns and their stability for the Allen–Arkolakis model in a racetrack economy. The uniqueness condition for the Allen–Arkolakis model is $\beta \leq -\alpha$ (i.e., Range 3 in the figure), which is a sufficient condition for the stability of $\bar{x}$. Figure 11 can be seen as a refinement of Figure I of Allen and Arkolakis (2014) that shows endogenous spatial patterns in a circular domain. ■



(a) Number of cities

(b) Spatial patterns

Figure 12: Class III model (Pflüger and Südekum, 2008)

5.2.3 Class III model

Finally, we consider a Class III model. Because both local and global dispersion forces exist, this class of models exhibits a rich and realistic interplay between the *number of cities, spacing between them* (as in Class I models), and the *spatial extent of each city* (as in Class II models).

Figure 12a shows the evolution of the number of cities in the course of increasing ϕ under the Pflüger and Südekum (2008)'s model (see Section D.2.3). The number of cities in a spatial distribution is defined by that of the local maxima therein. Figure 12a exhibits the mixed characteristics of Figures 9 and 10, as expected. When $\phi < \phi^*$ or $\phi > \phi^{**}$, $\bar{x}$ is stable. We interpret the number of cities in $\bar{x}$ as either 8 (for a low ϕ) or as 1 (for a high ϕ) to acknowledge that $\bar{x}$ at the low and high levels of ϕ are distinct. When ϕ gradually increases from $\phi \approx 0$, the number of cities reduces from $8 \to 4 \to 2 \to 1$ as in the Class I models (Figure 9), whereas it is always 1 in the latter stage as per the Class II models (Figure 10). The initial stage is governed by a decline in the global dispersion force, while the later stage is marked by a relative rise of the local dispersion force.

Figure 12b shows the spatial patterns associated with Figure 12a. Uniform pattern $\bar{x}$ is initially stable (Pattern A) and the first bifurcation at ϕ^* leads to a four-city pattern (B, C), whereas the second bifurcation to the formation of two cities (D, E). These transitions are in line with Figure 9 and are governed by the gradual decline in the global dispersion force. A further decline in the global dispersion force increases the relative importance of the local dispersion force. As a result, the two cities in Pattern E gradually increase their spatial extents (F, G). When ϕ increases further, the two cities gradually merge (H, I) to form a megacity (J, K). Finally, a gradual flattening of the single megacity occurs (L, M), followed by complete dispersion (N) after ϕ^{**} is reached.

Remark 5.3. The behavior of Class II and III models is related to empirical evidence for the flattening of once established economic clusters (i.e., cities) as a consequence of improved *interregional* access. Baum-Snow (2007) and Baum-Snow et al. (2017) presented evidence for US metro areas during 1950–

1990 and Chinese prefectures during 1990–2010, respectively. These studies addressed the changes in the population or production size of the central area within the larger region, both reporting a significantly negative effect of improvements in interregional access. As discussed in these studies, the local flattening of cities can also be interpreted as suburbanization in response to the improved *intraurban* transportation infrastructure in classical urban economic theory (e.g., Alonso, 1964). ■

Remark 5.4. In our analysis, the payoff function is *homogeneous* across mobile agents. The effects of introducing idiosyncratic payoff shocks are of interest, since it is a standard recipe in quantitative models (Redding and Rossi-Hansberg, 2017). For example, for additive random utility models,[10] the idiosyncratic payoff function is defined by $\hat{v}_{ni}(\boldsymbol{x}) = \tilde{v}_i(\boldsymbol{x}) + \epsilon_{ni}$ where $\tilde{v}_i(\boldsymbol{x})$ is the homogeneous component of payoff and ϵ_{ni} is an independent and identically distributed (i.i.d.) random payoff shock for an individual agent n for choosing region i. A spatial equilibrium is defined by $x_i = P_i(\boldsymbol{x})$ where $P_i(\boldsymbol{x}) \equiv \Pr\left(i = \arg\max_{j \in \mathcal{N}} \hat{v}_{nj}(\boldsymbol{x})\right) \in (0,1)$ is the probability for an agent to choose region i when the current spatial distribution is $\boldsymbol{x}$.

Such idiosyncratic heterogeneity acts as a *local dispersion force* and thus is related to Classes II and III. It is a well-known fact that random utility models can be represented on the basis of deterministic utility (Anderson et al., 1992; Hofbauer and Sandholm, 2002). There is a *deterministic* (or homogeneous) payoff function $\boldsymbol{v}(\boldsymbol{x}) = (v_i(\boldsymbol{x}))_{i \in \mathcal{N}}$, associated with the *stochastic* (or heterogeneous) payoff function $(\hat{v}_{ni}(\boldsymbol{x}))$, such that $\boldsymbol{x}^*$ is a deterministic spatial equilibrium under $\boldsymbol{v}(\boldsymbol{x})$ if and only if it satisfies $\boldsymbol{x}^* = \mathbf{P}(\boldsymbol{x}^*)$. Thus, the two spatial equilibrium concepts are "isomorphic" in terms of equilibrium spatial distribution of agents.[11] Hofbauer and Sandholm (2007) calls the deterministic version $\boldsymbol{v}$ the "virtual payoff"; their Appendix A explains that, for additive random utility models, we have $\boldsymbol{v}(\boldsymbol{x}) = \tilde{\boldsymbol{v}}(\boldsymbol{x}) - \nabla H(\boldsymbol{x})$ where H is an "admissible perturbation" that encapsulates the effects of considering payoff shocks. For example, when ϵ_{ni} is Gumbel-distributed, we can choose $H(\boldsymbol{x}) = \eta \sum_{i \in \mathcal{N}} x_i \log x_i$ where η is a constant that is proportional to the dispersion parameter for the distribution ϵ_{ni} are drawn from. This leads to $\boldsymbol{v}(\boldsymbol{x}) = \tilde{\boldsymbol{v}}(\boldsymbol{x}) - \eta \log \boldsymbol{x}$, where log is applied elementwise, and we have $\mathbf{V} \propto \tilde{\mathbf{V}} - \eta \mathbf{I}$, where $\mathbf{V}$ and $\tilde{\mathbf{V}}$ are, respectively, the elasticity matrix for $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$ evaluated at $\bar{\boldsymbol{x}}$. If G and $\tilde{G}$ are, respectively, the gain functions for $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$, then we have

$$G(t) \propto \tilde{G}(t) - \eta. \tag{28}$$

That is, the introduction of idiosyncratic payoff shocks is equivalent to adding a negative constant term (i.e., local dispersion force) to the gain function of the common component $\tilde{\boldsymbol{v}}$. This is a natural consequence of assuming that ϵ_{ni} is i.i.d. over n and i. The idiosyncratic payoff shock should act as some kind of dispersion force, but it has no connection to the underlying geography. As such, introducing idiosyncratic payoff shocks to a Class I model can change the model to Class III. ■

[10] Multiplicative random utility models are obtained by simple exponential transformation.

[11] See Behrens and Murata (2018) for a recent application of this fact in the context of spatial economic models.

6 Extensions

Proposition 1 builds on a complete geographical symmetry. Assumptions RE and S abstract away, respectively, geographical advantages of regions due to their relative position in the transport network and region-fixed advantages due to differences in local characteristics. Exogenous asymmetries are inherent in the real-world geography. This section illustrates the utility of distinguishing local and global dispersion forces and the model classes in asymmetric settings. Section 6.1 studies the marginal roles of local regional characteristics by relaxing Assumption S. Section 6.2 numerically explores the effects of exogenous geographic asymmetry in the presence of boundaries by dropping Assumption RE.

6.1 Exogenous regional characteristics

This section studies the sensitivity of spatial patterns to regional characteristics such as local amenities and productivity differences. We show that the spatial scale of dispersion force(s) in a model tends to determine whether the effects of exogenous advantages are amplified when transport cost varies. Throughout, we assume Assumption RE.

Let $\boldsymbol{a} = (a_i)_{i\in\mathcal{N}}$ with $a_i > 0$ be some exogenous regional characteristic, which may or may not affect the payoffs in other regions; the payoff function may be written as $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{a})$. For example, a_i may be the level of amenities exclusively enjoyed by region i's residents or region i's total factor productivity. In the latter case, interregional trade flows and the resulting payoffs in other regions can depend on a_i.

The regions are symmetric if $\boldsymbol{a} = \bar{\boldsymbol{a}} \equiv (\bar{a}, \bar{a}, \ldots, \bar{a})$, for some $\bar{a} > 0$. Therefore, pair $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})$ is an equilibrium. Consider a variation in the local characteristic so that $\boldsymbol{a} = \bar{\boldsymbol{a}} + \boldsymbol{\epsilon}$ with small $\boldsymbol{\epsilon} = (\epsilon_i)_{i\in\mathcal{N}}$. Then, there is a new equilibrium, say $\boldsymbol{x}(\boldsymbol{a})$, which is close to $\bar{\boldsymbol{x}}$. The "covariance" between region i's relative (dis)advantage $\epsilon_i = a_i - \bar{a}$ and the relative deviation of its population $x_i(\boldsymbol{a}) - \bar{x}$ is given by:

$$\rho \equiv (\boldsymbol{a} - \bar{\boldsymbol{a}})^\top (\boldsymbol{x}(\boldsymbol{a}) - \bar{\boldsymbol{x}}) = \sum_{i\in\mathcal{N}} (a_i - \bar{a}) (x_i(\boldsymbol{a}) - \bar{x}) . \tag{29}$$

We assume that $\bar{\boldsymbol{x}}$ is stable, since otherwise considering $\boldsymbol{x}(\boldsymbol{a})$ is nonsensical.

We expect $\rho > 0$ if regional characteristic $\boldsymbol{a}$ acts positively in the payoff for agents, that is, if $\boldsymbol{a}$ is "advantageous." To formalize this intuition, we focus on a class of local characteristics which encompass various standard specifications in the literature. Let $\mathbf{A} \equiv \frac{\bar{a}}{\bar{v}}[\frac{\partial v_i}{\partial a_j}]$ be the elasticity matrix of the payoff regarding $\boldsymbol{a}$, evaluated at $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})$. Analogous to Definition 1, we suppose the following.

Assumption A. For the local characteristic $\boldsymbol{a}$ under consideration, there exists a rational function $G^\natural$ that is continuous over $[0, 1]$, positive whenever $\bar{\boldsymbol{x}}$ is stable, and satisfies $\mathbf{A} = G^\natural(\bar{\mathbf{D}})$.

For each model in Examples 3.4, 3.5, and 3.6, there is $G^\natural$ that satisfies the hypotheses of Assumption A for each natural choice of a local characteristic vector. The simplest example is *heterogeneous local amenity*.

Example 6.1. Assume that the payoff function takes the form $v_i(\boldsymbol{x}, \boldsymbol{a}) = a_i v_i(\boldsymbol{x})$, where $a_i > 0$ is the exogenous level of regional amenities and $\boldsymbol{v}(\boldsymbol{x}) = (v_i(\boldsymbol{x}))_{i\in\mathcal{N}}$ is the homogeneous component of the payoff function that satisfies Assumption S. Then, $\mathbf{A} = \frac{\bar{a}}{\bar{v}}\bar{v}\mathbf{I} = \bar{a}\mathbf{I}$ and, thus, $G^\natural(t) = \bar{a} > 0$. ■

Analogous to gain function G of a model, $G^\natural$ encodes the effect of the marginal changes in local characteristics $\boldsymbol{a}$ on the regional payoffs $\boldsymbol{v}$. Particularly, condition $G^\natural(t) > 0$ implies $\rho > 0$.

Example 6.2. Consider a symmetric two-region economy as in Section 2.3. Whenever $\bar{\boldsymbol{x}}$ is stable,

$$\rho = c\delta(\chi) \quad \text{where} \quad \delta(t) \equiv -\frac{G^\natural(t)}{G(t)} \quad \text{and} \quad \chi = \frac{1-\phi}{1+\phi} \tag{30}$$

with some $c > 0$. If $\bar{\boldsymbol{x}}$ is stable, $\omega = G(\chi) < 0$. Thus, $\rho > 0$ if $G^\natural(\chi) > 0$ for all χ with $\omega = G(\chi) < 0$.

To show (30), observe first that the payoff gain due to exogenous advantage in region 1 can be summarized by the following elasticity:

$$\alpha \equiv \frac{\bar{a}}{\bar{v}} \left(\frac{\partial v_1(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})}{\partial a_1} - \frac{\partial v_2(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})}{\partial a_1} \right) \tag{31}$$

where α is the eigenvalue of $\mathbf{A}$ associated with $\boldsymbol{z} = (1, -1)$. Under Assumption A, we have

$$\alpha = G^\natural(\chi) > 0. \tag{32}$$

Suppose that $\bar{\boldsymbol{x}} = (\bar{x}, \bar{x})$ is perturbed to $\boldsymbol{x} = (\bar{x} + \xi, \bar{x} - \xi)$ due to an exogenous regional asymmetry of the form $\boldsymbol{a} = (\bar{a} + \epsilon, \bar{a} - \epsilon)$ with some scalars ξ and ϵ. For $\boldsymbol{x}$ to be an equilibrium, $v_1(\boldsymbol{x}) = v_2(\boldsymbol{x})$ must hold true. Thus, ξ and ϵ should cancel out two forces, namely, gain (or *loss*, since we assume $\bar{\boldsymbol{x}}$ is stable) $\omega = G(\chi) < 0$ from endogenous migration and gain $\alpha > 0$ from exogenous asymmetry:

$$\omega\xi + \alpha\epsilon = 0 \quad \Rightarrow \quad \zeta = -\frac{\alpha}{\omega}\epsilon = -\frac{G^\natural(\chi)}{G(\chi)}\epsilon. \tag{33}$$

We obtain (30) with $c = 2\epsilon^2$, because $\rho = \epsilon\xi + (-\epsilon)(-\xi) = 2\epsilon\xi$ by definition. The fraction $\delta(\chi) = \frac{G^\natural(\chi)}{|G(\chi)|} = \frac{\alpha}{|\omega|}$ compares the magnitudes of gain from marginal exogenous advantage and of loss from marginal endogenous migration, under the condition that the economy stays in equilibrium. ■

An important question is: *does ρ increase or decrease when ϕ increases?* In other words, does improved transportation access strengthen (weaken) the role of local characteristics and what are the responses of the spatial distribution of economic activities to an improvement in interregional access if $\boldsymbol{a}$ is fixed? The response of ρ is a prototypical version of the questions asked in counterfactual exercises employing calibrated quantitative spatial economic models (see, e.g., Redding and Rossi-Hansberg, 2017).

We characterize the marginal response of ρ to changes in ϕ as follows.

Proposition 2. *Suppose Assumption RE. Consider a canonical model with gain function G. Take local characteristic $\boldsymbol{a}$ that satisfies Assumption A with some $G^\natural$. Assume that $\bar{\boldsymbol{x}}$ is stable and define $\delta(t) = -\frac{G^\natural(t)}{G(t)}$. Then, the following hold true for ρ in (29):*

(a) $\rho'(\phi) > 0$, if $\delta'(t) < 0$ for all $t \in (0,1)$ such that $G(t) < 0$.

(b) $\rho'(\phi) < 0$, if $\delta'(t) > 0$ for all $t \in (0,1)$ such that $G(t) < 0$.

Obviously, the impacts of improved interregional access are model dependent. However, *model class matters*—the response of ρ under a given model may be inferred by the spatial scale of the dispersion force in the model.

Consider heterogeneous local amenity (Example 6.1), which is the simplest specification of $\boldsymbol{a}$. For this case, there is a clear contrast between the Krugman model (Class I) and Helpman model (Class II).

Example 6.3 (Example 6.1, continued)**.** Suppose $G^\natural(t) = \bar{a} > 0$. Then, $\operatorname{sgn}[\delta'(t)] = \operatorname{sgn}[\frac{\bar{a}G'(t)}{G(t)^2}] = \operatorname{sgn}[G'(t)]$. The Krugman model (Example 3.1) satisfies $G'(t) < 0$ whenever $G(t) < 0$ (cf. Figure 3a) and thus $\rho'(\phi) > 0$. On the other hand, the Helpman model (Example 3.2) satisfies $G'(t) > 0$ whenever $G(t) < 0$ (cf. Figure 3b) and thus $\rho'(\phi) < 0$ whenever the equilibrium is unique (as assumed in quantitative applications). See Appendix E for numerical illustrations. ■

To see why such contrast emerges, consider a two-region setup in Example 6.2.

Example 6.4 (Example 6.2, continued)**.** For $N = 2$, we have $\rho(\phi) = c\delta(\chi(\phi))$ with $\chi = \frac{1-\phi}{1+\phi}$. We have $\rho'(\phi) = c\delta'(\chi(\phi))\chi'(\phi)$ where $\chi'(\phi) < 0$. For example, suppose that $\delta'(\chi) < 0$. We observe that, in general, $\delta'(\chi) < 0$ means either gain $\alpha = G^\natural(\chi)$ from the exogenous regional asymmetry decreases in χ or the magnitude of loss $|\omega| = |G(\chi)|$ from the endogenous migration increases in χ. This is satisfied in Class I models. In Class I models, $|\omega|$ decreases in ϕ (i.e., increases in χ) as long as $\bar{\boldsymbol{x}}$ is stable, because ω approaches to zero at the bifurcation point. When $G^\natural$ is constant, then we expect $\delta'(t) < 0$ in Class I models because δ is inversely proportional to $|\omega|$. A similar discussion applies to Class II models, and we can expect that Class II models satisfy $\delta'(t) > 0$. ■

The contrast between Classes I and II generalizes to the regional characteristics that indirectly affect payoffs in the other regions through general equilibrium. For such cases, $G^\natural(\chi)$ become non-constant.

Example 6.5. Redding and Rossi-Hansberg (2017), §3, studies a Class II model, namely the Helpman model with the following market equilibrium condition:

$$w_i x_i = \sum_{j \in \mathcal{N}} \frac{x_i a_i w_i^{1-\sigma} \phi_{ij}}{\sum_{k \in \mathcal{N}} x_k a_k w_k^{1-\sigma} \phi_{kj}} w_j x_j \qquad \forall i \in \mathcal{N}, \tag{34}$$

where $\boldsymbol{a}$ corresponds to *productivity differences*. We can show that $G^\natural(t) > 0$ for all $t \in (0,1)$. When the equilibrium is unique, $\delta'(t) > 0$ for all $t \in (0,1)$ and thus $\rho'(\phi) < 0$. ■

Example 6.6. The Krugman model (Example 3.1) is of Class I. Immobile demand $\boldsymbol{l} = (l_i)_{i \in \mathcal{N}}$ in the model can be regarded as regional characteristic. We have $G^\natural(t) > 0$ if $t \in (0,1)$. Also, $\delta'(t) < 0$ for all $t \in (0,1)$, so that $\rho'(\phi) > 0$, if $\bar{\boldsymbol{x}}$ is stable. ■

Examples 6.3, 6.5, and 6.6 demonstrate that the class a model belongs to can govern whether the endogenous causation of the model boosts the exogenous advantages when interregional transport costs decrease. When interregional access improves, the endogenous mechanisms of a model strengthens (weakens) the effects of exogenous local advantages if the model has only a global (local) dispersion

force. If exogenous heterogeneity causes one region to attract more population, such effects will be magnified (reduced) for Class I (II) models. Analogous to **Proposition 1**, **Proposition 2** builds on sensitivity analysis at $\bar{x}$. Appendix E provides numerical examples to illustrate the difference between Class I and II in terms of the response of ρ.

The qualitative differences between Classes I and II can be understood from the basic properties of the local and global dispersion forces in Section 3.2. For a Class I model, a larger ϕ means a relatively smaller global dispersion force, which tends to amplify (both the endogenous and exogenous) location-specific advantages towards the concentration of mobile agents. However, in a Class II model, a larger ϕ means a relatively larger local dispersion force, which reduces not only the benefit from concentration due to endogenous agglomeration externalities but also that due to location-specific exogenous advantages.

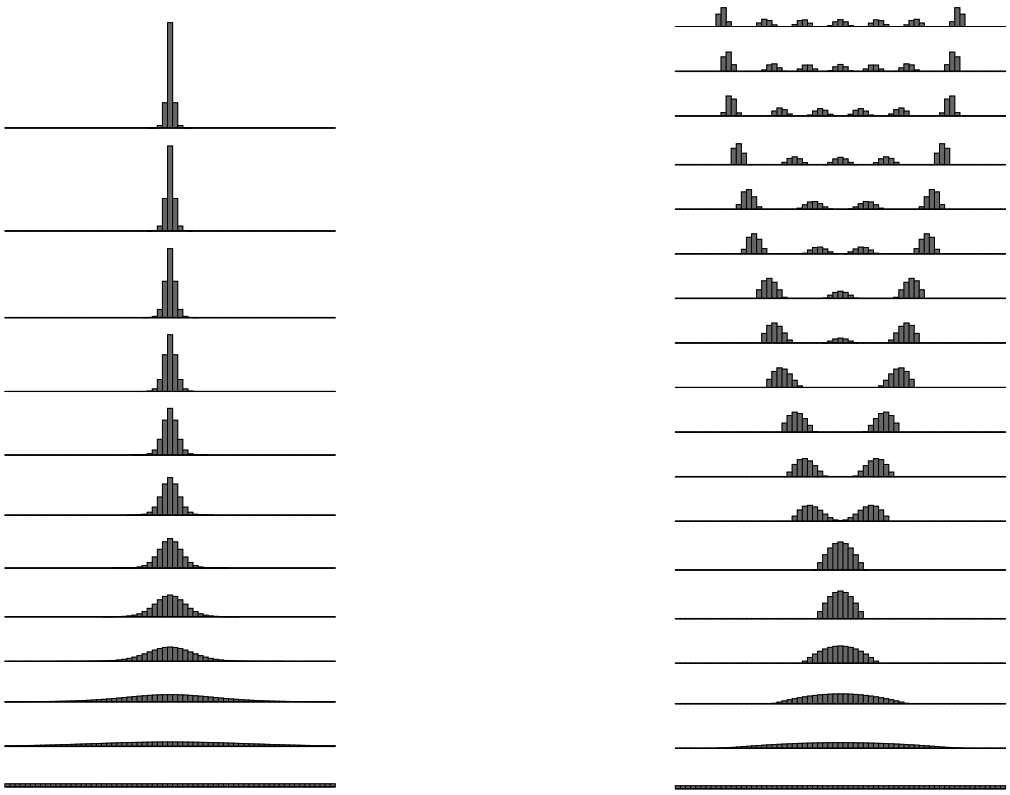
6.2 Shape of the transport network

This section illustrates that the implications of **Proposition 1** qualitatively generalize to various geographical setups (one-dimensional line segment, two-dimensional spaces with/without boundaries). In particular, the *polarity* of endogenous spatial patterns in each model class is unaffected; multiple cities can endogenously emerge from Class I, whereas a unimodal distribution emerges from Class II, even when we relax Assumption RE.

The simplest way to introduce geographical asymmetry into our one-dimensional setup is to consider a bounded line segment, which is a standard spatial setting in urban economic theory. Ikeda et al. (2017a) considered a Class I model (Forslid and Ottaviano, 2003) in a line segment. The paper showed that multiple cities emerges as in the racetrack economy, and demonstrated that the evolution of spatial structure in a line segment approximately follows the "period doubling" behavior discussed in Section 5.2.1. For Class II and III, Figure 13 shows examples of endogenous agglomeration patterns in the Helpman and Pflüger–Südekum models. For both the models, qualitative properties of the spatial patterns are consistent with those discussed in Section 5.2.

The real-world geography is two-dimensional. The two-dimensional counterpart of the racetrack economy is bounded lattices with periodic boundary conditions (i.e., flat torus), for which a basic theory of spatial agglomeration is provided in Ikeda and Murota (2014). For Class I models, it is now understood that they engender multiple disjointed cities and period-doubling behavior emerge from the model, as in the racetrack setup studied in this paper.[12] As concrete examples, Figure 14 shows endogenous equilibrium spatial patterns over a bounded square economy with $9 \times 9 = 81$ regions in the course of increasing ϕ for the Krugman and Allen–Arkolakis models. The parameters are the same as Figure 9 and Figure 10. The Krugman model (Class I) engender multiple disjointed cities. When ϕ increases, the number of cities gradually decreases, while the spacing between them enlarges. For the AA model (Class II), in contrast, the spatial pattern is initially monocentric, i.e., there is a single

[12] Ikeda et al. (2012b), Ikeda et al. (2014), and Ikeda et al. (2017b) analyze agglomeration patterns that emerge from a Class I model (Forslid and Ottaviano, 2003) in triangular lattice economy. Ikeda et al. (2018b) shows that agglomeration behavior of the same model in two-dimensional square lattice economy can be understood by analogy to the racetrack economy.



(a) The Helpman model (Class II)

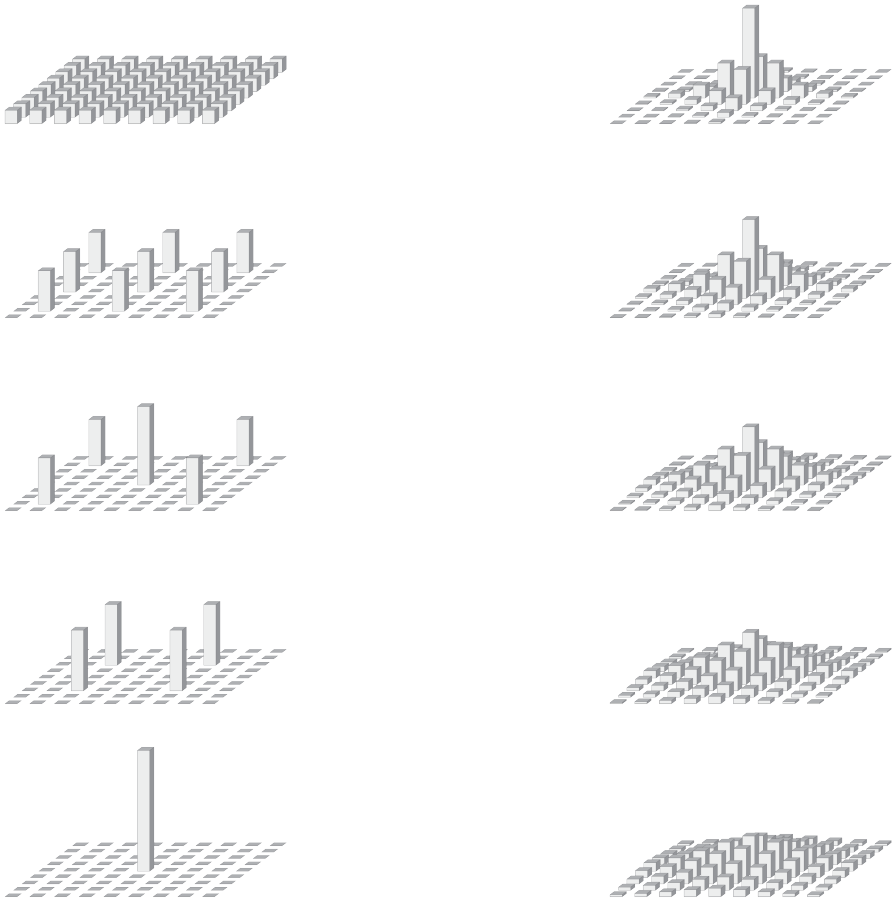
(b) The Pflüger–Südekum model (Class III)

Figure 13: Stable spatial patterns for Classes II and III in a line segment (number of regions $N = 65$). All regions share the same local parameters. The transport cost between every consecutive pair of regions is uniform. The level of transport cost monotonically decreases from top to bottom. For an extensive discussion on the behavior of a Class I model in a line segment, see Ikeda et al. (2017a).

big city. As ϕ increases, the city gradually flattens to exhibit suburbanization. All these behaviors are qualitatively consistent with **Proposition 1** and examples in Section 5.2, suggesting the robustness of qualitative implications of our theoretical developments.

Remark 6.1. Interestingly, the implications of **Proposition 1** seem to extend to different assumptions on transport technology that are not formally covered by our multiplicative specification of $\mathbf{D}$ in Assumption RE. For example, consider the models by Mossay and Picard (2011); Picard and Tabuchi (2013); Blanchet et al. (2016), all of which assume *linear* transport technology. Mossay and Picard (2011) considered a variant of the Beckmann model (Class II) and showed that the only possible equilibrium is a unimodal distribution in a continuous line segment.[13] Blanchet et al. (2016) considered a general Class II model over a continuous one- or two-dimensional space; they showed that equilibrium spatial pattern for the Beckmann model is unique and given by a concave regular paraboloid, i.e., a unimodal pattern. Picard and Tabuchi (2013) also considered a Class II general equilibrium model in a two-dimensional space and showed that spatial distribution become unimodal. For more complicated

[13] As shown by Akamatsu et al. (2017a), the model by Mossay and Picard (2011) can be regarded as a continuous limit of an appropriate discrete-space model.



(a) The Krugman model (Class I) (b) The Allen–Arkolakis model (Class II)

Figure 14: Stable spatial patterns in a square economy ($N = 9^2 = 81$).

examples, the numerical results of Anas and Kim (1996) and Anas et al. (1998) in line segments bear close resemblance to, respectively, agglomeration behaviors of Class I and II models; in fact, although their model of transport technology is by far more elaborated, economic forces in their models can be broadly understood in terms of local and global dispersion forces. These examples highlight the utility of distinguishing local and global dispersion forces. ■

Remark 6.2. Our observations in this section are consistent with findings in mathematics of pattern formation (see, e.g., Kondo and Miura, 2010, for a survey). From a broader scientific perspective, the racetrack setting (i.e., one-dimensional space with the periodic boundary conditions) is a standard starting point of theoretical analysis of pattern formation phenomena (as initiated by Turing, 1953). In particular, it is known that the properties of stable spatial patterns under a model in symmetric settings qualitatively generalize to asymmetric settings (e.g., Arcuri and Murray, 1986); if a model produces multimodal spatial patterns on the circular domain, then we can expect that the model is capable of producing multimodal spatial patterns on asymmetric domains. Furhter, it is now widely accepted

that a basic requirement for the formation of stable multimodal spatial patterns is the existence a short-range positive feedback combined with a *long-range negative feedback* for a concentration of mobile factors (Meinhardt and Gierer, 2000). In our paper, a global dispersion force plays the role of a long-range negative feedback. Thus, our results for spatial economic models seem to be manifestations of a universal mathematical principle. To provide a complete theoretical connection between our results and those in applied mathematics is, however, beyond the scope of this paper. ■

7 Concluding remarks

This paper introduced the dichotomy between "local" and "global" dispersion forces under a general framework that encompasses a wide range of extant economic geography models. The spatial scale of the dispersion force in a model significantly affects the endogenous spatial patterns and comparative statics of that model. Three prototypical model classes are defined according to the spatial scale(s) of their dispersion forces (i.e., only local, only global, and both local and global). The knowledge of the spatial scale of dispersion forces provides consistent interpretations to the empirical literature and qualitative characterizations for the comparative statics of structural models. We hope our results and methods can be extended to achieve a unified understanding of the robust properties of a broader class of economic geography models.

There are two directions that worth further research. First, the generalization of the theoretical results to asymmetric proximity structures is of importance. An efficient strategy is to fix a few representative models—instead of geography—as test pilots and identify general insights when proximity structure varies systematically (as in Matsuyama, 2017). The basic implications of **Proposition 1** for the polarity of endogenous spatial patterns—a unimodal pattern or multimodal pattern—may well be robust to the generalizations of an assumed geography, as discussed in Section 6.2. Second, the symmetric racetrack geography can be used as a standard testbed to investigate the implications of endogenous mechanisms for a given model. For instance, Dingel et al. (2018) employed a circular geography to theoretically characterize the welfare effects of exogenous productivity differences under a standard international trade model. Important topics include the consideration of multiple types of mobile agents that are subject to different proximity matrices and/or different degrees of increasing returns. Such a structure is ubiquitous in intra-city models with both firms and households (e.g., urban models after Fujita and Ogawa, 1982) or in multiple-sector models (Fujita et al., 1999a). A circular geography provides a canonical starting point for this type of models (Tabuchi and Thisse, 2011; Osawa and Akamatsu, 2020).

A Proofs

A.1 Proof of Proposition 1

We characterize stability of $\bar{\boldsymbol{x}} = (\bar{x}, \bar{x}, \ldots, \bar{x})$ and the destabilization of, and *bifurcation* from, it. Appendix B collects the technical facts referenced in the following.

To define stability of $\bar{\boldsymbol{x}}$, some myopic dynamics must be assumed. A myopic dynamic describes the rate of change in $\boldsymbol{x}$. Denote the dynamic that adjusts $\boldsymbol{x}$ over $\mathcal{X}$ by $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$, where $\dot{\boldsymbol{x}}$ represents the time derivative. For the majority of myopic dynamics in the literature, we have $\boldsymbol{f}(\boldsymbol{x}) \equiv \tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{v}(\boldsymbol{x}))$ where $\tilde{\boldsymbol{f}}$ maps each pair $(\boldsymbol{x}, \boldsymbol{v}(\boldsymbol{x}))$ of a state and its associated payoff to a motion vector $\dot{\boldsymbol{x}}$ that satisfies $\mathbf{1}^\top \dot{\boldsymbol{x}} = 0$. We will focus exclusively on such dynamics. Let *restricted equilibrium* be a state $\boldsymbol{x}^* \in \mathcal{X}$ such that $v_j(\boldsymbol{x}^*) = v_k(\boldsymbol{x}^*)$ for all $j, k \in \{i \in \mathcal{N} \mid x_i^* > 0\}$, that is, a spatial distribution in which all populated regions earn the same payoff level. A spatial equilibrium is always a restricted equilibrium.

We assume that $\boldsymbol{f}$ and $\tilde{\boldsymbol{f}}$ are differentiable and satisfy:

$$\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0} \text{ if } \boldsymbol{x} \text{ is a restricted equilibrium,} \tag{RS}$$

$$\text{if } \boldsymbol{f}(\boldsymbol{x}) \neq \mathbf{0}, \text{ then } \boldsymbol{v}(\boldsymbol{x})^\top \boldsymbol{f}(\boldsymbol{x}) > 0, \text{ and} \tag{PC}$$

$$\mathbf{P}\tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{v}(\boldsymbol{x})) = \tilde{\boldsymbol{f}}(\mathbf{P}\boldsymbol{x}, \mathbf{P}\boldsymbol{v}(\boldsymbol{x})) \text{ for all permutation matrices } \mathbf{P} \text{ with } \mathbf{PD} = \mathbf{DP}. \tag{Sym}$$

We call dynamics that satisfy (RS), (PC), and (Sym) *admissible dynamics*.

Remark A.1. The conditions (RS) and (PC) are, respectively, called *restricted stationality* and *positive correlation* (Sandholm, 2010), and are the most minimal assumptions on a dynamic $\boldsymbol{f}$ to be "consistent" with the underlying model $\boldsymbol{v}$. The symmetry assumption (Sym) ensures that $\boldsymbol{f}$ does not feature ex-ante preference over alternatives $\mathcal{N}$. We assume $\boldsymbol{f}$ is defined for all nonnegative orthant $\mathbb{R}^N_{\geq 0}$ to avoid unnecessary technical complication. Also, we suppose $\boldsymbol{f}$ is C^1 only because we employ linear stability as the definition of stability. ■

Example A.1. Admissible dynamics include, e.g., the *Brown–von Neumann–Nash dynamic* (Brown and von Neumann, 1950; Nash, 1951), the *Smith dynamic* (Smith, 1984), and *Riemannian game dynamics* (Mertikopoulos and Sandholm, 2018) that satisfy (Sym), e.g., the *Euclidian projection dynamic* (Dupuis and Nagurney, 1993) and the *replicator dynamic* (Taylor and Jonker, 1978). ■

Consider a rest point $\boldsymbol{x}^*$ of $\boldsymbol{f}$, i.e., $\boldsymbol{x}^* \in \mathcal{X}$ such that $\boldsymbol{f}(\boldsymbol{x}^*) = \mathbf{0}$. Denote the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x}^*$ by $\nabla \boldsymbol{f}(\boldsymbol{x}^*) = [\frac{\partial f_i}{\partial x_j}(\boldsymbol{x}^*)]$. Then, $\boldsymbol{x}^*$ is *linearly stable* if all the eigenvalues of $\nabla \boldsymbol{f}(\boldsymbol{x}^*)$, which we denote by $\{\eta_k\}$, have negative real parts (see, e.g., Hirsch et al., 2012). We say spatial equilibrium $\boldsymbol{x}^*$ to be *stable* (*unstable*) when it is linearly stable (unstable) under admissible dynamics.

Consider $\bar{\boldsymbol{x}}$. Suppose $\bar{\boldsymbol{x}}$ is an isolated spatial equilibrium. Then, (PC) implies that neighborhood $\mathcal{O} \subset \mathcal{X}$ of $\bar{\boldsymbol{x}}$ exists and $\boldsymbol{v}(\boldsymbol{x})^\top \boldsymbol{f}(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{O} \setminus \{\bar{\boldsymbol{x}}\}$. By expanding $\boldsymbol{v}$ and $\boldsymbol{f}$ about $\bar{\boldsymbol{x}}$, we see

$$\left(\boldsymbol{v}(\bar{\boldsymbol{x}}) + \nabla \boldsymbol{v}(\bar{\boldsymbol{x}})\boldsymbol{z}\right)^\top \left(\boldsymbol{f}(\bar{\boldsymbol{x}}) + \nabla \boldsymbol{f}(\bar{\boldsymbol{x}})\boldsymbol{z}\right) > 0. \tag{A.1}$$

Note that $v(\bar{x}) = \bar{v}\mathbf{1}$, $\nabla v(\bar{x}) = \frac{\bar{v}}{\bar{x}}\mathbf{V}$, $f(\bar{x}) = \mathbf{0}$ by (RS), and $\mathbf{1}^\top \nabla f(\bar{x}) z = 0$; the last equality $\mathbf{1}^\top \nabla f(\bar{x}) z = 0$ follows because $\dot{x} = f(x) \approx f(\bar{x}) + \nabla f(\bar{z}) z = \nabla f(\bar{x}) z$ and $\mathbf{1}^\top \dot{x} = \sum_{i \in \mathcal{N}} \dot{x}_i = 0$ must hold true for all x, since the total mass of agents is a constant. From (A.1), we then see

$$\tfrac{\bar{v}}{\bar{x}} (\mathbf{V} z)^\top (\nabla f(\bar{x}) z) > 0 \tag{A.2}$$

for any infinitesimal migration $z = x - \bar{x}$ from the uniform distribution.

For canonical models, there is a rational function $G(t) = \frac{G^\sharp(t)}{G^\flat(t)}$ such that $\mathbf{V} = G(\bar{\mathbf{D}}) = G^\flat(\bar{\mathbf{D}})^{-1} G^\sharp(\bar{\mathbf{D}})$, where $G^\sharp(t)$ and $G^\flat(t)$ are some polynomials. We assume $G^\flat(t) > 0$, so that $G^\sharp(t)$ is a net gain function. $\bar{\mathbf{D}}$ is real, symmetric, and *circulant* matrix under Assumption RE (see Appendix B). Then, by Fact B.2, $\mathbf{V}$ is real, symmetric, and circulant matrix. Because of (Sym), $\nabla f(\bar{x})$ is also real, symmetric, and circulant matrix. Then, by Fact B.3, $\mathbf{D}$, $\mathbf{V}$, and $\nabla f(\bar{x})$ share the same set of eigenvectors $\{z_k\}$.

For every eigenvector z_k (of $\mathbf{D}$, $\mathbf{V}$, or $\nabla f(\bar{x})$), (A.2) implies that

$$(\mathbf{V} z_k)^\top (\nabla f(\bar{x}) z_k) = \omega_k \eta_k > 0, \tag{A.3}$$

where ω_k and η_k are the (real) eigenvalues of $\mathbf{V}$ and $\nabla f(\bar{x})$ associated with z_k. Thus, $\operatorname{sgn}[\eta_k] = \operatorname{sgn}[\omega_k] = \operatorname{sgn}[G(\chi_k(\phi))] = \operatorname{sgn}[G^\sharp(\chi_k(\phi))]$ where $G^\sharp$ is a net gain function of the model. Therefore, $\bar{x}$ is stable spatial equilibrium under any admissible dynamic if and only if $\omega_k^\sharp \equiv G^\sharp(\chi_k(\phi)) < 0$ for all k. Note that η_k and ω_k are both real because $\nabla f(\bar{x})$ and $\mathbf{V}$ are both symmetric.

The eigenpairs $\{(\chi_k, z_k)\}$ of the row-normalized proximity matrix $\bar{\mathbf{D}}$ are derived in Akamatsu et al. (2012), where z_k is the eigenvector of $\bar{\mathbf{D}}$ associated to the eigenvalue χ_k. For self-containedness, Lemma B.1 and Lemma B.2 in Appendix B summarizes the relevant properties for $\{(\chi_k, z_k)\}$. When $\{(\chi_k, z_k)\}$ are available, the eigenpairs of $\mathbf{V} = G(\bar{\mathbf{D}})$ are given by $\{(G(\chi_k), z_k)\}$ (see Fact B.1). Akamatsu et al. (2012) shows that there are M relevant eigenvalues for $\bar{\mathbf{D}}$. Thus, $\bar{x}$ is stable if $\omega_k^\sharp \equiv G^\sharp(\chi_k(\phi)) < 0$ for all $k \in \mathcal{K} \equiv \{1, 2, \ldots, M\}$.

Below, we use the fact that

$$\max_{k \in \mathcal{K}} \chi_k(\phi) = \chi_1(\phi) \quad \text{and} \quad \min_{k \in \mathcal{K}} \chi_k(\phi) = \chi_M(\phi) \tag{A.4}$$

at any given level of $\phi \in (0, 1)$ under Assumption RE. Refer to Figure 15, which shows $\{\omega_k^\sharp\}$, $G^\sharp(\chi)$, and $\{\chi_k\}$, to understand the following arguments.

Class I. By assumption, there is χ^* such that $G^\sharp(\chi) < 0$ for all $\chi \in (\chi^*, 1)$, that $G^\sharp(\chi^*) = 0$, and that $G^\sharp(\chi) > 0$ for all $\chi \in (0, \chi^*)$. By Lemma B.1, $\{\chi_k(\phi)\}$ are strictly decreasing from 1. Thus, $\bar{x}$ is stable if and only if $\chi_k \in (\chi^*, 1)$, so that $\omega_k^\sharp \equiv G^\sharp(\chi_k) < 0$, for all $k \in \mathcal{K}$, i.e., if $\chi^* < \min_{k \in \mathcal{K}} \chi_k = \chi_M$. Thus, $\bar{x}$ is stable for all $(0, \phi_M^*)$ where $\phi_M^* = \frac{1 - \sqrt{\chi^*}}{1 + \sqrt{\chi^*}}$ is the unique solution for $\chi_M(\phi) = \chi^*$. Because $G^\sharp(\chi) > 0$ for all $\chi \in (0, \chi^*)$ and χ_M is strictly decreasing, $\bar{x}$ is unstable for all $(\phi_M^*, 1)$ because $\omega_M^\sharp > 0$ for the range.

Class II. By assumption, there is χ^{**} such that $G^\sharp(\chi) < 0$ for all $\chi \in (0, \chi^{**})$, that $G^\sharp(\chi^{**}) = 0$, and that $G^\sharp(\chi^{**}) > 0$ for all $\chi \in (\chi^{**}, 1)$. Thus, $\bar{x}$ is stable if and only if $\chi_k \in (0, \chi^{**})$, so that $\omega_k^\sharp = G^\sharp(\chi_k) < 0$, for all $k \in \mathcal{K}$, i.e., if $\chi^{**} > \max_{k \in \mathcal{K}} \chi_k = \chi_1$. Thus, $\bar{x}$ is stable for all $(\phi_1^*, 1)$ where

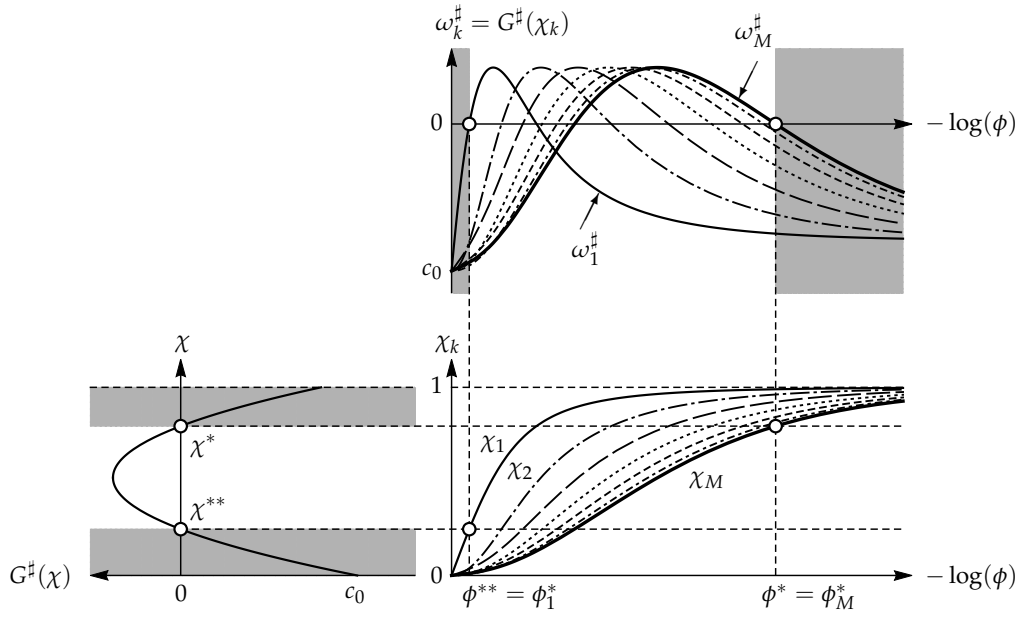


Figure 15: Net gain function $G^\sharp(\chi)$ and net agglomeration forces $\omega_k^\sharp$

Notes: Top: Graphs of $\omega_k^\sharp = G^\sharp(\chi_k)$. Bottom left: Net gain function $G^\sharp$ for Class III models with a quadratic net gain function of the form $G^\sharp(\chi) = c_0 + c_1\chi + c_2\chi^2$. Note that $G^\sharp(0) = c_0$. Bottom right: The eigenvalues $\{\chi_k(\phi)\}$ of $\bar{\mathbf{D}}$, which are model-independent. Uniform distribution $\bar{\boldsymbol{x}}$ is stable in the shaded regions of ϕ or χ. For the ϕ axis, negative log scale is used for better readability, although similar plots in the main text are shown with ordinary ϕ axis. Note that $\max\{\chi_k\} = \chi_1$ and $\min\{\chi_k\} = \chi_M$ at any given level of ϕ.

ϕ_1^* is the unique solution for $\chi_1(\phi) = \chi^{**}$. Because $G^\sharp(\chi) > 0$ for all $\chi \in (\chi^{**}, 1)$ and χ_1 is strictly decreasing, $\bar{\boldsymbol{x}}$ is unstable for all $(0, \phi_1^*)$.

Class III. Via a similar logic, we see $\bar{\boldsymbol{x}}$ is stable if $\phi \in (0, \phi_M^*) \cup (\phi_1^*, 1)$.

Consider a state where $\bar{\boldsymbol{x}}$ is stable. Suppose one and only one ω_k ($k \in \mathcal{K}$) switches its sign from negative to positive at ϕ_k^*. From (A.3), the corresponding eigenvector of the dynamic $\boldsymbol{f}$, η_k, must switch its sign from negative to positive at ϕ_k^*. It is a fact in bifurcation theory that, at such point, $\bar{\boldsymbol{x}}$ must deviate towards the direction of associated eigenvector $\boldsymbol{z}_k$; specifically, $\boldsymbol{z}_k$ is tangent to *unstable manifold* diverging from $\bar{\boldsymbol{x}}$ (see, e.g., Hirsch et al., 2012; Kuznetsov, 2004). Thus, a multimodal pattern with $M = \frac{N}{2}$ peaks emerges at ϕ_M^*, whereas a unimodal distribution emerges at ϕ_1^{**}.

Remark A.2. The bifurcation toward the unimodal direction ($k = 1$) is a *double bifurcation* at which the relevant eigenvalue, ω_1, has multiplicity two, as shown in (B.8). For this case, possible migration patterns are linear combinations of the form $c^+\boldsymbol{z}_1^+ + c^-\boldsymbol{z}_1^-$ with $c^+, c^- \in \mathbb{R}$. Under Assumptions RE and S, we have $(c^+, c^-) = (c, 0)$ or (c, c) for some $c \in \mathbb{R}$ (Ikeda et al., 2012a). Although which of the two possibility occur depends on specific functional form of v, the implication of **Proposition 1** is unaffected because any linear combination of $\boldsymbol{z}_1^+$ and $\boldsymbol{z}_1^-$ is a unimodal cosine curve. ■

Remark A.3. In Assumption RE, we assume that N is a multiple of four to ensure that $\min_{k\in\mathcal{K}}\{\chi_k\} = \chi_M$. Even when Assumption RE is violated, the essential implication of **Proposition 1** on the *polarity* of spatial patterns does not change. By Lemma B.2 (c), if N is even but not a multiple of four, then

$\min_{k\in\mathcal{K}}\{\chi_k\} = \min\{\chi_{M-1}, \chi_M\}$; therefore, the number of peaks in the emergent spatial pattern should be either $M-1$ or M. Similarly, if N is an odd, then $\min_{k\in\mathcal{K}}\{\chi_k\} = \min\{\chi_{\lfloor N/2\rfloor}, \chi_{\lfloor N/2\rfloor-1}\}$. Thus, $\min_{k\in\mathcal{K}}\{\chi_k\}$ corresponds to a multimodal pattern for any N (except for $N = 2$ or 3, which cannot express multimodal patterns). For expositional simplicity, we suppose N is a multiple of four. ■

A.2 Proof of Proposition 2

The equilibrium condition when all regions are populated is given by

$$\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{a}) - \bar{v}(\boldsymbol{x}, \boldsymbol{a})\mathbf{1} = \mathbf{0}, \tag{A.5}$$

where $\bar{v}(\boldsymbol{x}, \boldsymbol{a}) \equiv \sum_{i\in\mathcal{N}} v_i(\boldsymbol{x}, \boldsymbol{a})x_i$ is the average payoff and $\mathbf{1}$ is N-dimensional all-one vector. The pair $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})$ is a solution to (A.5). When $\boldsymbol{a} = \bar{\boldsymbol{a}} + \boldsymbol{\epsilon}$ with small $\boldsymbol{\epsilon} = (\epsilon_i)_{i\in\mathcal{N}}$, there is a spatial equilibrium nearby $\bar{\boldsymbol{x}}$ because $\boldsymbol{v}$ is differentiable. Let $\boldsymbol{x}(\boldsymbol{a})$ denote the perturbed version of the uniform distribution, which is a function in $\boldsymbol{a}$. In the following, we consider some level of ϕ such that $\bar{\boldsymbol{x}}$ is stable, because otherwise studying a perturbed version of $\bar{\boldsymbol{x}}$ is meaningless.

The covariance ρ discussed in Section 6.1 is evaluated as follows:

$$\rho \equiv (\boldsymbol{a} - \bar{\boldsymbol{a}})^\top (\boldsymbol{x}(\boldsymbol{a}) - \bar{\boldsymbol{x}}) = (\mathbf{C}\boldsymbol{a})^\top \mathbf{C}\boldsymbol{x}(\boldsymbol{a}) = \boldsymbol{a}^\top \mathbf{C}\boldsymbol{x}(\boldsymbol{a}) \tag{A.6}$$

where $\mathbf{C} \equiv \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$ is the N-dimensional centering matrix. Let $\mathbf{X} \equiv [\frac{\partial x_i}{\partial a_j}(\bar{\boldsymbol{a}})]$ is the Jacobian matrix of $\boldsymbol{x}$ with respect to $\boldsymbol{a}$ evaluated at $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})$ Then, $\boldsymbol{x}(\boldsymbol{a}) \approx \bar{\boldsymbol{x}} + \mathbf{X}(\boldsymbol{a} - \bar{\boldsymbol{a}}) = \bar{\boldsymbol{x}} + \mathbf{X}\mathbf{C}\boldsymbol{a}$ and

$$\rho = \boldsymbol{a}^\top \mathbf{C}\mathbf{X}\mathbf{C}\boldsymbol{a} \tag{A.7}$$

since $\mathbf{C}\bar{\boldsymbol{x}} = \mathbf{0}$. The implicit function theorem regarding (A.5) at $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})$ gives:

$$\begin{aligned}
\mathbf{X} &= -\left(\mathbf{V}_x - \mathbf{1}\bar{\boldsymbol{x}}^\top \mathbf{V}_x - \mathbf{1}\boldsymbol{v}(\bar{\boldsymbol{x}})^\top\right)^{-1}\left(\mathbf{V}_a - \mathbf{1}\bar{\boldsymbol{x}}^\top \mathbf{V}_a\right) && \text{(A.8)}\\
&= \left(\tfrac{\bar{v}}{\bar{x}}\tfrac{1}{N}\mathbf{1}\mathbf{1}^\top - \left(\mathbf{I} - \tfrac{1}{N}\mathbf{1}\mathbf{1}^\top\right)\mathbf{V}_x\right)^{-1}\left(\mathbf{I} - \tfrac{1}{N}\mathbf{1}\mathbf{1}^\top\right)\mathbf{V}_a && \text{(A.9)}\\
&= \tfrac{\bar{x}}{\bar{v}}\left((\mathbf{I} - \mathbf{C}) - \mathbf{C}\tfrac{\bar{x}}{\bar{v}}\mathbf{V}_x\right)^{-1}\mathbf{C}\tfrac{\bar{v}}{\bar{a}}\tfrac{\bar{a}}{\bar{v}}\mathbf{V}_a && \text{(A.10)}\\
&= \tfrac{\bar{x}}{\bar{a}}\left((\mathbf{I} - \mathbf{C}) - \mathbf{C}\mathbf{V}\right)^{-1}\mathbf{C}\mathbf{A} && \text{(A.11)}
\end{aligned}$$

where $\bar{v}$ is the payoff level, $\mathbf{V}_x \equiv [\frac{\partial v_i}{\partial x_j}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})]$, $\mathbf{V}_a \equiv [\frac{\partial v_i}{\partial a_j}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{a}})]$, $\mathbf{V} \equiv \frac{\bar{x}}{\bar{v}}\mathbf{V}_x$, and $\mathbf{A} \equiv \frac{\bar{a}}{\bar{v}}\mathbf{V}_a$. Under Assumptions RE, S, and A, $\mathbf{X}$ is real, symmetric, and circulant because all its components in (A.11) are.

The set of eigenvectors of $\mathbf{CXC}$ can be chosen as the same as in Lemma B.1 (a) because it is a circulant matrix of the same size as $\bar{\mathbf{D}}$ (see Fact B.3). Let $\{\lambda_k\}_{k=0}^M$ be the distinct eigenvalues of $\mathbf{CXC}$. Because $\mathbf{CXC}$ is symmetric, $\mathbf{CXC}$ admits the following eigenvalue decomposition:

$$\mathbf{C}\mathbf{X}\mathbf{C} = \lambda_0 \mathbf{1}\mathbf{1}^\top + \sum_{k=1}^{M-1} \lambda_k \left(\boldsymbol{z}_k^+ \boldsymbol{z}_k^{+\top} + \boldsymbol{z}_k^- \boldsymbol{z}_k^{-\top}\right) + \lambda_M \boldsymbol{z}_M \boldsymbol{z}_M^\top. \tag{A.12}$$

This fact yields the following representation of ρ:

$$\rho = \boldsymbol{a}^\top \mathbf{CXC}\boldsymbol{a} = \sum_{k\neq 0} \tilde{a}_k^2 \lambda_k, \tag{A.13}$$

where $\tilde{\boldsymbol{a}} \equiv (\tilde{a}_k)$ is the representation of $\boldsymbol{a}$ in the new coordinate system. We can omit $k = 0$ since $\lambda_0 = 0$, which reflects that $\boldsymbol{z}_0 = \mathbf{1}$ represents a uniform increase in $\boldsymbol{a}$ and thus is inconsequential. In concrete terms, as all the component matrices in (A.11) are circulant matrices and hence shares the same set of eigenvectors, we can translate the matrix relationship (A.11) to the following expression:

$$\lambda_k = \frac{\bar{x}}{\bar{a}} \frac{\kappa_k \alpha_k}{((1-\kappa_k) - \kappa_k \omega_k)} = -\frac{\bar{x}}{\bar{a}} \frac{\alpha_k}{\omega_k} \qquad \forall k \in \mathcal{K}, \tag{A.14}$$

where κ_k, ω_k, and α_k are the kth eigenvalues of $\mathbf{C}$, $\mathbf{V}$, and $\mathbf{A}$, respectively, and we have $\kappa_0 = 0$ and $\kappa_k = 1$ for all $k \neq 0$. We have $\omega_k = G(\chi_k)$ and $\alpha_k = G^\natural(\chi_k)$ because $\mathbf{G} = G(\bar{\mathbf{D}})$ and $\mathbf{A} = G^\natural(\bar{\mathbf{D}})$, thereby

$$\lambda_k = -\frac{\bar{x}}{\bar{a}} \frac{G^\natural(\chi_k)}{G(\chi_k)} = \frac{\bar{x}}{\bar{a}} \delta(\chi_k) \qquad \forall k \in \mathcal{K} \tag{A.15}$$

and $\lambda_0 = 0$ where $\delta(\chi) \equiv -\frac{G^\natural(\chi)}{G(\chi)}$, and $\{\chi_k\}_{k\in\mathcal{K}}$ are the eigenvalues of $\bar{\mathbf{D}}$.

Thus, $\rho > 0$ for all $\boldsymbol{a}$ if all $\{\lambda_k\}$ are positive except for $\lambda_0 = 0$. The denominator of (A.15), $G(\chi_k)$, must be negative for all k because $\bar{\boldsymbol{x}}$ is stable by assumption. Thus, we see that $\rho > 0$ if $G^\natural(\chi) > 0$ for all χ since $\chi_k \in (0,1)$ for all $k \in \mathcal{K}$.

Proposition 2 follows by noting that

$$\rho'(\phi) = \sum_{k\neq 0} \tilde{a}_k^2 \frac{\mathrm{d}\lambda_k}{\mathrm{d}\phi} = \frac{\bar{x}}{\bar{a}} \sum_{k\neq 0} \tilde{a}_k^2 \delta'(\chi_k) \frac{\mathrm{d}\chi_k}{\mathrm{d}\phi} = -\frac{\bar{x}}{\bar{a}} \sum_{k\neq 0} \tilde{a}_k^2 \delta'(\chi_k) \left|\frac{\mathrm{d}\chi_k}{\mathrm{d}\phi}\right|,$$

where we recall that $\{\chi_k\}_{k\in\mathcal{K}}$ are strictly decreasing in ϕ (Lemma B.2). Thus, if $\delta'(\chi) < 0$ ($\delta'(\chi) > 0$) for all relevant values of χ, then $\rho'(\phi) > 0$ ($\rho'(\phi) < 0$).

B Preliminaries

We collect the relevant facts from matrix analysis for self-containedness. See Horn and Johnson (2012) (henceforth HJ) for a concise reference. First, when we know the eigenpairs (eigenvalue–eigenvector pairs) of a square matrix $\mathbf{A}$, we know those of matrix polynomials of $\mathbf{A}$ (see HJ, Section 1.1).

Fact B.1. Consider a square matrix $\mathbf{A}$ with eigenpairs $\{(\lambda_k, \boldsymbol{z}_k)\}$. For a polynomial $P(t) = \sum_{l=0}^n p_l t^l$, let $P(\mathbf{A})$ be defined by $P(\mathbf{A}) = \sum_{l=0}^n p_l \mathbf{A}^l$ with $\mathbf{A}^0 \equiv \mathbf{I}$. Then, the eigenpairs of $P(\mathbf{A})$ are given by $\{(P(\lambda_k), \boldsymbol{z}_k)\}$. Take another polynomial $Q(t)$. If $Q(\mathbf{A})$ is nonsingular, then the eigenpairs of $Q(\mathbf{A})^{-1}$ are given by $\{(Q(\lambda_k)^{-1}, \boldsymbol{z}_k)\}$. Thus, the eigenpairs of the matrix $G(\mathbf{A}) \equiv Q(\mathbf{A})^{-1} P(\mathbf{A})$ is given by $\{(G(\lambda_k), \boldsymbol{z}_k)\}$ with $G(t) = \frac{P(t)}{Q(t)}$. ◊

Next, a *circulant matrix* $\mathbf{C}$ of size N generated by $\boldsymbol{c} = (c_i)_{i=0}^{N-1}$ is defined by

$$\mathbf{C} = \operatorname{circ}[\boldsymbol{c}] \equiv \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{N-2} & c_{N-1} \\ c_{N-1} & c_0 & c_1 & c_2 & \cdots & c_{N-2} \\ c_{N-2} & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & c_2 \\ c_2 & \cdots & c_{N-2} & c_{N-1} & c_0 & c_1 \\ c_1 & c_2 & \cdots & c_{N-2} & c_{N-1} & c_0 \end{bmatrix}. \tag{B.1}$$

Each row of $\mathbf{C}$ are identical to the previous row moved one position to the right and wrapped around. Every row sum equals to $\boldsymbol{c}^\top \mathbf{1}$ by definition. Circulant matrices are known to satisfy the following properties (see HJ, Section 0.9.6 and Problem 2.2.P10):

Fact B.2. Circulant matrices of size N form a commutative algebra: linear combinations and products of circulants ar circulants; the inverse of a nonsingular circulant is a circulant; any two circulants of the same size commute. ♢

Fact B.3. Let $\mathbf{C} = \operatorname{circ}[\boldsymbol{c}]$ be a real and *symmetric* circulant matrix of size N. Then, $\mathbf{C}$ is diagonalized by an orthogonal matrix $\mathbf{Z}$: $\operatorname{diag}[\boldsymbol{\lambda}] = \mathbf{Z}^\top \mathbf{C} \mathbf{Z}$. The column vectors of the matrix $\mathbf{Z}$ are the eigenvectors of $\mathbf{C}$. Let $\theta \equiv \frac{2\pi}{N}$. Eigenpairs $(\lambda_k, \boldsymbol{z}_k)$ can be chosen to be

$$\lambda_0 = \boldsymbol{c}^\top \mathbf{1}, \qquad \boldsymbol{z}_0 \equiv \langle 1 \rangle_{i=0}^{N-1}, \tag{B.2}$$

$$\lambda_k, \qquad \begin{cases} \boldsymbol{z}_k^+ \equiv \langle \cos(\theta k i) \rangle_{i=0}^{N-1}, \\ \boldsymbol{z}_k^- \equiv \langle \sin(\theta k i) \rangle_{i=0}^{N-1}, \end{cases} \qquad k = 1, 2, \ldots, \lfloor \tfrac{N}{2} \rfloor - 1, \tag{B.3}$$

$$\lambda_{\frac{N}{2}}, \qquad \boldsymbol{z}_{\frac{N}{2}} \equiv \langle (-1)^i \rangle_{i=0}^{N-1}, \qquad \text{if } N \text{ is an even}, \tag{B.4}$$

where $\langle z_i \rangle_{i=0}^{N-1} \equiv \|\boldsymbol{z}\|^{-1} (z_i)_{i=0}^{N-1}$ denotes the normalized version of real vector $\boldsymbol{z}$. Thus, the distinct eigenvalues of $\mathbf{C}$ are given by $\boldsymbol{\lambda} = \{\lambda_k\}_{k \in \mathcal{K}}$ where $\mathcal{K} \equiv \{0, 1, 2, \ldots, \lfloor \frac{N}{2} \rfloor\}$. Those eigenvalues with $k = 1, 2, \ldots, \lfloor \frac{N}{2} \rfloor - 1$ are multiplicity two. $\boldsymbol{z}_0$ is a uniform vector and λ_0 is the row-sum of $\mathbf{C}$; the elements of other eigenvectors sum up to zero. If in addition $c_i > 0$ for all i and $\boldsymbol{c}^\top \mathbf{1} = 1$, then $\mathbf{C} = \operatorname{circ}[\boldsymbol{c}]$ is positive and doubly stochastic (as we assume $\mathbf{C}$ is symmetric). We have $\mathbf{C}\mathbf{1} = \mathbf{1}$ and thus $\lambda_0 = 1$; λ_0 is the maximal eigenvalue (or the spectral radius) of $\mathbf{C}$ and $\mathbf{1}$ is the only strictly positive eigenvector (the Perron–Frobenius theorem). All real symmetric circulant matrices of size N share the same set of eigenvectors. For asymmetric circulant matrices, discrete Fourier transformation matrix can be employed for diagonalization (see, e.g., Akamatsu et al., 2012, for an application). ♢

We collect relevant facts for $\mathbf{D}$ under Assumption RE shown in Akamatsu et al. (2012). $\mathbf{D} = [\phi_{ij}]$ is a circulant matrix, as $\phi_{ij} = \phi^{\ell_{ij}}$ with $\ell_{ij} \equiv \{|i-j|, N - |i-j|\}$ and thus $\phi_{ij} = \phi^{\ell_{ij}} = \phi^{\ell_{i+1,j+1}} = \phi_{i+1,j+1}$ for all i, j (mod N for indices). The eigenvalues $\{\chi_k\}$ of $\bar{\mathbf{D}}$ are given as follows.

Lemma B.1 (Akamatsu et al. (2012), Lemma 4.2)**.** Assume that N is an even and let $M \equiv \frac{N}{2}$. Then,

$$\chi_k(\phi) = \begin{cases} \Psi_k(\phi)\Psi_M(\phi) & (k\text{: even}) \\ \Psi_k(\phi)\Psi_M(\phi)\bar{\Psi}(\phi) & (k\text{: odd}) \end{cases} \quad k = 0, 1, 2, \ldots, M, \tag{B.5}$$

$$\text{where} \quad \Psi_k(\phi) \equiv \frac{1-\phi^2}{1-2\phi\cos(\theta k)+\phi^2} \quad \text{and} \quad \bar{\Psi}(\phi) \equiv \frac{1+\phi^M}{1-\phi^M} \quad \text{with} \quad \theta = \frac{2\pi}{N}. \tag{B.6}$$

These formulae leads to the following lemma (see Akamatsu et al., 2012, Lemma 4.2):

Lemma B.2. Assume that N is an even and let $M \equiv \frac{N}{2}$. Then, $\bar{\mathbf{D}}$ satisfies the following properties:

(a) There are $M+1$ distinct eigenvalues. The eigenpairs $\{(\chi_k, \boldsymbol{z}_k)\}$ are

$$\chi_0 = 1, \qquad \boldsymbol{z}_0 \equiv \langle 1 \rangle_{i=0}^{N-1}, \tag{B.7}$$

$$\chi_k, \qquad \begin{cases} \boldsymbol{z}_k^+ \equiv \langle \cos(\theta k i) \rangle_{i=0}^{N-1}, \\ \boldsymbol{z}_k^- \equiv \langle \sin(\theta k i) \rangle_{i=0}^{N-1}, \end{cases} \qquad k = 1, 2, \ldots, M-1 \tag{B.8}$$

$$\chi_M, \qquad \boldsymbol{z}_M \equiv \langle (-1)^i \rangle_{i=0}^{N-1}. \tag{B.9}$$

where $\theta = \frac{2\pi}{N}$, and by $\langle z_i \rangle_{i=0}^{N-1} \equiv \frac{1}{\|\boldsymbol{z}\|}(z_i)_{i=0}^{N-1}$ we denote a normalized vector.

(b) Every $\chi_k(\phi)$ $(k \neq 0)$ is differentiable, strictly decreasing, and $\lim_{\phi\to 0}\chi_k = 1$ and $\lim_{\phi\to 1}\chi_k = 0$.

(c) For any given $\phi \in (0,1)$, $\{\chi_k\}$ $(k = 0, 1, 2, \ldots, M)$ are ordered as

$$\begin{cases} 1 = \chi_0 > \chi_2 > \cdots > \chi_{2k} > \cdots > \chi_M > 0, \\ 1 > \chi_1 > \chi_3 > \cdots > \chi_{2k+1} > \cdots > \chi_{M-1} > 0, \end{cases} \tag{B.10}$$

with $\chi_0 > \chi_1 > \chi_2$, so that $\max_{k\geq 1}\{\chi_k\} = \chi_1$ and $\min_{k\geq 1}\{\chi_k\} = \min\{\chi_{M-1}, \chi_M\}$.

(d) If N is a multiple of four, $\min_k\{\chi_k\} = \chi_M = \left(\frac{1-\phi}{1+\phi}\right)^2$.

In considering migration patterns from $\bar{\boldsymbol{x}}$, we can exclude $k = 0$ because $\boldsymbol{z}_0^\top \mathbf{1} \neq 0$.

C Cities in Japan: 1970–2015

Data. We use population count data of Japan from Statistics Bureau, Ministry of Internal Affairs and Communications of Japan (1970, 2015).

Method. We define an *urban agglomeration* (*UA*) as a set of contiguous 1 km-by-1 km cells whose population density is at least 1000/km^2 and total population of at least 10,000. The results discussed below remain basically the same for alternative threshold densities and populations.

Below, UA i in year s is said to be *associated with* UA j in year t $(\neq s)$ if the intersection of the spatial coverage of i and that of j accounts for the largest population of i among all the UAs in year t. For years $s < t$, if i and j are associated with each other, they are considered to be *the same* UA. If i is associated

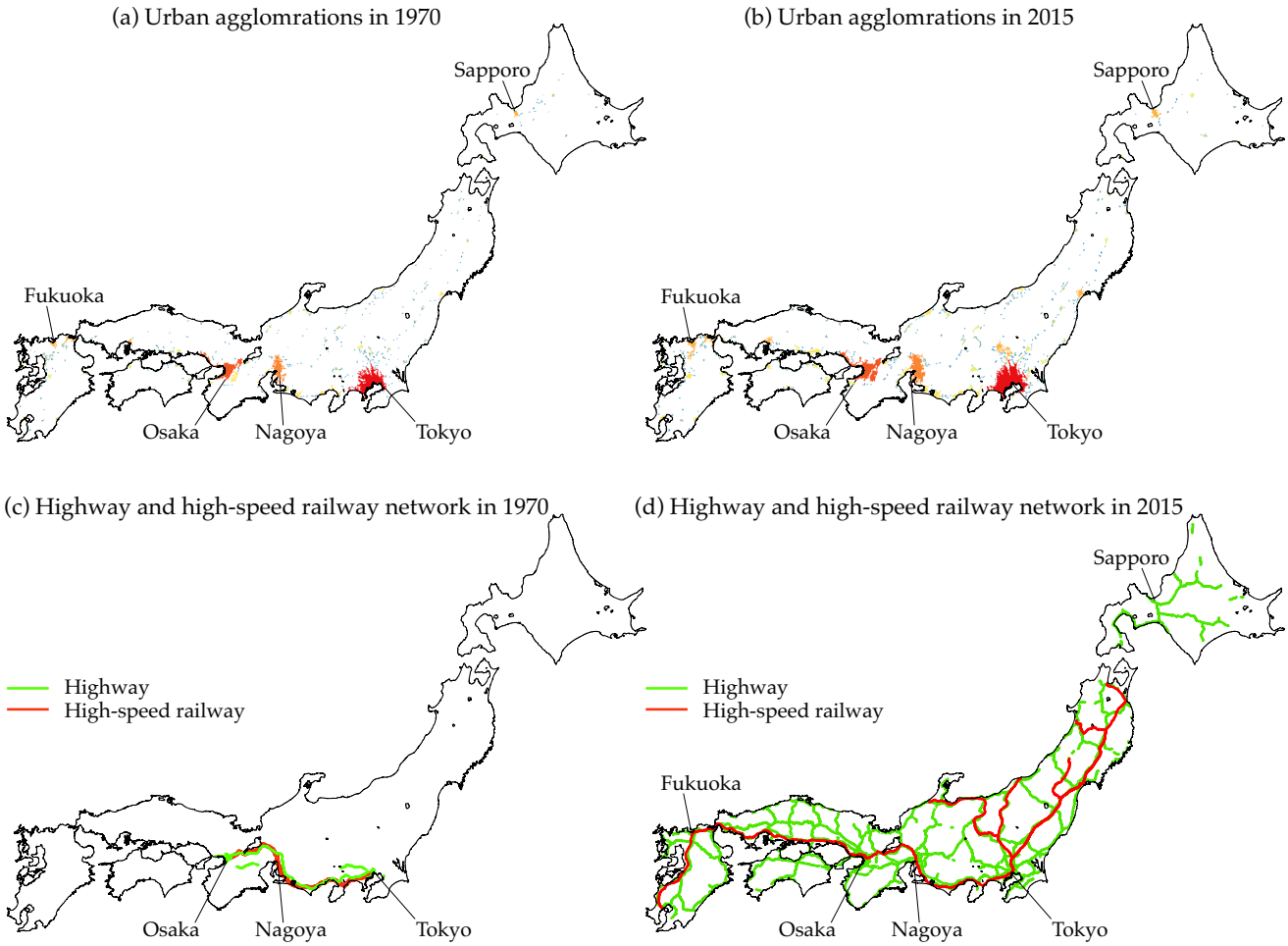


Figure 16: UAs and transport network in Japan

with j but not vice versa, then i is considered to have been *absorbed* into j, while if j is associated with i but not vice versa, then j is considered to have *separated* from i. If i is not associated with any UA in year t, then i is considered to have *disappeared* by year t, while if j is not associated with any UA in year s, then j is considered to have newly *emerged* by year t.

For the part of Japan contiguous by roads to at least one of the four major islands (Hokkaido, Honshu, Shikoku, and Kyushu), 503 and 450 UAs are identified, as depicted in Panels (a) and (b) of Figure 16 for 1970 and 2015, respectively, where the warmer color indicates a larger population. These together account for 64% and 78% of the total population in 1970 and 2015, respectively. Thus, there is a substantial 18% increase in the urban share over these 40 years. Of the 503 UAs that existed in 1970, 302 survived, while 201 either disappeared or integrated with other UAs by 2015. Of the 450 UAs that existed in 2015, 148 were newly formed after 1970 (including those split from existing UAs).

Panels (c) and (d) of Figure 16 show the highway and high-speed railway networks in use in 1970 and 2015, respectively. They show a substantial expansion of these networks during these 45 year.

Panels (a), (b), and (c) of Figure 17 show, respectively, the distributions of the growth rates of population share (in the national population), the areal size, and population density of individual UAs

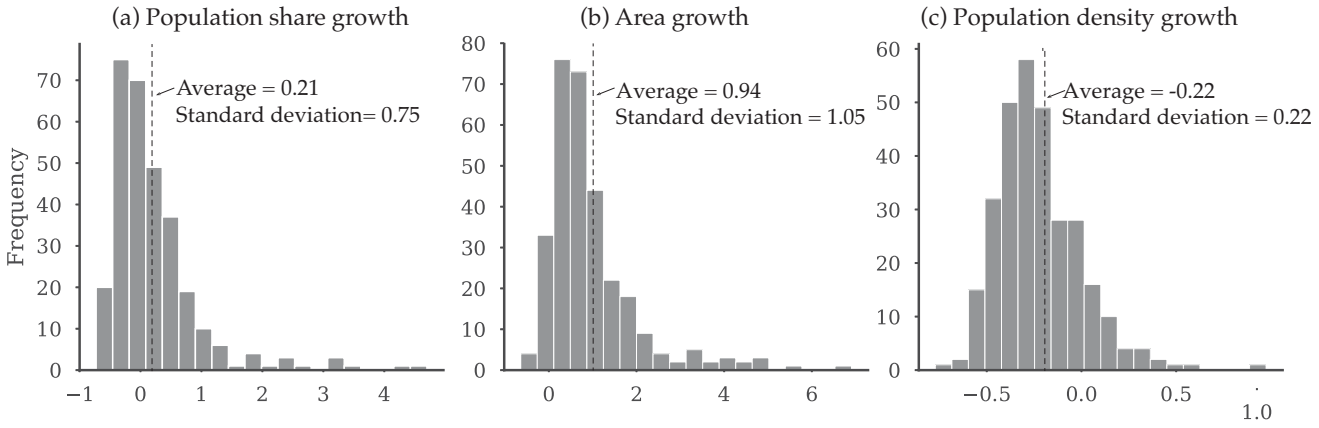


Figure 17: Growth rates of the sizes of UAs in Japan.

for the set of the 302 UAs that survived throughout the 45-year period. A UA experienced an average growth rate of 21% (75%) of population share, 94% (105%) of areal size, and -22% (22%) of population density (per km^2), respectively, where the numbers in parentheses are the standard deviations.

As a larger population share was concentrated in a smaller number of UAs in 2015 than in 1970, the spatial size of an individual UA almost doubled on average. These spatial expansions are not simply due to the shortage of available land in UAs, as population density decreased by 22% on average. We take this as evidence of a decline in the number of major population concentrations combined with local flattening of each concentration in the course of the improvement in interregional transport access.

D Supporting Computations

We collect supporting computations for the examples provided in the main text. Below, $\mathbf{F}_x = [\frac{\partial f_i}{\partial x_j}]$ denotes the partial derivative (the Jacobian matrix) of a vector-valued function $\boldsymbol{f}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$. For example, $\mathbf{V}_x \equiv [\frac{\partial v_i}{\partial x_j}]$, $\tilde{\mathbf{V}}_w \equiv [\frac{\partial \tilde{v}_i}{\partial w_j}]$, and $\mathbf{W}_x \equiv [\frac{\partial w_i}{\partial x_j}]$. They are evaluated at $\bar{\boldsymbol{x}}$ unless otherwise noted. Throughout, $\bar{v}$, $\bar{w}$, $\bar{e}$ and so on represent v_i, w_i, e_i evaluated at $\boldsymbol{x} = \bar{\boldsymbol{x}} = \bar{x}\mathbf{1}$, unless otherwise noted.

D.1 General Observations

D.1.1 The payoff elasticity matrix $\mathbf{V} = \frac{\bar{x}}{\bar{v}}\mathbf{V}_x$

The payoff functions for the models we referenced in the main text reduce to the following form:

$$\boldsymbol{v}(\boldsymbol{x}) = \tilde{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{w}), \tag{D.1}$$

$$\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{w}) = \mathbf{0}. \tag{D.2}$$

The condition (D.2) represents, e.g., the market equilibrium conditions for a given $\boldsymbol{x}$ that defines endogenous variable $\boldsymbol{w}$ other than $\boldsymbol{x}$ as an implicit function of $\boldsymbol{x}$. For $\boldsymbol{v}(\boldsymbol{x})$ to be well-defined, (D.2) must admit a unique solution of $\boldsymbol{w}$ for all $\boldsymbol{x} \in \mathcal{X}$. We assume that (D.2) has a unique solution for all $\boldsymbol{x} \in \mathcal{X}^\circ$,

where $\mathcal{X}^\circ \equiv \{\boldsymbol{x} \in \mathcal{X} \mid x_i > 0 \ \forall i \in \mathcal{N}\}$ denote the interior of $\mathcal{X}$. In general, we have

$$\mathbf{V}_x(\boldsymbol{x}) = \tilde{\mathbf{V}}_x(\boldsymbol{x}) + \tilde{\mathbf{V}}_w(\boldsymbol{x})\mathbf{W}_x(\boldsymbol{x}), \tag{D.3}$$

$$\mathbf{W}_x(\boldsymbol{x}) = -\mathbf{S}_w(\boldsymbol{x})^{-1}\mathbf{S}_x(\boldsymbol{x}), \tag{D.4}$$

where $\mathbf{W}_x(\boldsymbol{x})$ is obtained by the implicit function theorem regarding (D.2).

Under Assumptions RE and S, if $\boldsymbol{x} = \bar{\boldsymbol{x}}$, $\mathbf{V}_x = \mathbf{S}_w^{-1}(\mathbf{S}_w\tilde{\mathbf{V}}_x - \tilde{\mathbf{V}}_w\mathbf{S}_x)$, since all matrices commute (they are real, symmetric, and circulant at $\bar{\boldsymbol{x}}$). In the Krugman and Helpman models, $G^\flat(\bar{\mathbf{D}})$ arises from $\mathbf{S}_w$ and represents general equilibrium effects through (D.2). For any model whose payoff function reduces to the equations of the form (D.1) and (D.2), $\mathbf{V} = G^\flat(\bar{\mathbf{D}})^{-1}G^\sharp(\bar{\mathbf{D}})$ where polynomials $G^\sharp(t)$ and $G^\flat(t)$ are chosen such that $G^\sharp(\bar{\mathbf{D}}) = \mathbf{S}_w\tilde{\mathbf{V}}_x - \tilde{\mathbf{V}}_w\mathbf{S}_x$ and $G^\flat(\bar{\mathbf{D}}) = \mathbf{S}_w$.

Example D.1. In Examples 3.1 and 3.2, (D.2) is given by

$$s_i(\boldsymbol{x}, \boldsymbol{w}) = w_i x_i - \sum_{j\in\mathcal{N}} m_{ij}e_j = 0, \tag{D.5}$$

where $e_i = e(w_i, x_i)$ with some nonnegative function e and $\mathbf{M} = \mathbf{M}(\boldsymbol{x}) = [m_{ij}]$ is

$$m_{ij} = \frac{x_i w_i^{1-\sigma}\phi_{ij}}{\sum_{k\in\mathcal{N}} x_k w_k^{1-\sigma}\phi_{kj}}. \tag{D.6}$$

In matrix form, we may write (D.5) as $\boldsymbol{y} - \mathbf{M}\boldsymbol{e} = \mathbf{0}$ where $\boldsymbol{y} = (w_i x_i)_{i\in\mathcal{N}}$. It gives

$$\mathbf{S}_x(\boldsymbol{x}) = \operatorname{diag}[\boldsymbol{w}] - \left(\operatorname{diag}[\mathbf{M}\boldsymbol{e}] - \mathbf{M}\operatorname{diag}[\boldsymbol{e}]\mathbf{M}^\top\right)\operatorname{diag}[\boldsymbol{x}]^{-1} - \mathbf{M}\mathbf{E}_x, \tag{D.7a}$$

$$\mathbf{S}_w(\boldsymbol{x}) = \operatorname{diag}[\boldsymbol{x}] + (\sigma-1)\left(\operatorname{diag}[\mathbf{M}\boldsymbol{e}] - \mathbf{M}\operatorname{diag}[\boldsymbol{e}]\mathbf{M}^\top\right)\operatorname{diag}[\boldsymbol{w}]^{-1} - \mathbf{M}\mathbf{E}_w. \tag{D.7b}$$

Suppose Assumptions RE and S. Suppose $\boldsymbol{x} = \bar{\boldsymbol{x}}$ and let $\bar{w}$ be the uniform level of wage at $\bar{\boldsymbol{x}}$. Note that $\mathbf{M} = \bar{\mathbf{D}}$ when $\boldsymbol{x} = \bar{\boldsymbol{x}}$. Let $\mathbf{E}_x = \epsilon_x\bar{w}\mathbf{I}$ and $\mathbf{E}_w = \epsilon_w\bar{x}\mathbf{I}$ at $\bar{\boldsymbol{x}}$. Let $\bar{e} = e(\bar{w}, \bar{x})$ and $\zeta \equiv \frac{\bar{e}}{\bar{w}\bar{x}}$. We see that

$$\mathbf{S}_x = -\bar{w}\left((\zeta-1)\mathbf{I} + \epsilon_x\bar{\mathbf{D}} - \zeta\bar{\mathbf{D}}^2\right), \tag{D.8a}$$

$$\mathbf{S}_w = \bar{x}\left((1+\zeta(\sigma-1))\mathbf{I} - \epsilon_w\bar{\mathbf{D}} - \zeta(\sigma-1)\bar{\mathbf{D}}^2\right). \tag{D.8b}$$

If $e(w_i, x_i) = w_i x_i$, then $\zeta = 1$ and $\epsilon_x = \epsilon_w = 1$ at $\bar{\boldsymbol{x}}$, thereby $\mathbf{W}_x = \frac{\bar{w}}{\bar{x}}(\sigma\mathbf{I} + (\sigma-1)\bar{\mathbf{D}})^{-1}\bar{\mathbf{D}}$. ■

D.1.2 The matrix $\mathbf{A} = \frac{\bar{x}}{\bar{a}}\mathbf{V}_a$

In (A.11), $\mathbf{X} = [\partial x_i(\bar{\boldsymbol{a}})/\partial a_i] = \mathbf{X}_a$ acts as $\hat{\mathbf{X}} \equiv -\mathbf{V}_x^{-1}\mathbf{V}_a$ for $\boldsymbol{z}$ such that $\boldsymbol{z}^\top\mathbf{1} = \mathbf{0}$. Thus, $\mathbf{V}_a$ is of interest.

For purely local characteristics (Example 6.1), since $v_i(\boldsymbol{x}, \boldsymbol{a}) = a_i v_i(\boldsymbol{x})$, it follows that $\mathbf{V}_a = \operatorname{diag}[\boldsymbol{v}(\boldsymbol{x})]$. At $\bar{\boldsymbol{x}}$, we have $\mathbf{V}_a = \bar{v}\mathbf{I}$. Thus, $\hat{\mathbf{X}} = -\bar{v}\mathbf{V}_x^{-1}$.

For regional characteristics that affect trade flows (Examples 6.5 and 6.6), the payoff function and the market equilibrium condition are, respectively, modified to $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{a}) = \tilde{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{a})$ and $\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{a}) = \mathbf{0}$. By

applying the implicit function theorem, we see $\mathbf{V}_a = \tilde{\mathbf{V}}_a + \tilde{\mathbf{V}}_w \mathbf{W}_a = \tilde{\mathbf{V}}_a - \tilde{\mathbf{V}}_w \mathbf{S}_w^{-1} \mathbf{S}_a$. Since all matrices commute at $\bar{\boldsymbol{x}}$ under Assumption RE, it is equivalent to consider

$$\hat{\mathbf{X}} = -\left(\tilde{\mathbf{V}}_x - \tilde{\mathbf{V}}_w \mathbf{S}_w^{-1} \mathbf{S}_x\right)^{-1} \left(\tilde{\mathbf{V}}_a - \tilde{\mathbf{V}}_w \mathbf{S}_w^{-1} \mathbf{S}_a\right) = \left(\mathbf{S}_w \tilde{\mathbf{V}}_x - \tilde{\mathbf{V}}_w \mathbf{S}_x\right)^{-1} \left(\tilde{\mathbf{V}}_w \mathbf{S}_a - \mathbf{S}_w \tilde{\mathbf{V}}_a\right). \tag{D.9}$$

Example D.2. In Example 6.5, we have

$$s_i(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{a}) = w_i x_i - \sum_{j \in \mathcal{N}} \frac{x_i a_i w_i^{1-\sigma} \phi_{ij}}{\sum_{k \in \mathcal{N}} x_k a_k w_k^{1-\sigma} \phi_{kj}} e_j = 0. \tag{D.10}$$

We have $\mathbf{S}_a = -\left(\operatorname{diag}[\mathbf{M}\boldsymbol{y}] - \mathbf{M} \operatorname{diag}[\boldsymbol{y}] \mathbf{M}^\top\right) \operatorname{diag}[\boldsymbol{a}]^{-1}$ and $\mathbf{S}_a = -\frac{\bar{e}}{\bar{a}} \left(\mathbf{I} - \bar{\mathbf{D}}^2\right) = -\frac{\bar{e}}{\bar{a}} \left(\mathbf{I} - \bar{\mathbf{D}}\right) \left(\mathbf{I} + \bar{\mathbf{D}}\right)$ at $\boldsymbol{x} = \bar{\boldsymbol{x}}$. See also Section D.2.2. ■

Example D.3. In Example 6.6, we have

$$s_i(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{a}) = w_i x_i - \sum_{j \in \mathcal{N}} \frac{x_i w_i^{1-\sigma} \phi_{ij}}{\sum_{k \in \mathcal{N}} x_k w_k^{1-\sigma} \phi_{kj}} e(w_j, x_j, a_j) = 0 \tag{D.11}$$

where e maps the tuple (w_j, x_j, a_j) to the regional expenditure. Then, we have $\mathbf{S}_a = -\mathbf{M}\mathbf{E}_a$, or $\mathbf{S}_a = -\epsilon_a \bar{\mathbf{D}}$ at $\bar{\boldsymbol{x}}$ where $\epsilon_a = \frac{\partial e(\bar{x}, \bar{w}, \bar{a})}{\partial a_i}$. See also Section D.2.1. ■

D.2 Model-by-model Analyses

We provide omitted derivations of the net gain functions $G^\sharp$ for the main examples in the main text. See Akamatsu et al. (2017b, 2019) for derivations for other models in Examples 3.4, 3.5, and 3.6.

D.2.1 Krugman (1991b) model (Examples 3.1 and 6.6)

There are two types of workers, mobile and immobile, and their total masses are 1 and L, respectively. $\boldsymbol{x} \equiv (x_i)_{i \in \mathcal{N}}$ is the distribution of mobile workers. Each worker supplies one unit of labor inelastically.

There are two industrial sectors: agriculture (abbreviated as A) and manufacturing (abbreviated as M). The A-sector is perfectly competitive and a unit input of immobile labor is required to produce one unit of goods. The M-sector follows Dixit–Stiglitz monopolistic competition. M-sector goods are horizontally differentiated and produced under increasing returns to scale using mobile labor as the input. The goods of both sectors are transported. Transportation of A-sector goods is frictionless, while that of M-sector goods is of an iceberg form. For each unit of M-sector goods transported from region i to j, only the proportion $1/\tau_{ij}$ arrives, where $\tau_{ij} > 1$ for $i \neq j$ and $\tau_{ii} = 1$.

All workers have an identical preference for both M- and A-sector goods. The utility of a worker in region i is given by a two-tier form. The upper tier is Cobb–Douglas over the consumption of A-sector goods C_i^{A} and that of M-sector constant-elasticity-of-substitution (CES) aggregate C_i^{M} with $\sigma > 1$

$$C_i^{\mathrm{M}} \equiv \left(\sum_{j \in \mathcal{N}} \int_0^{n_j} q_{ji}(\xi)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\xi \right)^{\frac{\sigma}{\sigma-1}}, \tag{D.12}$$

that is, $u_i = (C_i^{\mathrm{A}})^{\mu}(C_i^{\mathrm{A}})^{1-\mu}$ where $\mu \in (0,1)$ is the constant expenditure of the latter. With free trade in the A-sector, the wage of the immobile worker is equalized, and we normalize it to unity by taking A-sector goods as the numéraire. Consequently, region i's expenditure on the M-sector goods is given by $e_i = \mu(w_i x_i + l_i)$ where l_i denotes the mass of immobile workers in region i.

In the M-sector, to produce q units, a firm requires $\alpha + \beta q$ units of mobile labor. Profit maximization of firms yields the price of differentiated goods produced in region i and exported to j as $p_{ij} = \frac{\sigma\beta}{\sigma-1} w_i \tau_{ij}$, which in turn determines gravity trade flow from j to i. That is, when X_{ij} denotes the price of M-sector goods produced in region i and sold in region j, $X_{ij} = m_{ij} e_j$ where the share $m_{ij} \in (0,1)$ is defined by (D.6) with $\phi_{ij} \equiv \tau_{ij}^{1-\sigma}$. The proximity matrix is thus $\mathbf{D} = [\phi_{ij}] = [\tau_{ij}^{1-\sigma}]$.

Given $\boldsymbol{x}$, we determine the market wage $\boldsymbol{w} \equiv (w_i)_{i\in\mathcal{N}}$ by the M-sector product market-clearing condition, zero-profit condition, and mobile labor market-clearing condition. These conditions are summarized by the trade balance $w_i x_i = \sum_{j\in\mathcal{N}} X_{ij}$, or (D.5) with $e(x_i, w_i) = \mu(w_i x_i + l_i)$. By adding up (D.5) for the Krugman model, we see $\sum_{i\in\mathcal{N}} w_i x_i = \frac{\mu}{1-\mu} L$, which constrains the total income of mobile workers at any configuration $\boldsymbol{x}$. The existence and uniqueness of the solution for (D.5) follow from standard arguments (e.g., Facchinei and Pang, 2007). Given the solution $\boldsymbol{w}(\boldsymbol{x})$ of (D.5), we have the indirect utility of mobile workes is given by $v_i = \Delta_i^{\frac{\mu}{\sigma-1}} w_i$, where $\Delta_i \equiv \sum_{k\in\mathcal{N}} x_k w_k^{1-\sigma} d_{ki}$.

To satisfy Assumption S, let $l_i = l \equiv \frac{L}{N}$ for all $i \in \mathcal{N}$. We have

$$\nabla \log \boldsymbol{v}(\bar{\boldsymbol{x}}) = \underbrace{\frac{\mu}{\sigma-1}\mathbf{M}^{\top}\operatorname{diag}[\boldsymbol{x}]^{-1}}_{\bar{v}^{-1}\tilde{\mathbf{V}}_x} + \underbrace{\left(\mathbf{I} - \mu\mathbf{M}^{\top}\right)\operatorname{diag}[\boldsymbol{w}]^{-1}}_{\bar{v}^{-1}\tilde{\mathbf{V}}_w}\mathbf{W}_x = \frac{1}{\bar{x}}\frac{\mu}{\sigma-1}\bar{\mathbf{D}} + \frac{1}{\bar{w}}\left(\mathbf{I} - \mu\bar{\mathbf{D}}\right)\mathbf{W}_x, \quad \text{(D.13)}$$

where (D.4) and (D.8) give $\mathbf{W}_x$. By plugging $\delta = \frac{\mu(\bar{w}\bar{x}+l)}{\bar{w}\bar{x}} = 1$ (as $\bar{w}\bar{x} = \frac{\mu}{1-\mu} l$) and $\epsilon_x = \epsilon_w = \mu$ to (D.8),

$$\mathbf{W}_x = \frac{\bar{w}}{\bar{x}} \underbrace{\left(\sigma\mathbf{I} - \mu\bar{\mathbf{D}} - (\sigma-1)\bar{\mathbf{D}}^2\right)^{-1}}_{\bar{x}\mathbf{S}_w^{-1}} \underbrace{\bar{\mathbf{D}}\left(\mu\mathbf{I} - \bar{\mathbf{D}}\right)}_{\bar{w}^{-1}\mathbf{S}_x}. \quad \text{(D.14)}$$

By Fact B.2, (D.13) and (D.14) imply $\mathbf{V} = \bar{x}\nabla \log \boldsymbol{v}(\bar{\boldsymbol{x}}) = G^{\flat}(\bar{\mathbf{D}})^{-1} G^{\sharp}(\bar{\mathbf{D}})$, where we define

$$G^{\sharp}(\chi) \equiv \mu\left(\frac{1}{\sigma-1} + \frac{1}{\sigma}\right)\chi - \left(\frac{\mu^2}{\sigma-1} + \frac{1}{\sigma}\right)\chi^2, \quad \text{(D.15)}$$

$$G^{\flat}(\chi) \equiv 1 - \frac{\mu}{\sigma}\chi - \frac{\sigma-1}{\sigma}\chi^2. \quad \text{(D.16)}$$

Remark D.1. In Figure 9, we set $\mu = 0.5$, $\sigma = 10$, and $L = 8$. ■

Remark D.2 (Example 6.6)**.** To obtain $G^{\natural}$ for $\boldsymbol{l} = (l_i)_{i\in\mathcal{N}}$, we need to evaluate $\mathbf{V}_l = -\tilde{\mathbf{V}}_w \mathbf{S}_w^{-1}\mathbf{S}_l$ since $\mathbf{A} = \frac{l}{\bar{v}}\mathbf{V}_l$. From Example D.3, $\mathbf{S}_l = -\mu\bar{\mathbf{D}}$. Also, $\tilde{\mathbf{V}}_w = \bar{v}\frac{\partial}{\partial \boldsymbol{w}} \log \boldsymbol{v}(\bar{\boldsymbol{x}}) = \frac{\bar{v}}{\bar{w}}(\mathbf{I} - \mu\bar{\mathbf{D}})$ and $\tilde{\mathbf{V}}_l = \mathbf{0}$. Thus,

$$G^{\natural}(\chi) = c\frac{\chi(1-\mu\chi)}{G^{\flat}(\chi)} > 0 \quad \text{(D.17)}$$

where $c = \frac{l}{\bar{w}}\frac{\mu}{\sigma} = \frac{1-\mu}{\sigma}\bar{x} > 0$. It then follows that

$$\delta(\chi) = -\frac{\bar{x}}{\bar{a}}\frac{G^\natural(\chi)}{G(\chi)} = -\frac{c\bar{x}}{\bar{a}}\frac{\chi(1-\mu\chi)}{G^\sharp(\chi)}. \tag{D.18}$$

Straightforward algebra veryfies that $\delta'(\chi) < 0$ if $G^\sharp(\chi) > 0$. ■

D.2.2 Helpman (1998) model (Examples 3.2 and 6.5)

Helpman (1998) removed the A-sector in the Krugman model and assumed that all workers are mobile. Instead of the A-sector, the model introduces the housing (abbreviated as H) sector. Each region i is endowed with a fixed stock a_i of housing. Workers' preference is Cobb–Douglas of M-sector CES aggregate C_i^{M} and H-sector goods C_i^{H}, $u_i = (C_i^{\mathrm{M}})^\mu (C_i^{\mathrm{H}})^\gamma$, where $\mu \in (0,1)$ is the expenditure share of the former and $\gamma = 1-\mu \in (0,1)$ is that for the latter. There are two variants for assumptions on how housing stocks are owned: *public landownership* (abbreviated as PL) and *local landownership* (LL). The original formulation by Helpman (1998) supposes housing stocks are equally owned by all workers (i.e., PL). The income of a worker in region i is the sum of the wage and an equal dividend $r > 0$ of rental revenue *over the economy*. On the other hand, Ottaviano et al. (2002), Murata and Thisse (2005), and Redding and Sturm (2008) assumed that housing stocks are locally owned (i.e., LL). The income of a worker in region i is the sum of the wage and an equal dividend of rental revenue *in each region*.

Regarding the market equilibrium conditions, the only difference from the Krugman model is regional expenditure e_i on M-sector goods in each region:

$$\text{[PL]} \quad e_i = \mu\,(w_i + r)\,x_i, \tag{D.19}$$

$$\text{[LL]} \quad e_i = w_i x_i, \tag{D.20}$$

and market wage is given as the solution for (D.5). For the LL case, $\boldsymbol{w}(\boldsymbol{x})$ is uniquely given up to normalization. The indirect utility function is, with $\Delta_i \equiv \sum_{j\in\mathcal{N}} x_j w_j^{1-\sigma}\phi_{ji}$ and $r > 0$,

$$\text{[PL]} \quad v_i = \left(\frac{x_i}{a_i}\right)^{-\gamma}(w_i + r)^\mu \Delta_i^{\frac{\mu}{\sigma-1}}, \tag{D.21}$$

$$\text{[LL]} \quad v_i = \left(\frac{x_i}{a_i}\right)^{-\gamma} w_i^\mu \Delta_i^{\frac{\mu}{\sigma-1}}. \tag{D.22}$$

Let $a_i = 1$ to satisfy Assumption S. We compute that

$$\mathbf{V} = \bar{x}\left(\frac{\mu}{\sigma-1}\mathbf{M}^\top \operatorname{diag}[\boldsymbol{x}]^{-1} + \hat{\mathbf{V}}_w \mathbf{W}_x - \gamma \operatorname{diag}[\boldsymbol{x}]^{-1}\right), \tag{D.23}$$

$$\text{where} \quad \text{[PL]} \quad \hat{\mathbf{V}}_w \equiv \mu\left(\operatorname{diag}[\boldsymbol{w} + r\mathbf{1}]^{-1} - \mathbf{M}^\top \operatorname{diag}[\boldsymbol{w}]^{-1}\right), \tag{D.24}$$

$$\text{[LL]} \quad \hat{\mathbf{V}}_w \equiv \mu\left(\mathbf{I} - \mathbf{M}^\top\right)\operatorname{diag}[\boldsymbol{w}]^{-1}, \tag{D.25}$$

and $\mathbf{M}$ is defined by (D.6). From (D.7), $\mathbf{V} = G^\flat(\bar{\mathbf{D}})^{-1} G^\sharp(\bar{\mathbf{D}})$ with

$$G^\sharp(\chi) \equiv -\gamma + \mu\left(\frac{1}{\sigma-1} + \frac{1}{\sigma}\right)\chi - \left(\left(\frac{\mu^2}{\sigma-1} + \frac{1}{\sigma}\right) - \gamma\right)\chi^2, \tag{D.26}$$

$$G^\flat(\chi) \equiv 1 - \frac{\mu}{\sigma}\chi - \frac{\sigma-1}{\sigma}\chi^2 \tag{D.27}$$

for the PL case, whereas for the LL case

$$G^\sharp(\chi) \equiv (1-\chi)\left(-\gamma + \left(\mu\left(\frac{1}{\sigma-1} + \frac{1}{\sigma}\right) - \gamma\frac{\sigma-1}{\sigma}\right)\chi\right), \tag{D.28}$$

$$G^\flat(\chi) \equiv (1-\chi)\left(1 + \frac{\sigma-1}{\sigma}\chi\right). \tag{D.29}$$

Remark D.3. Equilibrium is unique when $\gamma\sigma = (1-\mu)\sigma > 1$ (Redding and Sturm, 2008). For both PL and LL, it implies that $G^\sharp(\chi) < 0$ for all $\chi \in (0,1)$. ■

Remark D.4 (Example 6.5)**.** The regional model in §3 of Redding and Rossi-Hansberg (2017) is a variant of the Helpman model with LL, in which variable input of mobile labor depends on region i (i.e., productivity differs across regions). The cost function of firms in region i is given by $C_i(q) = w_i(\alpha + \beta_i q)$. The market equilibrium condition for this case is, with $a_i \equiv \beta_i^{1-\sigma} > 0$, given by

$$s_i(\boldsymbol{x}, \boldsymbol{w}) = w_i x_i - \sum_{j\in\mathcal{N}} \frac{x_i a_i w_i^{1-\sigma}\phi_{ij}}{\sum_{k\in\mathcal{N}} x_k a_k w_k^{1-\sigma}\phi_{kj}} w_j x_j = 0. \tag{D.30}$$

The payoff function is given by (D.25) with $\Delta_i = \sum_{k\in\mathcal{N}} x_k a_k w_k^{1-\sigma}\phi_{ki}$.

From Example D.2, $\mathbf{S}_a = -\frac{\bar{w}\bar{x}}{\bar{a}}(\mathbf{I} - \bar{\mathbf{D}})(\mathbf{I} + \bar{\mathbf{D}})$ as $\bar{e} = \bar{w}\bar{x}$. Also, we have $\tilde{\mathbf{V}}_w = \frac{\bar{v}}{\bar{w}}\mu(\mathbf{I} - \bar{\mathbf{D}})$, $\tilde{\mathbf{V}}_a = \frac{\bar{v}}{\bar{a}}\frac{\mu}{\sigma-1}\bar{\mathbf{D}}$, and $\mathbf{S}_w = \sigma\bar{x}G^\flat(\bar{\mathbf{D}})$. Since $\mathbf{V}_a = \tilde{\mathbf{V}}_a - \tilde{\mathbf{V}}_w\mathbf{S}_w^{-1}\mathbf{S}_a$ and $\mathbf{A} = \frac{\bar{a}}{\bar{v}}\mathbf{V}_a = G^\natural(\bar{\mathbf{D}})$, we compute

$$G^\natural(\chi) = c\frac{(\sigma-1) + \sigma\chi}{G^\flat(\chi)} > 0 \tag{D.31}$$

where $c \equiv \frac{\bar{v}}{\bar{a}}\frac{\mu}{\sigma} > 0$. This in turn implies

$$\delta(\chi) = -\frac{\bar{x}}{\bar{a}}\frac{G^\natural(\chi)}{G(\chi)} = -\frac{c\bar{x}}{\bar{a}}\frac{(\sigma-1) + \sigma\chi}{G^\sharp(\chi)} \tag{D.32}$$

where $G^\sharp(\chi)$ is that for the LL case (D.29). We have $\delta'(\chi) > 0$ for all χ whenever $(1-\mu)\sigma > 1$. ■

D.2.3 Pflüger and Südekum (2008) (PS) model (Example 3.6)

The Pflüger–Südekum model builds on Pflüger (2004) and introduces the housing sector (again denoted by H). In this model, a quasi-linear upper-tier utility is assumed: $u_i = C_i^{\mathrm{A}} + \mu\log C_i^{\mathrm{M}} + \gamma\log C_i^{\mathrm{H}}$. The

production cost for a firm in $i \in \mathcal{N}$ is $\alpha w_i + \beta q$. Then, $\boldsymbol{w}$ is given as follows:

$$w_i = \frac{\mu}{\sigma} \sum_{j \in \mathcal{N}} \frac{\phi_{ij}}{\sum_{k \in \mathcal{N}} \phi_{kj} x_k} (x_j + l_j). \tag{D.33}$$

The indirect utility of a mobile worker in region i is

$$v_i(\boldsymbol{x}) = \frac{\mu}{\sigma - 1} \ln[\Delta_i] - \gamma \ln \frac{x_i + l_i}{a_i} + w_i, \tag{D.34}$$

where $\Delta_i = \sum_{j \in \mathcal{N}} \phi_{ji} x_j$, and l_i and a_i denote the mass of immobile workers and amount of housing stock in region i, respectively. The nominal wage in region i is given by (D.33). Let $l_i = l$ and $a_i = a$ for all i to meet Assumption S. Then, we see that $\mathbf{V} = \frac{1}{\bar{v}} G^\sharp(\bar{\mathbf{D}})$ with

$$G^\sharp(\chi) = -\frac{\gamma}{1+L} + \mu \left(\frac{1}{\sigma - 1} + \frac{1}{\sigma} \right) \chi - \frac{\mu}{\sigma}(1+L)\chi^2. \tag{D.35}$$

Remark D.5. In Figure 12a and Figure 12b, we set $\mu = 0.4$, $\sigma = 2.5$, $L = 4$, $\gamma = 0.5$, and $a_i = 1$. ■

D.2.4 Allen and Arkolakis (2014) (AA) (Example 3.5)

The AA model is a perfectly competitive Armington (1969)-based framework with positive and negative local externalities. We consider a discrete-space version of the model and abstract away all exogenous differences across regions. In the model, productivity of a location is proportional to x_i^α with $\alpha > 0$, representing positive externalities. The market equilibrium condition is

$$s_i(\boldsymbol{x}, \boldsymbol{w}) = w_i x_i - \sum_{j \in \mathcal{N}} \frac{w_i^{1-\sigma} x_i^{\alpha(\sigma-1)} \phi_{ij}}{\sum_{k \in \mathcal{N}} w_k^{1-\sigma} x_k^{\alpha(\sigma-1)} \phi_{kj}} w_j x_j = 0. \tag{D.36}$$

With market wage $\boldsymbol{w}$, the payoff function is $v_i(\boldsymbol{x}) = x_i^\beta w_i \Delta_i^{\frac{1}{\sigma-1}}$ with $\Delta_i \equiv \sum_{k \in \mathcal{N}} w_k^{1-\sigma} x_k^{\alpha(\sigma-1)} \phi_{ki}$; x_i^β ($\beta < 0$) represents negative externalities from congestion. We have $\mathbf{V} = G^\flat(\bar{\mathbf{D}})^{-1} G^\sharp(\bar{\mathbf{D}})$ with

$$G^\sharp(\chi) = -\left(\alpha + \beta - \frac{1+\alpha}{\sigma} \right) + \left(\alpha + \beta + \frac{1-\beta}{\sigma} \right) \chi, \tag{D.37}$$

$$G^\flat(\chi) = (\sigma + (\sigma - 1)\chi)(1 - \chi). \tag{D.38}$$

Remark D.6. In Figure 10, $\alpha = 0.5$, $\beta = -0.3$, and $\sigma = 6$. In Figure 18a, $\beta = -0.6$ so that $\alpha + \beta < 0$. ■

E Numerical examples for Proposition 2

This section provides numerical examples for **Proposition 2** for the $N = 8$ case. As in Example 6.1, we multiply the payoff in region 1 by $a_1 \geq 1$, whereas we let $a_i = 1$ for all $i \neq 1$. We consider the Krugman and Allen–Arkolakis (AA) models, and basic model parameters are the same as Figure 9 and Figure 10

except that region 1 has exogenous advantage.

Figure 18a reports the evolutionary paths of x_1 for the AA model (Class II) under the uniqueness of the equilibrium. The curves depicts region 1's population share, x_1, at stable equilibria against ϕ. Four incremental settings $a_1 \in \{1.000, 1.001, 1.005, 1.010\}$ are considered, including the baseline case with no location-fixed advantage ($a_1 = 1.000$). We have $\delta'(\chi) > 0$ for all $\chi \in (0,1)$ and see that $x_1 - \bar{x} > 0$ when $a_1 > 1$ and $x_1 - \bar{x}$ increases as a_1 increases, which are intuitive. Additionally, $x_1 - \bar{x}$ decreases as ϕ increases. We confirm that $\rho(\phi) > 0$ and $\rho'(\phi) < 0$ for all ϕ.

Similarly, Figure 18b and Figure 18c considers, respectively, the Krugman and AA models under multiplicity of equilibria. Unlike in the AA model, the Krugman model admits multiple equilibria for some ϕ for *any* pair of the structural parameters (μ, σ). **Proposition 2** correctly predicts the sign of $\rho'(\phi)$ for the range of ϕ such that $\bar{x}$ is stable when $a_1 = 1$; we have $\rho'(\phi) > 0$ when $\phi \in (0, \phi^*)$ for the Km model, whereas $\rho'(\phi) < 0$ when $\phi \in (\phi^{**}, 1)$ for the AA model.

Although **Proposition 2** does not cover the ranges where $\bar{x}$ is unstable, it tends to predict the general tendency of the evolution of ρ. In Figure 18b, in particular, $\rho'(\phi) > 0$ holds true except for the transitional phase after ϕ^{**}.

In Figure 18b, the definition of ρ is modified for spatial patterns with unpopulated regions. For the range $\phi \in (\phi^*, \phi^{**})$, ρ is evaluated with respect to the four-centric pattern $(2\bar{x}, 0, 2\bar{x}, 0, 2\bar{x}, 0, 2\bar{x}, 0)$:

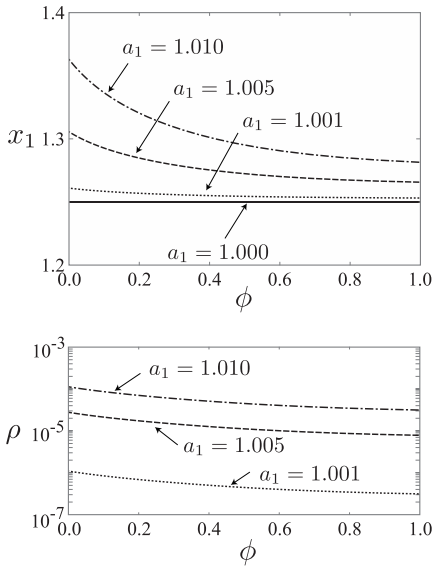
$$\rho \equiv \sum_{i \in \mathcal{I}(\boldsymbol{x})} (x_i - 2\bar{x})(a_i - \bar{a}(\boldsymbol{x})), \tag{E.1}$$

where $\mathcal{I}(\boldsymbol{x}) = \{i \in \mathcal{I} \mid x_i > 0\}$ is the set of populated regions, and $\bar{a}(\boldsymbol{x}) \equiv \frac{1}{|\mathcal{I}(\boldsymbol{x})|} \sum_{i \in \mathcal{I}(\boldsymbol{x})} a_i$. We define ρ for two-centric pattern $(4\bar{x}, 0, 0, 0, 4\bar{x}, 0, 0, 0)$ similary. For the transitional phase after ϕ^{**} we let

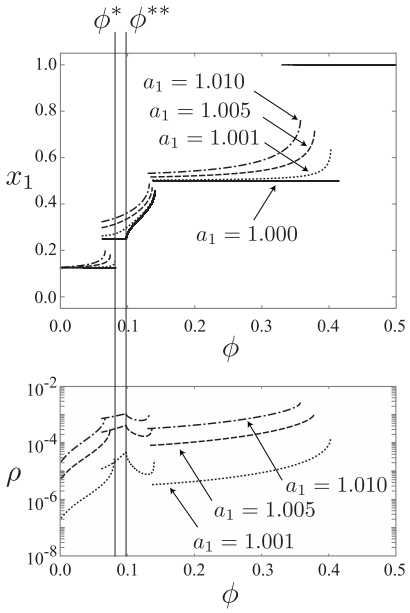
$$\rho \equiv \sum_{i \in \mathcal{I}(\boldsymbol{x})} (x_i - x_i^*)(a_i - \bar{a}(\boldsymbol{x})), \tag{E.2}$$

where x_i^* corresponds to the stable solution for $a_1 = 1$. Note that $\rho = (x_1 - 1)(a_1 - a_1) = 0$ for the complete monopolar pattern $(1, 0, 0, 0, 0, 0, 0, 0)$. It is natural that we have $\rho'(\phi) > 0$ for the four- and two-centric patterns because these patterns can be regarded as the uniform distribution on the four- and two-region cases, respectively.

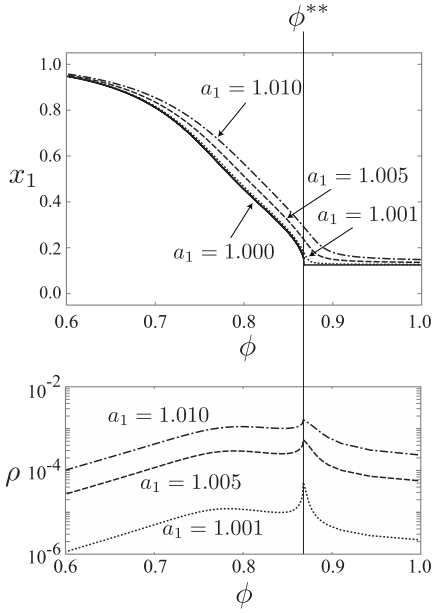
For Figure 18c, we employ (E.2) as the definition of ρ for the case $\phi \in (0, \phi^{**})$, i.e., we consider the deviation from the baseline equilibrium ($a_1 = 1$). We observe that $\rho'(\phi) < 0$ does not necessarily hold true for $\phi \in (0, \phi^{**})$. For instance, $\rho'(\phi) > 0$ when local dispersion force is relatively weak (when ϕ is small). Nonetheless, when ϕ is relatively large, we see ρ is generally decreasing in ϕ except for the vicinity of the bifurcation point ϕ^{**}.



(a) The Allen–Arkolakis model (unique equilibrium)



(b) The Krugman model



(c) The Allen–Arkolakis model

Figure 18: Population share of the advantageous region 1 and covariance ρ

References

Akamatsu, T., Fujishima, S., and Takayama, Y. (2017a). Discrete-space agglomeration model with social interactions: Multiplicity, stability, and continuous limit of equilibria. *Journal of Mathematical Economics*, 69:22–37.

Akamatsu, T., Mori, T., Osawa, M., and Takayama, Y. (2017b). Spatial scale of agglomeration and dispersion: Theoretical foundations and empirical implications. Unpublished manuscript, MPRA Paper No. 84145.

Akamatsu, T., Mori, T., Osawa, M., and Takayama, Y. (2019). Endogenous agglomeration in a many-region world. Unpublished manuscript, arXiv:1912.05113.

Akamatsu, T., Takayama, Y., and Ikeda, K. (2012). Spatial discounting, fourier, and racetrack economy: A recipe for the analysis of spatial agglomeration models. *Journal of Economic Dynamics and Control*, 99(11):32–52.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Allen, T., Arkolakis, C., and Takahashi, Y. (2019). Universal gravity. *Journal of Political Economy*, Forthcoming.

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press, Cambridge, MA.

Anas, A., Arnott, R., and Small, K. A. (1998). Urban spatial structure. *Journal of Economic Literature*, 36(3):1426–1464.

Anas, A. and Kim, I. (1996). General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. *Journal of Urban Economics*, 40(2):232–256.

Anderson, S. P., de Palma, A., and Thisse, J. F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.

Arcuri, P. and Murray, J. D. (1986). Pattern sensitivity to boundary and initial conditions in reaction-diffusion models. *Journal of Mathematical Biology*, 24(2):141–165.

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

Armington, P. S. (1969). A theory of demand for product distinguished by place of production. *International Monetary Fund Staff Papers*, 16(1):159–178.

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press, Ann Arbor, MI.

Baldwin, R. E., Forslid, R., Martin, P., Ottaviano, G. I. P., and Robert-Nicoud, F. (2003). *Economic Geography and Public Policy*. Princeton University Press.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, railroads and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448.

Beckmann, M. J. (1976). Spatial equilibrium in the dispersed city. In Papageorgiou, Y., editor, *Mathematical Land Use Theory*, pages 117–125. Lexington Book.

Behrens, K. and Murata, Y. (2018). On quantitative spatial economic models. *Higher School of Economics Research Paper No. WP BRP*, 184.

Blanchet, A., Mossay, P., and Santambrogio, F. (2016). Existence and uniqueness of equilibrium for a spatial model of social interactions. *International Economic Review*, 57(1):36–60.

Bleakley, H. and Lin, J. (2012). Portage and path dependence. *The Quarterly Journal of Economics*, 127(2):587–644.

Brown, G. W. and von Neumann, J. (1950). Solutions of games by differential equations. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games I*, pages 73–80. Princeton University Press.

de Palma, A., Papageorgiou, Y. Y., Thisse, J.-F., and Ushchev, P. (2019). About the origin of cities. *Journal of Urban Economics*, 111:1–13.

Dingel, J. I., Meng, K. C., and Hsiang, S. M. (2018). The spatial structure of productivity, trade, and inequality: Evidence from the global climate. Unpublished manuscript.

Dupuis, P. and Nagurney, A. (1993). Dynamical systems and variational inequalities. *Annals of Operations Research*, 44(1):7–42.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J. V. and Thisse, J.-F., editors, *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. North-Holland.

Duranton, G. and Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 79(4):1407–1440.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Facchinei, F. and Pang, J.-S. (2007). *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media.

Forslid, R. and Ottaviano, G. I. P. (2003). An analytically solvable core-periphery model. *Journal of Economic Geography*, 33(3):229–240.

Fujita, M. and Krugman, P. R. (1995). When is the economy monocentric?: von thuünen and chamberlin unified. *Regional Science and Urban Economics*, 25(4):505–528.

Fujita, M., Krugman, P. R., and Mori, T. (1999a). On the evolution of hierarchical urban systems. *European Economic Review*, 43(2):209–251.

Fujita, M., Krugman, P. R., and Venables, A. (1999b). *The Spatial Economy: Cities, Regions, and International Trade*. Princeton University Press.

Fujita, M. and Ogawa, H. (1982). Multiple equilibria and structural transformation of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12:161–196.

Fujita, M. and Thisse, J.-F. (2013). *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth (Second Edition)*. Cambridge University Press.

Golubitsky, M. and Stewart, I. (2003). *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space*, volume 200. Springer Science & Business Media.

Harris, B. and Wilson, A. G. (1978). Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environment and Planning A*, 10(4):371–388.

Helpman, E. (1998). The size of regions. In Pines, D., Sadka, E., and Zilcha, I., editors, *Topics in Public Economics: Theoretical and Applied Analysis*, pages 33–54. Cambridge University Press.

Hirsch, M. W., Smale, S., and Devaney, R. L. (2012). *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press.

Hofbauer, J. and Sandholm, W. H. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294.

Hofbauer, J. and Sandholm, W. H. (2007). Evolution in games with randomly disturbed payoffs. *Journal of economic theory*, 132(1):47–69.

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis (Second Edition)*. Cambridge University Press.

Hummels, D. and Skiba, A. (2004). Shipping the good apples out? an empirical confirmation of the alchian-allen conjecture. *Journal of Political Economy*, 112(6):1384–1402.

Ikeda, K., Akamatsu, T., and Kono, T. (2012a). Spatial period-doubling agglomeration of a core–periphery model with a system of cities. *Journal of Economic Dynamics and Control*, 36(5):754–778.

Ikeda, K. and Murota, K. (2014). *Bifurcation Theory for Hexagonal Agglomeration in Economic Geography*. Springer.

Ikeda, K., Murota, K., and Akamatsu, T. (2012b). Self-organization of lösch's hexagons in economic agglomeration for core-periphery models. *International Journal of Bifurcation and Chaos*, 22(08):1230026.

Ikeda, K., Murota, K., Akamatsu, T., Kono, T., and Takayama, Y. (2014). Self-organization of hexagonal agglomeration patterns in new economic geography models. *Journal of Economic Behavior & Organization*, 99:32–52.

Ikeda, K., Murota, K., Akamatsu, T., and Takayama, Y. (2017a). Agglomeration patterns in a long narrow economy of a new economic geography model: Analogy to a racetrack economy. *International Journal of Economic Theory*, 13:113–145.

Ikeda, K., Murota, K., and Takayama, Y. (2017b). Stable economic agglomeration patterns in two dimensions: Beyond the scope of central place theory. *Journal of Regional Science*, 57(1):132–172.

Ikeda, K., Onda, M., and Takayama, Y. (2018a). Bifurcation theory of a racetrack economy in a spatial economy model. *Networks and Spatial Economics*, pages 1–26.

Ikeda, K., Onda, M., and Takayama, Y. (2018b). Spatial period doubling, invariant pattern, and break point in economic agglomeration in two dimensions. *Journal of Economic Dynamics and Control*, 92:129–152.

Irarrazabal, A., Moxnes, A., and Opromolla, L. D. (2015). The tip of the iceberg: A quantitative framework for estimating trade costs. *Review of Economics and Statistics*, 97(4):777–792.

Kondo, S. and Miura, T. (2010). Reaction–diffusion model as a framework for understanding biological pattern formation. *Science*, 329(5999):1616–1620.

Krugman, P. R. (1991a). History versus expectations. *The Quarterly Journal of Economics*, 106(2):651–667.

Krugman, P. R. (1991b). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.

Krugman, P. R. (1993). On the number and location of cities. *European Economic Review*, 37(2):293–298.

Kuznetsov, Y. A. (2004). *Elements of Applied Bifurcation Theory (3rd Eds.)*. Springer-Verlag.

Matsuyama, K. (2017). Geographical advantage: Home market effect in a multi-region world. *Research in Economics*, 71:740–758.

Meinhardt, H. and Gierer, A. (2000). Pattern formation by local self-activation and lateral inhibition. *Bioessays*, 22(8):753–760.

Mertikopoulos, P. and Sandholm, W. H. (2018). Riemannian game dynamics. *Journal of Economic Theory*, 177:315–364.

Michaels, G. and Rauch, F. (2018). Resetting the urban network: 117–2012. *The Economic Journal*, 128(608):378–412.

Mossay, P. and Picard, P. M. (2011). On spatial equilibria in a social interaction model. *Journal of Economic Theory*, 146(6):2455–2477.

Murata, Y. and Thisse, J.-F. (2005). A simple model of economic geography à la helpman–tabucbi. *Journal of Urban Economics*, 58(1):137–155.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.

Osawa, M. and Akamatsu, T. (2020). Equilibrium refinement for a model of non-monocentric internal structures of cities: A potential game approach. *Journal of Economic Theory*, 187:105025.

Osawa, M., Akamatsu, T., and Takayama, Y. (2017). Harris and wilson (1978) model revisited: The spatial period-doubling cascade in an urban retail model. *Journal of Regional Science*, 57(3):442–466.

Ottaviano, G. I. P., Tabuchi, T., and Thisse, J.-F. (2002). Agglomeration and trade revisited. *International Economic Review*, 43:409–436.

Papageorgiou, Y. Y. and Smith, T. R. (1983). Agglomeration as local instability of spatially uniform steady-states. *Econometrica*, pages 1109–1119.

Pflüger, M. (2004). A simple, analytically solvable, chamberlinian agglomeration model. *Regional Science and Urban Economics*, 34(5):565–573.

Pflüger, M. and Südekum, J. (2008). Integration, agglomeration and welfare. *Journal of Urban Economics*, 63(2):544–566.

Picard, P. M. and Tabuchi, T. (2013). On microfoundations of the city. *Journal of Economic Theory*, 148(6):2561–2582.

Proost, S. and Thisse, J.-F. (2019). What can be learned from spatial economics? *Journal of Economic Literature*, 57:575–643.

Puga, D. (1999). The rise and fall of regional inequalities. *European Economic Review*, 43(2):303–334.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9:21–58.

Redding, S. J. and Sturm, D. (2008). The cost of remoteness: Evidence from german division and reunification. *American Economic Review*, 98(5):1766–1797.

Sandholm, W. H. (2010). *Population Games and Evolutionary Dynamics*. MIT Press.

Smith, M. J. (1984). The stability of a dynamic model of traffic assignment: An application of a method of Lyapunov. *Transportation Science*, 18(3):245–252.

Statistics Bureau, Ministry of Internal Affairs and Communications of Japan (1970). Population census.

Statistics Bureau, Ministry of Internal Affairs and Communications of Japan (2015). Population census.

Tabuchi, T. (1986). Urban agglomeration economies in a linear city. *Regional Science and Urban Economics*, 16(3):421–436.

Tabuchi, T. (1998). Urban agglomeration and dispersion: A synthesis of alonso and krugman. *Journal of Urban Economics*, 44(3):333–351.

Tabuchi, T. and Thisse, J.-F. (2011). A new economic geography model of central places. *Journal of Urban Economics*, 69(2):240–252.

Takayama, Y. and Akamatsu, T. (2011). Emergence of polycentric urban configurations from combination of communication externality and spatial competition. *Journal of JSCE Series D3: Infrastructure Planning and Management*, 67(1):001–020.

Taylor, P. D. and Jonker, L. B. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156.

Turing, A. M. (1953). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society (part B)*, 237:37–72.