Travel Circle: A Model of Supply Chains

Lau, C. Oscar

Massey University

6 November 2022

Travel Circle: A Model of Supply Chains

By C. Oscar Lau*

Draft: December 17, 2022

Travel circle is a metaphor for supply chains: in travel circle, travelers are transported by carriers in multiple legs from the center to diverse destinations on the circumference; in supply chains, goods are transformed in multiple stages by firms from natural resources into differentiated products. The model is generated using only three cost parameters. At the start of supply chains, a few firms mass-produce standardized commodities at low unit costs; at the end, many firms produce distinctive products in small scales at high unit costs. As an extension, the circle's size is endogenized to account for consumers' preferences for varieties.

JEL: L11, L13, L23, L25

Keywords: Product differentiation, Scale economies, Entry game

Consider a distribution network. As goods move from a central warehouse to regional distribution centers, then to local centers, then to retailers' storerooms, and finally to the shelves, three related trends occur in parallel. First, the locations of the storage facilities get increasingly localized because the goods have to ultimately make their way to consumers, who are diversely located. Second, as the locations decentralize, the number of storage facilities increases, and their capacities and delivery sizes shrink. Third, as the scale of operations shrinks, the unit cost of inventory and transportation soars.

The central warehouse operates with massive volumes to exploit economies of

* School of Economics and Finance, Massey University, New Zealand (Email: c.l.lau@massey.ac.nz). I thank Jay Pil Choi, Thomas Jeitschko, Arijit Mukherjee and seminar participants at Massey University for their helpful comments.

scale. In contrast, the retail outlets have to shelf the goods in small batches and sell them individually to consumers, leading to much higher unit costs of distribution. If the consumers choose home deliveries instead of in-store shopping, the individual orders have to be delivered to their exact addresses, which is also an expensive operation commonly known as "last mile" delivery.

The distribution network exhibits a central-to-local pattern. The premise of this paper is that production of goods and services also follows a similar generic-to-specific pattern, in which generic natural resources are transformed into increasingly distinctive intermediate goods step by step, and are eventually differentiated into unique final goods. As the intermediate goods become more distinctive, the scale of production decreases, and so does the economies of scale.

This paper develops a model of supply chains that follows this generic-to-specific pattern. The supply chains are constructed using only three fundamental cost parameters. Given the final goods, the intermediate goods are endogenously generated. The degree of differentiation, the number of firms, and their levels of investments are determined for all the final and intermediate goods. As an extension, even the final goods themselves are endogenized by incorporating consumers' tastes for varieties.

This paper contributes to the industrial organization literature in two ways. First, as just described, the model generates multiple layers of firms and goods in supply chains endogenously, demonstrating the generic-to-specific pattern of increasing differentiation. In doing so, the model also offers new perspectives on differentiation, scale economies, and welfare analysis of markets with differentiated goods. Second, the model makes two contributions in methodologies: (i) it introduces a cost function that links a firm's investment to its marginal cost, thus endogenizing the investment; (ii) in the extension, it offers a new approach to modeling consumers' preference for varieties, as an alternative to Dixit-Stiglitz utility functions.

Productions as differentiation

All final goods begin their journey of creation as natural resources. Natural resources are typically mass processed in vast scales into commodities such as oils and metals. Commodities are usually highly standardized and versatile; they can take shape in a wide variety of intermediate goods. Aluminum alloys, for instance, can be cast into anything from car engines to cellphone bodies, or rolled into anything from soda cans to aluminum foils.

Once the intermediate goods have taken shape—as engines cast from aluminum, as semiconductors sliced from silicon, as textiles woven from cotton threads, etc.—they are inevitably far less versatile than the commodities from which they are made. However, they are usually still generic enough to be used in multiple products, and thus are still produced in sizable scales. Producers specializing in intermediate goods take advantage of the economies of scale by producing large quantities, and the economies of scope by producing closely related goods. They then sell their outputs to multiple buyers for use in multiple products, so each intermediate good finds its way into many products. For example, the same engine can be used to power many different models of vehicles, the same memory microchip can be put inside a variety of electronic gadgets; the same textile can be cut into any shape of cloths. Some intermediate goods exist exactly for their versatility: bolts and nuts for constructions, buttons and zippers for garments, etc.

The further down the supply chain, the more distinctive the intermediate goods become. Finally, the goods are differentiated into finished goods. These final steps require certain components to be designed, developed, and manufactured specifically for the product, e.g., body panels for a car model's unique exterior design; chassis for a computer's particular form factor; clothing parts for a garment's distinctive styling in specific sizes. These components are unique to the product and are therefore produced in relatively low volumes. The more customized a product is, the smaller the production scale, and the higher the cost. Thus tailored suits cost more than off-the-rack suits, custom-made cakes cost more than pre-made cakes, and unique houses cost more than cookie-cutter houses, even if the customized items use the same quality of materials or ingredients as their standardized counterparts.

The generic-to-specific pattern applies not only to goods, but to services as well. Consider the training that economics students receive, from undergraduate to PhD level. ECON101, covering standard economic principles, is taught to hundreds of students in a lecture hall. Then, an undergraduate elective course focuses on a particular subject for a medium-sized class. Next, a graduate field course is even more specialized for an even smaller class. Lastly, in the research stage, a PhD advisor works closely with the individual student on their unique research agenda. Along the process, the subject gets increasingly specialized, the class size shrinks, and the per-student cost soars. Early education exhibits a general-to-specific pattern too: a child would learn general subjects like math and reading at school, then go to after-school classes for particular music instruments or sports classes, and then get lectured about family history and personal values by parents at dinner.

A similar pattern is replicated in many other services. In health care, general practitioners provide general and preventive care to many patients, and if necessary, refer them to specialists for specialized and personalized treatments. In news service, local news outlets receive international and national news from major news agencies such as the Associated Press and Reuters, and then add their own reporting for local news. The pattern is not restricted only to professional services. General services such as table waiting, banking, and hair dressing all have standardized elements, but they can also be made more personalized at extra cost. Actually, services are in general more individualized than physical goods because many of them necessarily involve personal interactions. (The stark contrast in costs between personalized service and mass-produced goods is a vivid demonstration of the Baumol effect (Baumol and Bowen 1965), which predicts that costs of personal services would rise relative to costs of mass manufacturing because productivity of the former benefits little from mass production technologies.)

To sum up, production can be seen as an elaborate process in which generic inputs differentiate into increasingly specific outputs with decreasing scale of production.

Before proceeding, I should briefly mention new trade theory (Krugman 1980, Lancaster 1980, Helpman 1981), which, like the travel circle model, is premised on the notions of differentiation and scale economies. New trade theory offers an explanation of how scale economies drive trade flows between countries with similar productivities and factor endowments. In essence, consumers prefer diversity of goods, and bilateral trading allows both countries' firms to sell their varieties to markets beyond their own home, thus taking advantage of economies of scale. In contrast, the travel circle model focuses on the intricacies of the supply chains, e.g., how differentiation takes place in multiple stages of production and culminate in the final goods.

Productions as journeys

Much like the distribution network, a transportation network also exhibits a central-to-local pattern; so transportation can be viewed as an analogy for the generic-to-specific pattern in supply chains. In fact, the essence of the generic-to-specific pattern can be illustrated by a multi-leg trip from a major hub to a local destination, for instance from Dallas airport in Texas to the Economics Department of Michigan State University. The traveler would first take a large plane to fly from Dallas (a major hub) to Detroit, Michigan (a regional hub), then take a small plane to Lansing, Michigan (a local hub), then catch a bus to the bus station in East Lansing, and finally hire a taxi to the Economics Department of MSU. Along the journey, the transit point gets increasingly distinctive and localized, the size of the carrier shrinks, the distance of the leg drops, and the cost per passenger-mile soars. A single-passenger taxi ride would cost at least a few dozen times more than a long-haul flight in per passenger-mile term.[1]

Figure 1 shows the stylized travel pattern. There are several large planes, each flying from the major hub (the center) to a regional hub. Each traveler will board the plane that is heading to the regional hub closest to their destination. So the travelers split into several pools. When a large plane reaches its regional hub, the travelers split again into several smaller pools, each of which will take a small plane to the local hub closest to their destinations. Then each pool splits yet again into even smaller pools to take buses to different local towns. Finally each individual

[1]The taxi ride is expensive not only because the route is customized, but also because the service timing is personalized—the taxi stands by for taking passengers whenever they please. Passengers wait for the plane, but taxis wait for passengers most of the time.

takes taxis to their exact destinations.

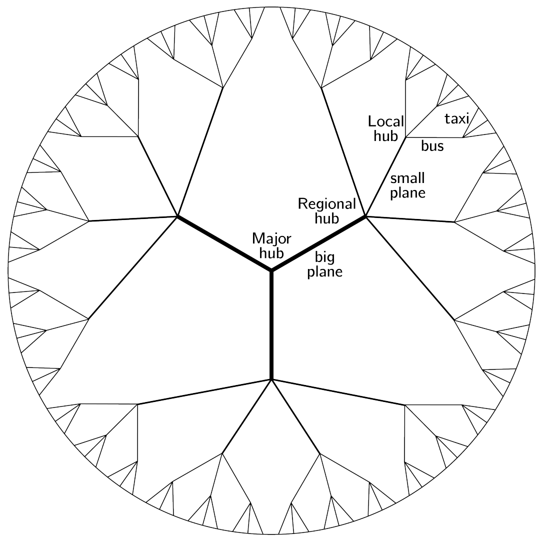
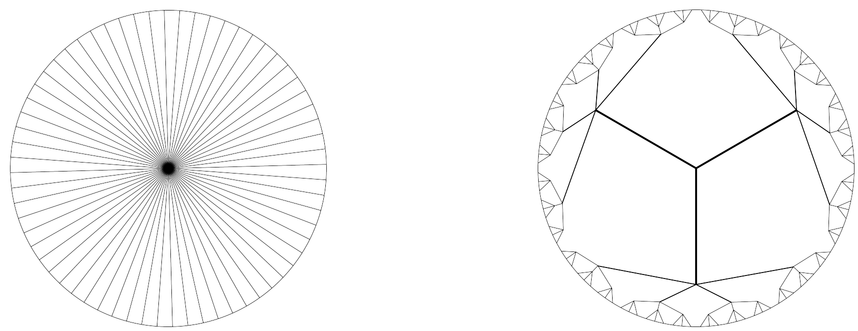


Figure 1. : Travel in legs

The total transportation cost is minimized by optimizing the level of pooling, i.e., the economies of scale, in various legs of the journey. Clearly zero pooling throughout the journeys would not be optimal. In Figure 2a, each traveler takes a taxi from the center all the way to their destination, which is prohibitively costly. It corresponds to zero economies of scale.[2] Figure 2b shows a travel pattern that gravitate towards the other direction. The first leg, which has the highest degree of pooling, is elongated, so that the overall pooling is higher than that in Figure 1. Which one of the two configurations is more efficient depends on the particular circumstances.

The traveler's journey is a metaphor for a product's journey of creation. The traveler moves from one point to another towards the destination; the good transforms from one intermediate good to another towards the final good. The travelers are moved by carriers; the goods are transformed by machines. A large crowd is

[2]Renting and self-driving a car would be much cheaper than taking a taxi, but it would still cost a lot of time and effort. More importantly, the rental car and the road network are themselves results of mass productions. In a world truly devoid of any mass production, the travelers will be left with nothing more than the most rudimentary equipment and their own feet.



(a) Zero pooling (b) Long first leg

Figure 2. : Alternative routes

carried by a large plane at low per passenger-mile cost; a commodity is produced en masse at low marginal cost. The large crowd splits into smaller and smaller groups, and ultimately individuals; the commodity differentiates into more and more distinctive goods, and ultimately final goods. An individual traveler is chauffeured to their specific destination at high cost; an intermediate good is refined into a unique product at high cost.

I. Setup

The travel circle is a spatial differentiation model built on Salop's (1979) circle. Salop's circle focuses on product differentiation of final goods, skipping over the supply chains behind. In his model, all firms are located on the circle, leaving the circle's interior empty. The travel circle model fills in the void by adding multiple layers of circles inside the original one, with each layer representing one stage of transformation of goods. It opens up the supply chains by covering the full production process from start (center) to finish (circumference).[3] While Salop's circle models product differentiation directly, the travel circle uses traveling as an analogy

[3]I am not aware of any other similar multiple-level model. However, there are several models in which firms may be located at the center of the circle in addition to on the circumference, with the centrally located firms interpreted as mail-order or online sellers (Balasubramanian 1998, Bouckaert 2000, Loginova 2009, and Madden and Pezzino 2011). In another direction, Chen and Riordan (2007), followed by Firgo, Pennerstorfer and Weiss (2015), study a "spokes model." In the spokes model, there are a number of spokes with equal lengths, each extending from the center to a unique point. The end of each spoke represents one variety of good; and the consumers are uniformly distributed on the spokes.

for production.

Basics: To begin building the travel circle model, consider a circle with a radius of one mile (think of it as a very long "mile"). A unit mass of travelers concentrate at the center. They all must go home. Their homes are uniformly distributed on the circle.

Each traveler's journey is broken down into L legs. Imagine that $L-1$ concentric circles of increasing sizes are nested inside the original circle (see Figure 3). Call them, from small to big, circle 1, circle 2, ..., circle $(L-1)$; and call the original circle circle L. Define circle 0 as the center.

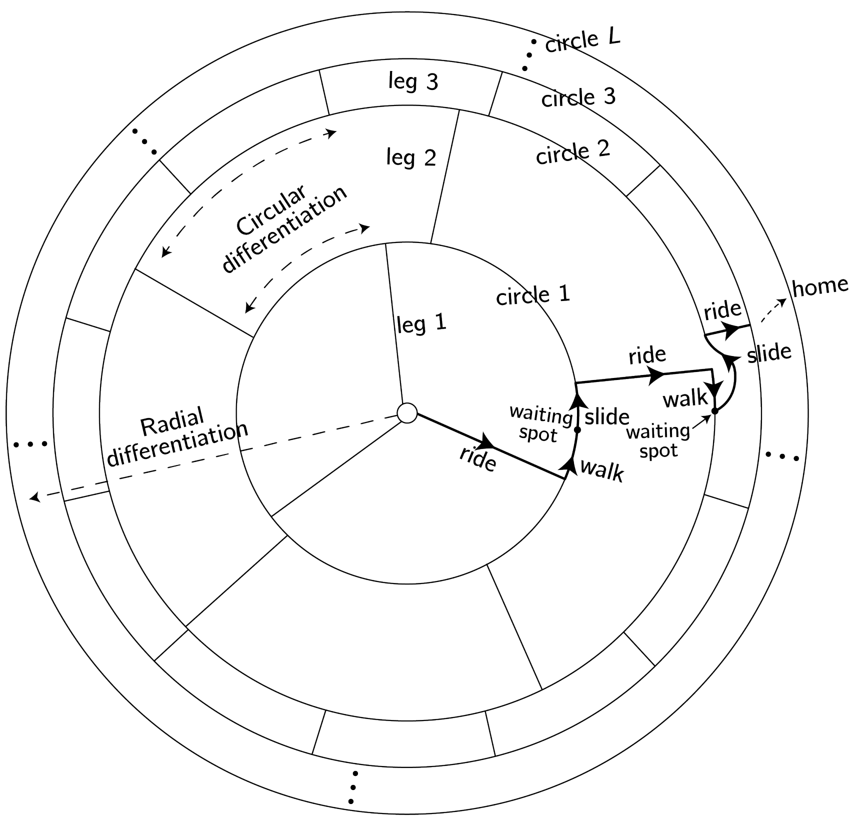


Figure 3. : Travel circle

For $l = 1, 2, ..., L$, leg l covers a radial distance (i.e. distance along a radius) of r_l miles from circle $(l-1)$ to circle l, so $\sum_{l=1}^{L} r_l = 1$. Each leg can be served by multiple routes, represented by the "spokes" in Figure 3. Each route is operated by a different firm, which runs one carrier. A leg-l firm runs its carrier radially from its starting point on circle $(l-1)$ to its ending point on circle l. All carriers in the same leg depart from their starting points at the same time, and reach their ending points at the same time. All carriers run only once. The entire transportation system enclosed in circle L is known as the travel circle (although it is really a disc).

Travelers: As mentioned above, a unit mass of travelers have to travel from the center to their homes, which are uniformly distributed on circle L. Every traveler's objective is to go home at the lowest cost. Each traveler's journey proceed as follows (see the arrowed path in Figure 3 as an example). From the center, they board one of the leg-1 carriers, which takes them to its ending point on circle 1. They then get off and walk along the circle to their designated waiting spot for their leg-2 carrier. Every traveler's waiting spot is unique. When their leg-2 carrier arrives at its starting point on circle 1, they slide along the circle to the carrier and board. Then the carrier takes them to circle 2, and they walk circularly again to their designated waiting spot for their leg-3 carrier. Then their leg-3 carrier arrives, and they again slide and board. (In the example in Figure 3, the traveler walks clockwise and then slides counter-clockwise on circle 2.) The process continues until their leg-L carrier takes them to circle L. From there they walk home circularly (unless the carrier happens to stop at exactly their home).

Sliding is effortless, walking is not. As just described, the travelers can slide from the waiting spots to the carriers, but before that they have to walk to their waiting spots. They cannot slide to their waiting spots (or homes in case of the final leg) because of a "crashing condition" assumed in the model: Travelers sliding in opposite directions will crash into each other and die; but if they walk, they can walk around each other to avoid crashing. As we will see in the next subsection, the waiting spots for passengers of a ride are randomly assigned within a waiting zone allocated to the ride. So when the travelers go to their waiting spots, some of them go clockwise and others counter-clockwise, depending on where they got off their previous ride and where they will wait for their next. Therefore, they will crash into each other if they slide. But then why can they slide from their waiting spot to their carrier without crashing? This is because, again as we will see in the next subsection, the waiting spots are assigned such that passengers of each carrier will wait on either side of its starting point, separate from passengers of other carriers. When a carrier arrives, its passengers waiting on either side will slide into it in tandem and in the same direction, so they do not risk crashing into each other. (The motivation for introducing the crashing condition will become apparent in the next section. It will be addressed in a remark at the end of the proof for Proposition 1.)

The total cost to a traveler is the sum of the ticket prices for all the rides they take plus their walking costs. The ticket prices will be set by the firms that run them. The walking cost is w per mile for a unit measure of travelers, where $w > 0$. Sliding is costless.

Firms: A large number of identical firms consider whether to enter the transportation market. Each firm is allowed to run at most one route in the whole travel circle. To enter, a firm has to make an investment to buy a carrier. The firm decides the amount of investment, which can be any positive number. All firms face constant marginal costs; but their magnitudes can be different. A firm's marginal cost depends on how much it invests. If it invests f, then its marginal cost with respect to passenger-miles is $\frac{\alpha}{f^\beta}$, where α and β are both positive parameters. So if it invests f to carry x passengers through a radial distance of m miles, its total cost is

$$c(f, x, m) = f + \frac{\alpha}{f^\beta} \cdot xm, \tag{1}$$

where f is the firm's (fixed) investment cost, and $\frac{\alpha}{f^\beta} \cdot xm$ is its (variable) operating cost.

A firm can invest in any carrier, from a rickshaw to a taxi to a bus to an airplane. The bigger the investment, the lower the marginal cost. A rickshaw is cheap, but it is backbreaking to pull. An airplane is expensive, but it costs little to carry an additional passenger-mile.

The investment's impact on the marginal cost is shaped by α and β. First, α scales the magnitude of the marginal cost relative to the investment cost; α is the marginal cost when $f = 1$. Second, β is the elasticity of marginal cost with respect to investment: $\frac{\partial \ln(\alpha/f^{\beta})}{\partial \ln f} = -\beta$.

In this setup, investment is not some exogenous entry cost, as in Salop (1979). Instead, each firm chooses the optimal investment to minimize its total cost (as Lemma 2 will show). So the model features endogenous investments.

Travelers cannot invest—only firms can. As mentioned earlier, travelers can walk circularly. They can also, in principle, walk radially. But if they do, they face the same cost equation (1) as firms for their radial walks. With zero investment, their marginal cost of walking radially would be infinite. Thus in practice they have to take carriers for radial movements.

The game

Many identical firms play a complete information entry game. The game has L stages, running in descending order from leg L to 1. Each stage has two steps. The two steps for leg l proceed as follows.

Step 1: Entry, investment, and assignment of carrier locations

In Step 1, potential entrants simultaneously decide whether to enter leg l and, if so, how much to invest in a carrier. Each firm can at most enter one leg and run one ride in the entire travel circle.

Following Salop (1979), all adjacent routes in the leg are assumed to be equidistant from each other. Also following Salop, the entrants do not choose their routes' locations. Subject to equidistance, the exact starting points of the routes on circle $(l-1)$, or correspondingly their ending points on circle l, are drawn randomly with uniform distribution along the circle. The locations are then revealed to the entrants and travelers. The drawings for all legs are mutually independent events.

Step 2: Price competition, sales, and assignment of waiting spots

In Step 2, firms compete in prices simultaneously; and travelers choose their leg-l ride and buy tickets. All tickets are non-refundable and non-transferable.

After all travelers have bought their leg-l tickets, each ride is allocated an arc of the circle as its waiting zone, with the mid-point of the waiting zone set at the ride's starting point. Each waiting zone's share of the circle equals the ride's market share in the leg. So if all rides have the same number of passengers, the waiting zones will divide up the circle's circumference equally.[4] Each traveler is then randomly assigned a waiting spot on their ride's waiting zone, with uniform distribution along the zone.

Upon completion of all stages of the game, the actual traveling takes place, from leg 1 to leg L. Leg l begins with the travelers sitting inside the carriers of their leg-l rides, waiting for the carriers to start, and ends with the travelers boarding their leg-$(l+1)$ ride (except for leg L, which ends with the travelers arriving home).

A subgame perfect Nash equilibrium will be identified by Proposition 1 in the next section.

Interpretations

The travel circle is an abstraction of supply chains. The center symbolizes natural resources. The legs are stages of production. The carriers are facilities, machines and equipment. So as the travelers are transported from the center to increasingly local destinations by carriers; the goods are transformed from natural resources to increasingly distinctive products by machines. The travelers' walking represents adaptation and customization made for the particular intermediate good to prepare it for mass production (to be elaborated in Subsection VI.A).

[4]The circumference of each circle is just "wide" enough to accommodate all travelers. If the rides have different market share, there will be overlaps and gaps between the waiting zones, which means that the waiting zones of some rides will not be long enough to accommodate all their passengers. This is not a concern for the purpose of this paper, because I focus only on equilibria in which all firms in the same leg have identical sales. But for completeness the following rule of waiting zone allocation will be followed if the rides have different sales: Start with the longest zone. (If more than one zones are tied for the longest, then randomly pick one of them.) If it has overlaps or gaps with any of its two neighboring zones, then move the neighboring zone(s) clockwise or counter-clockwise so that the overlaps or gaps just disappear. Repeat the process further in both directions circularly until all overlaps and gaps disappear.

Each journey represents a final product. Each journey is unique, so is each product. It is tempting to equate a destination with a final product. But a final product is the aggregation of all the steps taken to bring it to fruition, so the product encompasses the whole journey, not just the destination. Different routes have different costs, even if they lead to the same destination. Just like life, a product is a journey, not a destination.

The travel circle can be thought as the supply chains for a particular good category, e.g., computers. But it can also be viewed as the aggregate production system of the whole economy in a nutshell (again, see Subsection VI.A for details).

In the model, a traveler buys tickets directly from multiple carriers. Meanwhile, in reality, a consumer pays one final price for the good to the retailer, who passes part of the proceeds to its suppliers, who do the same to their suppliers, and so on. The operational procedures are different, but the end results are the same. But if one wants to reconcile even the operational difference, one could imagine that in the model there is a zero-cost travel agent automaton who handles all the ticket purchases for the travelers and charges each of them one full price, and then passes the proceeds down the supply chain.

Circular and radial differentiation

A final good is the cumulation of differentiation carried out at various intermediate goods, just as a journey is the aggregate of many legs. As labeled on Figure 3, there are two ways to view differentiation in the travel circle: circular and radial.

Circular differentiation corresponds to the familiar notion of spatial differentiation as per Salop (1979). (I consider only horizontal differentiation, not vertical differentiation.) The concept is applied to intermediate goods as well as final products.

In the travel circle, the degree of circular differentiation for each leg is measured by the extent of entry, i.e., the number of firms that enter the leg. The more firms that enter, the more localized the routes become, and hence the higher the degree of differentiation. Circular differentiation can literally denote spatial diversity, or more generally represent varieties in attributes such as colors, features, and designs.

Meanwhile, radial differentiation traces the increase in cumulative differentiation across stages: generic commodities differentiate into intermediate goods, which in turn differentiate into finished products. Through the process, works-in-progress are shaped, sliced, divided and combined repeatedly, thus becoming more distinctive in each stage, and culminating in final goods eventually.

In the travel circle, the pattern of radial differentiation is captured by the partition of the radius into multiple legs of various lengths. The number of legs equals the number of steps of radial differentiation. The length of each leg indicates the degree of radial differentiation. A long first leg, for instance, takes the travelers a long way from the center to highly local areas. (see the discussion following Proposition 3 for a more general interpretation of the lengths of legs).

Circular and radial differentiation progress in parallel, much as water ripples spread circularly and radially in tandem. As a carrier travels radially, it achieves a higher degree of radial differentiation. But as it pulls forward, it is also approaching a circle with a higher degree of circular differentiation. For instance, turning aluminum into engines represents radial differentiation, but the variety of the resulting engines exhibits circular differentiation.

The remainder of the paper is organized as follows. Section II solves for circular differentiation, with radial differentiation taken as given. Using that result, Section III in turn endogenizes radial differentiation. Section IV then combines the two dimensions of differentiation to complete the travel circle and characterizes its properties. Section V presents an extension in which travelers can choose how far to travel. Finally, Section VI interprets the model in more detail and explores the insights it offers.

II. Circular Differentiation

In this section, I solve for circular differentiation, i.e., the extent of entry, for all legs. The number of legs L and the lengths of all legs r_l, for all $l = 1, 2, ..., L$, are taken as given.

Suppose n_l firms will enter the market for leg l. Since all firms are identical, and the routes will be located symmetrically, it makes sense to look for an equilibrium in which all entrants for leg l invest the same amount f_l and set the same price p_l for their tickets (with p_l being the price charged to a population of one).

Let R_l be the circumference of circle l, so $R_l = 2\pi \sum_{k=1}^{l} r_k$.

LEMMA 1 (Salop 1979): *In equilibrium, the number of entrants in leg l is* $n_l = \sqrt{\frac{wR_l}{f_l}}$ *(zero-profit condition).*

Proofs of all results, except for Theorem 1, are relegated to the Appendix. As in Salop (1979), free entry leads to zero profit (up to the integer constraint). A higher investment cost means that in order to break-even, each entrant needs more sales or a higher markup, both of which would happen only if fewer firms enter. Therefore, n_l is inversely related to f_l. (I assume that the investment is low enough such that there are at least three entrants, so that each entrant is competing with two surrounding firms.)

The investment cost in Salop (1979) is exogenous. In contrast, firms in the travel circle choose the optimal investment to minimize their total cost, given by the cost equation (1).

LEMMA 2: *The total cost borne by a firm to transport x travelers through m miles is minimized by investing $f^{min}(x,m) = (\alpha\beta xm)^{\frac{1}{\beta+1}}$ (cost-minimization condition).*

The cost function is $c^{min}(x,m) := c\left(f^{min}(x,m), x, m\right) = \left(1+\frac{1}{\beta}\right)(\alpha\beta xm)^{\frac{1}{\beta+1}}$. Also, f^{min} is increasing in β iff $xm < \frac{\exp\left(\frac{1}{\beta}+1\right)}{\alpha\beta}$.

The average cost per passenger-mile is $\frac{c^{min}(x,m)}{xm} = \left(1+\frac{1}{\beta}\right)\left(\frac{\alpha\beta}{x^\beta m^\beta}\right)^{\frac{1}{\beta+1}}$, which is decreasing in passenger-mile, thus exhibiting economies of scale. The endogenized marginal cost is $\frac{dc^{min}(x,m)}{d(xm)} = \left[\frac{\alpha}{(\beta xm)^\beta}\right]^{\frac{1}{\beta+1}}$, which is also decreasing in passenger-mile.

Note that f^{min} is not monotonically increasing in β. An increase in β has two effects on f^{min}, as shown by the cost-minimization condition: the β inside the parentheses encourages more investment, but the β in the exponent reins it in when $\alpha\beta xm > 1$.

The cost-minimization condition shows that the lower the sales, the lower the optimal investment. In equilibrium, all entrants have a sales volume of $1/n_l$; so the more entrants there are, the less each of them will sell, and the less each of them will invest. Therefore, like the zero-profit condition in Lemma 1, the cost-minimization condition in Lemma 2 also relates n_l to f_l negatively. When put together, the two conditions jointly determine the leg's market equilibrium allocation, defined below alongside with the leg's efficient allocation.

DEFINITION 1: *A* circular differentiation program *(CDP) for leg l, denoted by (n_l, f_l), specifies the number of firms n_l and the firm-level investment f_l in the leg.*

The market CDP *for leg l, denoted by (n_l^*, f_l^*), specifies the equilibrium allocation with free entry, with n_l^* and f_l^* determined by the parameters $\{\alpha, \beta, w, r_l, R_l\}$. The* efficient CDP *for leg l, denoted by (n_l^{**}, f_l^{**}), specifies the efficient allocation, with n_l^{**} and f_l^{**} determined by the same set of parameters.*

PROPOSITION 1 (Market CDP): *The equilibrium number of entrants in leg l is*

$$n_l^* = \left(\frac{w^{\beta+1} R_l^{\beta+1}}{\alpha\beta r_l} \right)^{\frac{1}{2\beta+1}}, \tag{2}$$

with each entrant investing

$$f_l^* = \left(\frac{\alpha^2 \beta^2 r_l^2}{w R_l} \right)^{\frac{1}{2\beta+1}}. \tag{3}$$

The total investment in leg l is thus

$$n_l^* f_l^* = \left(\alpha\beta r_l w^\beta R_l^\beta \right)^{\frac{1}{2\beta+1}}. \tag{4}$$

The equilibrium price in leg l is

$$p_l^* = \Big(\underbrace{1}_{inv.} + \underbrace{\frac{1}{\beta}}_{op.}\Big) n_l^* f_l^* \tag{5}$$

with markup ratio β. Finally, the market total cost of leg l (firms' costs plus walking costs) is

$$C_l^* = \Big(\underbrace{1}_{inv.} + \underbrace{\frac{1}{\beta}}_{op.} + \underbrace{\frac{1}{4}}_{walk}\Big) n_l^* f_l^*. \tag{6}$$

In the parentheses above, inv., op., and walk are due to investment costs, operating costs, and walking costs respectively.

With the equilibrium outcomes solved, an equilibrium for the entry game is also identified. A subgame perfect Nash equilibrium is: n_l^* firms enter leg l, each investing f_l^* and setting the price at p_l^*, for $l = 1, 2, ..., L$.

Consider the effects of α and r_l. A higher marginal cost scaler α or a longer ride distance r_l amplifies the return of investment, therefore inducing a higher f_l^* (equation (3)). A higher f_l^* in turn means that the market can only sustain fewer entrants, i.e., a smaller n_l^* (equation (2)). The higher α or r_l leads to a higher total investment $n_l^* f_l^*$ as the negative effect on n_l^* is overwhelmed by the positive effect on f_l^* (equation (4)).

Next, consider the effects of w and R_l. For a given number of entrants, a higher walking cost w or a longer circumference R_l (which is proportional to the total walking distance for the given number of entrants) allows the firms to command a higher markup. So a higher w or R_l means more entrants can be accommodated in the market, i.e., a higher n^* (equation (2)). Intuitively, if the walking costs are high, either due to a high w or a long R_l, the market will populate the leg with many firms to reduce the walking distances. However, a higher n^* implies that each entrant will face a lower demand and therefore invest less, i.e., a lower f^* (equation

(3)). Like α and r_l, a higher w or R_l also leads to a higher $n_l^* f_l^*$, but this time around it is because their negative effect on f_l^* is overwhelmed by their positive effect on n_l^* (equation (4)).

As shown by equation (5), every firm's operating cost is $1/\beta$ times its investment cost. The firm breaks even by recouping its investment cost through the markup, so the markup ratio is simply β, regardless of the leg.

The market total cost in equation (6) is also the travelers' total cost for the leg because the ticket prices they pay cover exactly the costs of investment and operation.

When w approaches infinity, or when α or r_l approaches zero, the number of entrants tends to infinity and the firm-level investment tends to zero. However, just because n^* approaches infinity does not mean that the market is competitive—the markup ratio is still β. The market will become competitive only if β approaches zero. But in such an environment, investment is virtually useless; and with nearly zero investment, the cost of traveling radially is astronomical. In effect, the economy has almost zero economies of scale and every traveler essentially walks all the way from the center to home at punishing cost.

Although the price is set above the marginal cost, the markup introduces no distortions on the consumption side because the travelers have inelastic demands. Distortions come only from the production side. The next proposition compares the efficient allocation with the market equilibrium allocation.

PROPOSITION 2 (Efficient CDP): *The efficient number of entrants in leg l is*

$$n_l^{**} = \frac{1}{2}\left(\frac{w^{\beta+1} R_l^{\beta+1}}{2\alpha\beta r_l}\right)^{\frac{1}{2\beta+1}} = \frac{1}{2^{1+\frac{1}{2\beta+1}}} \cdot n_l^* \in \Big(\underbrace{\frac{n_l^*}{4}}_{\beta\to 0}, \underbrace{\frac{n_l^*}{2}}_{\beta\to\infty}\Big), \tag{7}$$

with each entrant investing

$$f_l^{**} = \left(\frac{4\alpha^2\beta^2 r_l^2}{wR_l}\right)^{\frac{1}{2\beta+1}} = 4^{\frac{1}{2\beta+1}} f_l^* \in (\underbrace{f_l^*}_{\beta\to\infty}, \underbrace{4f_l^*}_{\beta\to 0}). \tag{8}$$

The efficient total investment in leg l is thus

$$n_l^{**} f_l^{**} = \frac{1}{2}\left(2\alpha\beta r_l w^\beta R_l^\beta\right)^{\frac{1}{2\beta+1}} = \frac{1}{2^{1-\frac{1}{2\beta+1}}} \cdot n_l^* f_l^* \in \Big(\underbrace{\frac{n_l^* f_l^*}{2}}_{\beta\to\infty}, \underbrace{n_l^* f_l^*}_{\beta\to 0}\Big). \tag{9}$$

The efficient total cost of leg l (firms' costs plus walking costs) is

$$C_l^{**} = \Big(\underbrace{1}_{inv.} + \underbrace{\frac{1}{\beta}}_{op.} + \underbrace{1}_{walk}\Big) n_l^{**} f_l^{**} = \widetilde{\beta} C_l^* \in \Big(\underbrace{\frac{4}{5} C_l^*}_{\beta\to\infty}, \underbrace{C_l^*}_{\beta\to 0}\Big) \tag{10}$$

where $\widetilde{\beta} = \frac{1}{2^{1-\frac{1}{2\beta+1}}} \frac{1/\beta+2}{1/\beta+5/4} \in \left(\frac{4}{5}, 1\right)$ *is strictly decreasing in* β.

Compared to the efficient allocation, the market under free entry "over-differentiates" by generating too many firms in the leg. Although each firm invests less than the efficient amount, in aggregate the firms invest too much. Firms under market allocation also incur higher aggregate operating cost, which equals $1/\beta$ times the aggregate investment cost under both allocations. However, due to over-differentiation, travelers walk less under market allocation, resulting in lower walking costs. In the market allocation, the total walking costs amount to only a quarter of the total investment cost (recall equation (6)), whereas in the efficient allocation the total walking cost is as much as the total investment cost. Although the market has a higher total investment cost, it is at most twice as much as that under efficient allocation. So overall, the total walking cost is lower under market allocation.

Notice that all distortions, when measured as ratios of the equilibrium quantities to their efficient counterparts, are determined solely by β. When β gets close to zero, the market is approximately competitive, as mentioned earlier. In this case, the equilibrium number of firms approaches four times the efficient number of firms, but each firm invests only about a quarter of the efficient level, resulting in about the same total investment as the efficient level. Despite that the market is over-differentiated, the total cost approximately matches the efficient level. (But again, the total cost is enormous when β approaches zero.)

Conversely, when β tends to infinity, the equilibrium number of firms is about double the efficient number of firms, and each firm invests close to the efficient level, so the total investment is nearly double the efficient level. The total cost is about a quarter higher than the efficient level, which represents the highest deadweight loss in proportional term.

In the standard Salop circle with fixed entry cost, the equilibrium number of firms is double the efficient number of firms, the total investment is double the efficient level, and the total cost is a quarter higher than the efficient cost.[5] In the present model with endogenous investment, more firms will enter because the investment will be driven down. Now the equilibrium number of firms is between double and quadruple the efficient number, the firm level investment is between a quarter and one time the efficient level, the total investment is less than double the efficient level, and the total cost is less than a quarter higher than the efficient cost. So under endogenous investment, in proportional terms over-differentiation is even more pronounced; but thanks to under-investments at firm level, the over-investment at aggregate level is alleviated, and so is the deadweight loss.

III. Radial Differentiation

A radial differentiation program (RDP) for the travel circle specifies the number of legs and their radial lengths. (The RDP will be defined formally shortly.) In the previous section, I study the entry of firms in each leg, taking the RDP as given. But how does the RDP arise in the first place? This section answers the question.

For the socially efficient outcome, obviously the social planner chooses an RDP that minimizes the total cost of the travel circle. For the market outcome, however, it is less apparent how the RDP comes about spontaneously. In reality, of course an industry's supply chain does not emerge suddenly. Rather, the supply chain develops and evolves over time, with firms entering and exiting continuously. I

[5] Refer to p. 284 of Tirole (1988). It is straightforward to work out that, using the textbook's notations, the efficient total cost is $\sqrt{tf}$, where t is the consumers' unit transport cost and f is the fixed entry cost, and the market total cost is $\frac{5}{4}\sqrt{tf}$.

make no attempt to model such development processes; rather my goal is to predict the resulting market RDP in a plausible and simple way.

With that goal in mind, I contend that it is reasonable to postulate that the market RDP minimizes the sum of the market total costs across all legs. In other words, the market RDP minimizes the total cost of the travel circle, subject to market allocations at the leg levels as detailed in the last section. The rationale is straightforward: if the prevailing RDP does not minimize the sum of market total costs, then there are opportunities for firms in adjacent legs to adjust the radial distances between them to achieve a lower total cost, and split the saving. Therefore, as the supply chain evolves, firms would always gravitate towards a more efficient RDP. For example, when an aluminum producer supplies aluminum to an automaker for car bodies, does it supply aluminum ingots (short distance for the leg), aluminum sheets made from ingots (medium distance), or aluminum sheets readily cut into shapes for the car models (long distance)? Obviously the firms would choose the option that minimizes their total cost, otherwise there is room for re-negotiation that would result in savings for both.

(The market total cost of each leg includes the walking costs as well as the firms' costs. But the walking costs are proportional to the firms' costs (equation 6). So by minimizing the firms' costs, the market RDP also minimizes the market total costs.)

One may wonder why this efficiency argument applies to radial differentiation, but not to circular differentiation in the last section. In other words, why does inefficiency arise from circular differentiation, but not from radial differentiation? The answer is that firms in different legs do not compete with each other, while firms in the same leg do. At the leg level, firms impose negative business-stealing externalities on each other, resulting in excessive entries and aggregate over-investment. There is no such competition in the radial direction, so the firms are not subject to the negative externalities in this direction.

There is also a secondary rationale for arguing that the market RDP tends to be efficient. In the real world, even a market oriented economy inevitably involves some central planning, most notably infrastructure planning. Firms would anchor their locations around the infrastructures. For instance, logistics firms would build warehouses based on the transportation network. Since the infrastructures are supposedly designed to maximize social efficiency, the firms that follow the infrastructures would also position themselves efficiently.

With the background for the market and efficient RDPs explained, we are ready for the following definitions and results.

DEFINITION 2: *A* radial differentiation program *(RDP) of the travel circle specifies the number of legs and their radial lengths. An RDP is denoted by* $(r_1, r_2, ..., r_L)$ *with* $r_l > 0$ *for every* $l = 1, ..., L$ *and* $\sum_{l=1}^{L} r_l = r$*, where* r *is the radius of the travel circle in miles, and* L *can be* ∞*. The RDP is* finite *if* L *is finite; it is* infinite *if* L *is* ∞*.*

A market RDP *is an RDP that minimizes* $\sum_{l=1}^{L} C_l^*$*. An* efficient RDP *is an RDP that minimizes* $\sum_{l=1}^{L} C_l^{**}$*.*

PROPOSITION 3 (Market and efficient RDPs): *The market RDP and the efficient RDP are both unique and infinite. For the radius of one mile, denote the two RDPs by* $(r_1^*, r_2^*, ...)$ *and* $(r_1^{**}, r_2^{**}, ...)$ *respectively. The lengths of their legs follow the algorithm*

$$\frac{r_l}{r_{l+1}} = -\frac{\widehat{r}_l}{2} + \sqrt{\left(\frac{\widehat{r}_l}{2}\right)^2 + \left(1 + \beta\widehat{r}_l\right)^{2+\frac{1}{\beta}}} \tag{11}$$

where

$$\widehat{r}_l := \frac{r_l}{\sum_{k=1}^{l} r_k} \text{ for } l = 1, 2, ...$$

Given the radius of the circle, the algorithm (11) produces the lengths of all legs, and hence the RDP. For the transportation and distribution industries, the lengths of legs can be interpreted literally as distances, so the RDP pinpoints the locations of the transportation and distribution centers along the networks. For goods and services in general, the lengths can be understood as the degree of transformation—

the longer the leg, the "bigger" the transformation, so the RDP identifies the "positions" of intermediate goods along the supply chains. Although there is no universal measure of the degree of transformation, the lengths of legs still have intuitive appeals. As item (i) of Theorem 2 will assert, the lengths of legs always decrease along the journeys. Turning silicon into silicon wafers is a bigger transformation than turning wafers into microchips in the sense that more "taking shape" occurs in the first process—wafers are unrecognizable from raw silicon, but microchips are more recognizable from wafers. Likewise, turning wafers into microchips is a bigger transformation than merely assembling the microchips into computers.

The RDPs $(r_1^*, r_2^*, ...)$ and $(r_1^{**}, r_2^{**}, ...)$ are identical because their lengths of legs follow the same algorithm and they both add up to one (the radius).[6] (But the radius of the circle is not restricted to one in Section V). Note that $(r_1^*, r_2^*, ...)$ and $(r_1^{**}, r_2^{**}, ...)$ are both dictated by β, so they can also be written as $(r_l^*(\beta))_{l=1}^{\infty}$ and $(r_l^{**}(\beta))_{l=1}^{\infty}$. The two RDPs are infinite, but we can construct a finite RDP as an approximation. First iterate the algorithm (11) many times to obtain a sequence of ratios of lengths, then solve for the lengths by combining those ratios with the constraint that the lengths add up to one.

In typical infinite games such as repeated prisoners' dilemma, there is no definite final stage, so the game never ends. In contrast, the travel circle entry game starts with the final leg, but there is no such a final leg under an infinite RDP, so the game never begins. While a game does not necessarily have a definite end point, a game that does not begin is a non-starter. To address this issue, I contend that for all purposes, theoretical or practical, it is sufficient to use the approximate RDP described in the previous paragraph in place of the market and efficient RDPs. The approximate RDP is finite, so the game now has a final leg to begin with. It can be made arbitrarily close to the two RDPs, so the approximation is innocuous for any practical purpose. Actually the approximation is more realistic than the infinite RDPs because all journeys in reality have finite legs.

[6] Although they are identical, I stick with the r_l^* and r_l^{**} notations in order to continue the practice of using * to denote market and ** efficiency.

Another issue related to infinite RDPs is that the travelers will only get arbitrarily close to home no matter how many legs they go through, so technically they never reach home. Again, the issue can be resolved by using the approximate RDP instead. Under the approximate RDP, the carriers in the final leg take each traveler to a point on the circumference of the travel circle, and from there the traveler walks home through a tiny distance.

IV. Travel Circle

This section completes the circle.

DEFINITION 3: *A* travel circle program *(TCP), denoted by* $(n_l, f_l, r_l)_{l=1}^{L}$*, consists of an RDP of* L *legs, denoted by* $(r_1, r_2, ..., r_L)$*, and* L *number of CDPs, one for each leg, denoted by* (n_l, f_l) *for* $l = 1, ..., L$*, where* L *can be* ∞*.*

The market TCP*, denoted by* $(n_l^*, f_l^*, r_l^*)_{l=1}^{\infty}$*, is defined jointly by the market RDP* $(r_1^*, r_2^*, ...)$ *and the market CDPs* $(n_l^*, f_l^*)_{l=1}^{\infty}$*. The* efficient TCP*, denoted by* $(n_l^{**}, f_l^{**}, r_l^{**})_{l=1}^{\infty}$*, is defined jointly by the efficient RDP* $(r_1^{**}, r_2^{**}, ...)$ *and the efficient CDPs* $(n_l^{**}, f_l^{**})_{l=1}^{\infty}$*.*

So a TCP renders a full picture of the travel circle by specifying how many legs there are, how long each leg is, how many firms there are in each leg, and how much each firm invests. The market TCP depicts the market allocation; the efficient TCP depicts the efficient allocation.

THEOREM 1 (Travel circle): *Three cost parameters,* α*,* β*, and* w*, are sufficient to generate the market TCP and the efficient TCP.*

Proof: The theorem is a culmination of Propositions 1 (Market CDP), 2 (Efficient CDP), and 3 (Market and efficient RDPs). The RDPs of both TCPs are determined by β in accordance with algorithm (11). The market CDPs follow equations (2) and (3) in Proposition 1; the efficient CDPs follow equations (7) and (8) in Proposition 2. The CDPs are functions of α, β, w, r_l and R_l. But r_l and hence R_l are defined by the RDPs, which in turn are dictated by β. So the CDPs depend on α, β, and w only. ■

Recall that when the radius is fixed at one mile, the RDPs depend solely on β. So the market TCP can be written as $(n_l^*(\alpha,\beta,w), f_l^*(\alpha,\beta,w), r_l^*(\beta))_{l=1}^{\infty}$ and similarly the efficient TCP as $(n_l^{**}(\alpha,\beta,w), f_l^{**}(\alpha,\beta,w), r_l^{**}(\beta))_{l=1}^{\infty}$.

THEOREM 2 (Properties of travel circle): *The market TCP and the efficient TCP follow these properties when l increases from 1 to ∞:*

(i) r_l^* *and* r_l^{**} *decrease and approach* 0*;*

(ii) $\frac{r_l^*}{r_{l+1}^*}$ *and* $\frac{r_l^{**}}{r_{l+1}^{**}}$ *decrease and approach* 1*;*

(iii) n_l^* *and* n_l^{**} *increase and approach* ∞*;*

(iv) f_l^* *and* f_l^{**} *decrease and approach* 0*;*

(v) $n_l^* f_l^*$*,* C_l^**,* $n_l^{**} f_l^{**}$ *and* C_l^{**} *decrease and approach* 0*;*

(vi) p_l^* *decreases and approaches* 0*;*

(vii) $\frac{p_l^*}{r_l^*}$*,* $\frac{C_l^*}{r_l^*}$ *and* $\frac{C_l^{**}}{r_l^{**}}$ *increase and approach* ∞*.*

In early legs of the journeys, there are only a few firms, each investing in a big carrier to carry many passengers through long distances at low per-mile cost and price. Towards later legs, more and more firms invest in smaller and smaller carriers to carry fewer and fewer passengers through shorter and shorter distances to more and more specific locations at higher and higher per-mile costs and prices. This pattern is true for both the market allocation and the efficient allocation (except that there are no prices under the efficient allocation).

Theorems 1 and 2 are illustrated in Figure 4, which shows four pairs of travel circle, each corresponding to a set of values for the parameters α, β, and w. The upper circle in each pair represents market allocation and the lower one efficient allocation. The first pair is the baseline, while each of the other three pairs have one of the parameters altered from the baseline. For clarity, only the first five legs are shown for the first, second, and fourth pairs, all for which $\beta = 1$. For the third pair, where $\beta = 1.2$, only the first three legs are shown because they are already longer than the first five legs in the other pairs (recall that the RDP is dictated by β).[7]

The widths of the routes are proportional to the numbers of passengers; and the sizes of the dots at the end of the routes are proportional to the investments for the carriers. All travel circles are generated with only α, β, and w, in accordance with Theorem 1. They all display properties (i) to (iv) listed in Theorem 2. For each pair, the market circle has more firms in every leg than the corresponding efficient circle, with each firm investing less, as claimed by Proposition 2.

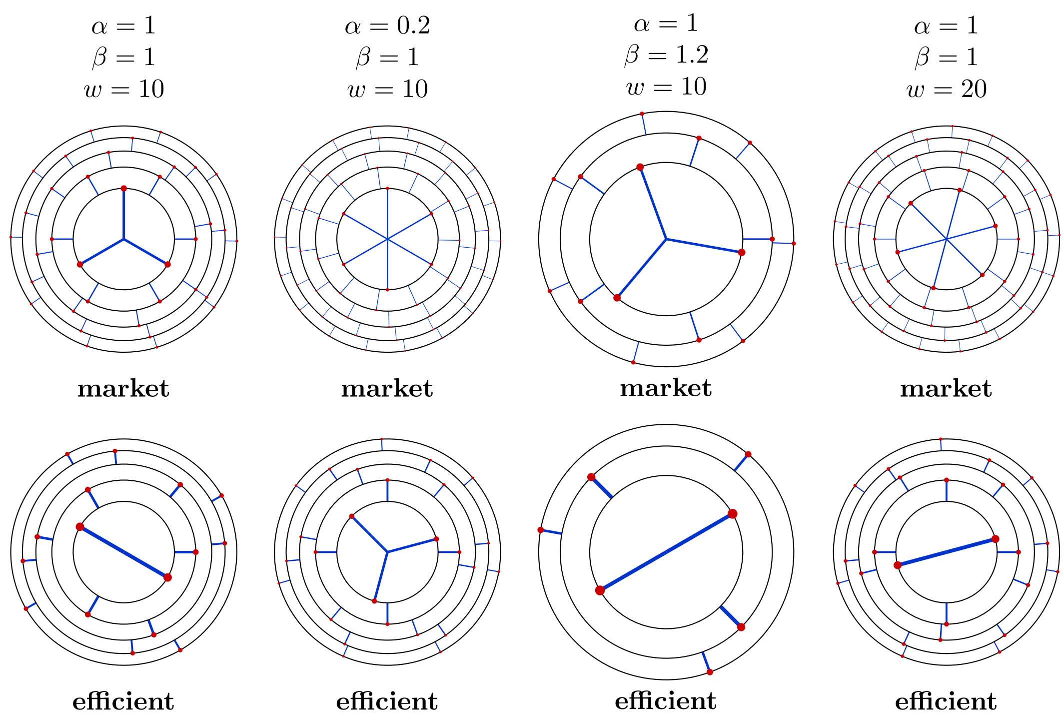


Figure 4. : Differentiation patterns

[7]The circles are created by iterating the RDP algorithm (11) until $\widehat{r}$ is smaller than 0.0001. The number of firms calculated is rounded *down* to the nearest integer for market allocations, and rounded *off* to the nearest integer for efficient allocations.

V. Extension: Endogenous Radius

So far, the model focuses only on the supply side. The travelers' demand is inelastic—all travelers simply must go home. This section extends the model by considering another scenario: instead of going home, the travelers at the center are going on vacations, and they can choose how far to travel. From now on, call the original model "homecoming" and this extended model "vacation." In the vacation model, each traveler can choose a vacation destination anywhere along the ray starting from the center and passing through what used to be the location of their home. The further they go, the happier they are. They can also consume a generic good along with their travel. All traveler's have the same Cobb-Douglas utility function

$$u(r, s) = r^{\theta} s^{1-\theta}, \tag{12}$$

where r is the radial distance in miles of their vacation destination from the center, s is the amount of the generic good they consume, and $\theta \in (0, 1)$ is a parameter.

I will focus on outcomes where all travelers choose the same distance. Suppose all travelers choose distance r in equilibrium, so that the vacation travel circle has a radius of r miles. Let $\mathcal{C}_r^*(\alpha, \beta, w)$ be the total cost of the travel circle (firms' costs plus walking costs) with radius r miles under market allocation. As in the homecoming model, the market total cost results from picking the RDP that minimizes the total cost of the travel circle. The rationale remains the same as that offered in Section III—through negotiations, firms in different legs would gravitate towards the cost-minimizing RDP to minimize their total cost. Also as before, the travelers fully bear this market total cost because the ticket prices they pay cover exactly the firms' costs, and they are the ones who do the walking. The next lemma implies that $\mathcal{C}_r^*$ is increasing and concave in r.

LEMMA 3: $\mathcal{C}_r^*(\alpha, \beta, w) = r^{\widehat{\beta}} \cdot \mathcal{C}_1^*(\alpha, \beta, w)$ *where* $\widehat{\beta} := \frac{1}{2} + \frac{1}{4\beta+2} \in \left(\frac{1}{2}, 1\right)$.

So far we think of the travelers as many individuals, but they can also be interpreted collectively as one representative consumer who chooses a combination of differentiated goods (traveling) and the generic good. So θ measures the consumer's preference for the diversity of goods. The radius of the travel circle represents the degree of differentiation of the differentiated goods. The bigger the circle, the more diverse (i.e., circularly differentiated) and the more refined (i.e., radially differentiated) are the myriad of differentiated goods they consume.

Suppose the representative traveler/consumer has a total income of I to spend on traveling and the generic good. Let the price of the generic goods be p_s. As mentioned, the cost of traveling is the sum of ticket prices and walking costs. (Income is "spent" on walking in the sense that the time and efforts spent on walking could have been used to earn more income.) Since the market total cost is a function of r, when the travelers choose r, in effect they are also choosing their traveling costs. The next proposition solves for the equilibrium radius.

PROPOSITION 4 (Endogenous market radius): *The market equilibrium radius of the vacation model is*

$$r^* = \left[\frac{I}{\mathcal{C}_1^* \left(1 + \frac{1-\theta}{\theta}\widehat{\beta}\right)} \right]^{\frac{1}{\beta}}.$$

Note that $\mathcal{C}_1^*$ is just the market total cost of the homecoming travel circle. Since $\mathcal{C}_1^*$ is a function of (α, β, w), r^* is a function of $(\alpha, \beta, w, \theta, I)$.

COROLLARY 1: *The market equilibrium radius r^* is increasing in θ and I. When $\theta \to 0$, $r^* \to 0$; when $\theta \to 1$, $r^* \to \left(\frac{I}{\mathcal{C}_1^*}\right)^{\frac{1}{\beta}}$.*

Travel circles with different radii can be used to illustrate differences across good categories or changes over time. For example, one might expect that a travel circle modeled for women's clothing would have a higher θ and thus longer radius than that for men's, showing more varieties and higher sophistication. Meanwhile, both circles would grow in size over time as I rises with economic growth and $\mathcal{C}_1^*$ declines with technological improvements.

Next, I find the market RDP, CDP, and TCP of the vacation model.

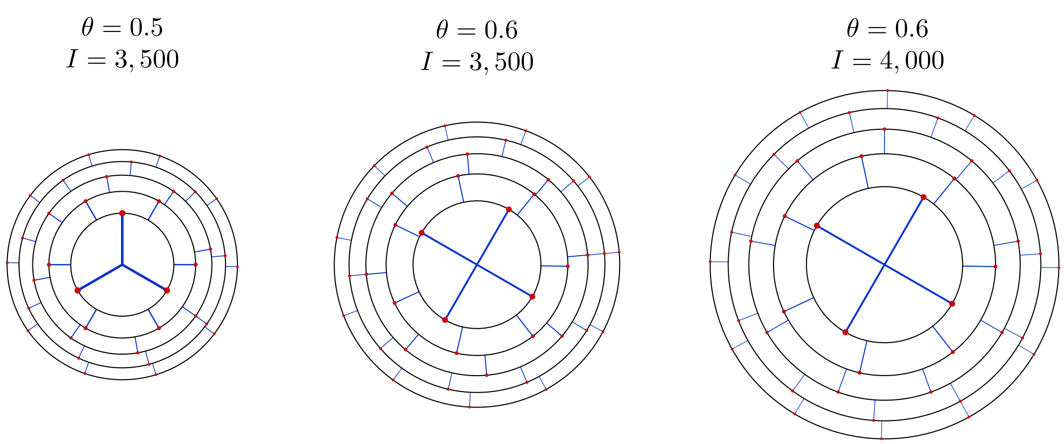
LEMMA 4: *The market RDP of the vacation model is* $(r^* r_1^*, r^* r_2^*, ...)$.

While the RDPs of the homecoming model depend only on β, the market RDP of the vacation model depend on all of α, β, w, and θ, due to its dependence on r^*.

With the market RDP figured out, the market CDP of the vacation model simply follow from Proposition 1. Therefore, the market TCP of the vacation model is also solved. Moreover, none of the properties for travel circles under market allocation listed in Theorem 2 rely on the radius being one, so they apply to the vacation model too. Following Theorems 1 and 2 for the homecoming model, the corresponding version for the vacation model is presented as follows.

PROPOSITION 5 (Vacation travel circle): *Five parameters,* α, β, w, θ, *and* I, *are sufficient to generate the market TCP for the vacation model. Moreover, the properties of TCPs listed in Theorem 2 for the homecoming model under market allocation apply to the TCPs for the vacation model too.*

Propositions 4 and 5 and Corollary 1 are illustrated in Figure 5.



All three circles are market allocations with $\alpha = 1$, $\beta = 1$, and $w = 10$.

Figure 5. : Endogenous radius

Figure 5 displays three vacation travel circles under market allocations in a similar way as Figure 4 does for the homecoming model. They all use the same baseline values for α, β, and w, but different values for θ and I. The second circle has a higher θ than the first; and the third one keeps the same θ as the second but has a higher I. The circles show that the radius of the circle is increasing in both θ and I.

VI. Discussion

A. Further interpretations

This subsection interprets several aspects of the model in detail.

Walking: The travelers travel en masse in carriers to benefit from economies of scale, but in between rides each of them has to walk individually to a unique waiting spot for their next ride. Similarly, goods are mass-produced, but in between mass productions an intermediate good needs to be prepared specifically to be ready for the next mass processing. Walking can therefore be interpreted as adaptation, customization, product design, and other efforts made specifically for an intermediate good to prepare it for the next stage of mass production. So radial movements are transformations; circular movements are preparations for transformations. For preparations, the costs are split between the buyer and the supplier of the intermediate good in the most efficient way. So the "walking" is done by the buyer, the supplier, or both. A short walk means the supplier's outputs fit the buyer's need closely and require little efforts to adopt. Conversely, a long walk means a lot of efforts are necessary.

A carrier has limited waiting spots, so it assigns specific spots to each passenger, who has to adapt by walking there. Similarly, a seafood monger supplying to local restaurants has a limited delivery window, so they assign a specific delivery time for each restaurant, which has to work around it. This is an example of adaptation by the buyer. An example of adaptation by the seller is that a laundry contractor for hotels collects and return the linens at times that fit the hotels' housekeeping schedules. As for customization, examples include: an aluminum can producer, which supplies to multiple soda companies, print different graphics on the cans for each beverage; a retailer switching to a new point-of-sales system coverts their files and data into a compatible format; an automaker buying microchips from a microchip manufacturer write customized programs on the microchips for their vehicles.[8] The customization is done by the supplier in the first example, and by the buyer in the other two. Lastly, product-specific design is infused into various stages of production, just as walking takes place in various legs. The design jobs can be done by contractors in the supply chain. However, the owners of the brands, such as electronics giants and high fashion labels, often keep the key designs in-house because they create high values, and contract out only the low value-adding manufacturing.

Walking also has another significance in the model—it enables travelers to choose between different carriers. Such freedom of choice is ruled out in a hub-and-spoke setup without walking, such as the one in Figure 1, in which each destination can be reached via only one pathway. Williamson's (1975) notion of site specificity highlights the benefits of trading with a nearby supplier or buyer. For some industries, especially natural resources, it is crucial for the buyer to be located close to the supplier. Likewise, the travelers would save walking costs by choosing the carrier that will stop closest to their waiting spot for their next ride. However, if the nearest carrier's ticket is too expensive, they would choose a farther but cheaper carrier if the combined cost of ticket and walking is lower. (In equilibrium, all carriers in the same leg charge the same price, so all travelers would just choose the nearest ride; but the point remains that they are free to choose.) By the same token, a builder may buy from a cheaper materials supplier even if it is further away; a car mechanic may use cheaper third-party parts even if they require modifications; a factory may opt for a cheaper automation system even if it requires more customization.[9]

[8] In fact, most automakers rely on contractors to write their software. When the supply of microchips for vehicles fell short during the pandemic, they could not switch to other microchips easily because that would require rewriting of software. But Tesla Motors write their own software, i.e., they do their own "walking," and therefore was able rewrite them on the microchips that were available. As a result, Tesla weathered the global microchip supply shortage better than other automakers in 2021. See Ewing, Jack. 2022. "Why Tesla Soared as Other Automakers Struggled to Make Cars." *The New York Times*, January 8. https://www.nytimes.com/2022/01/08/business/teslas-computer-chips-supply-chain.html.

[9] As a real-world example, the food company General Mills has revamped the recipes of their products in response to shortages of ingredients following the pandemic. See Creswell, Julie. 2022. "Why Totino's

Multiple materials: Real-world productions involve thousands of different materials being transformed in countless ways. The travel circle reduces any form of transformation of any material into spatial movements in the circle. Such an abstraction is necessary to boil down the complexity of supply chains into anything tractable. But despite the abstraction, we can incorporate multiple materials into the model, at least informally, to give it a more tangible interpretation.

One way is to think of the natural resources at the center as a composite of all raw materials. The composite materials are then split and transformed along the routes into different composite intermediate goods, which continue to transform and ultimately differentiate into the myriad of final goods.

Alternatively, we can incorporate multiple materials more explicitly. Suppose for simplicity that computers are made of only two materials: plastic and silicon. Think of two travel circles, each representing one of the materials, stacked together. The materials first differentiate along their respective routes into plastic and silicon components for different computer models. Then their routes overlap, where the plastic and silicon components are combined into new intermediate goods. The combining and differentiating of intermediate goods continue until they culminate in the finished products.

Investments and labors: Travelers going in the same general direction are pooled together in the same carrier. So a carrier can be seen as a common vehicle that unites its individual passengers for the ride. Likewise, goods that are similar to each other went through the same machines and equipment that process their common components. Therefore the machines and equipment contribute to the generic part of the products. The shipping container is a prime example of a generic and versatile equipment; it can carry any goods that fit in through the global containerized freight system. In manufacturing, generic materials are often matched with versatile equipment. For example, an injection molding machine can shape generic plastic pellets into any form defined by the mold (but the mold itself

Needs 25 Ways to Make Pizza Rolls." *The New York Times*, August 31. https://www.nytimes.com/2022/08/31/business/totinos-pizza-rolls-ingredients.html.

is custom-made for the output). Some machines are more specialized. For instance, auto-plant robots are custom-built for a particular automaker; but even they can be programmed for different car models of the automaker.

The carriers in the model need to be driven by drivers. Likewise, the equipment in real-world production needs to be operated by workers. Human resources, like natural resources, are generic and versatile. The same factory worker can process any category of goods in the assembly line; the same office clerk can handle documents for any type of business. Skilled workers tend to be more specialized; but even they can apply their skills to numerous products. An industrial designer can create a variety of consumer products; an accountant can manage the finances of many different trades.[10]

B. *Insights from the model*

This subsection discusses some key insights drawn from the model. (The insights on welfare have already been covered below Proposition 2.)

Radial differentiation: This paper broadens the concept of product differentiation. The conventional notion of (horizontal) product differentiation corresponds to circular differentiation in the model. The model shows that circular differentiation always goes hand in hand with radial differentiation. On one hand, differentiation as a status refers to variations in attributes of intermediate or final goods (circular differentiation). On the other hand, differentiation as a process refers to generic-to-specific transformations whereby commodities differentiate into intermediate goods, and intermediate goods differentiate into final goods (radial differentiation). In this sense, production *is* differentiation, and products are cumulations of differentiation in multiple steps. For instance, oil is refined into different grades, some of which are processed into various types of plastic pellets, some of which are moulded into certain toys, some of which are adorned with particular features.

[10]The versatility of labor partly explains why labor economics is in general less theory-oriented than industrial organization. Labor is more versatile and hence generic than goods. So the labor market tends to be more competitive than the goods market, which means it requires less complicated market structures.

By interpreting production as differentiation, the notion of radial differentiation also sheds light on the mechanism of scale economies. In the model, a large airplane feeds its passengers to multiple small planes, each of which in turn feeds their passengers to multiple buses, and so on. Likewise, a commodity is used in many intermediate goods, each of which is used in multiple intermediate goods further down the supply chain, and so on. So while differentiation accumulates downstream (outward from the center of the circle), scale economies accumulates upstream (inward). The large airplane attract many passengers because its route is common to the journeys of many travelers. Similarly, commodities and upstream intermediate goods are typically produced in huge scales because they are versatile and are used in numerous downstream intermediate goods. For instance, the economy produces many products with aluminum contents, so aluminum is produced in massive scales. Aluminum product producers thus create external economies of scale for each other by increasing the economies of scale of aluminum production. In this sense, economies of scale upstream (mass production of aluminum) are derived from economies of scope downstream (many products with aluminum contents).

Endogenous investments: The cost equation (1) is crucial to the construction of the model. By linking a firm's marginal cost to its investment cost, it lets each firm choose the optimal investment for their route. It is the universality of this link that enables the whole travel circle in the homecoming model to be generated with just three cost parameters (Theorem 1). As the next paragraph argues, this link is universal not just in the model, but in essentially all productions in the real world. Yet scarcely any existing production models in the literature exploits the link. It could be worthwhile to explore incorporating the link in many existing and future models.

For sure, there are countless production functions and cost functions in the real world. But every production process allows some degree of trade-off between the upfront investment and the average variable cost per unit, as depicted by the cost equation. As noted before, passengers in the travel circle can be carried by anything from a rickshaw to a taxi to a large airplane; the higher the investment, the lower the marginal cost per passenger mile (which equals the average variable cost because the marginal cost is constant). Similarly, most consumption goods like foods and clothes can be produced using a wide range of tools and equipment, from primitive to basic to advanced. Just consider that many goods we consume today, especially the necessities, have been produced in some forms throughout history. The goods produced with simple equipment may be different from their mass-produced counterparts—not necessarily inferior, as handcrafted products are often considered superior nowadays—but the point remains that simple equipment is an option. Even operations that are typically regarded as investment intensive—such as mining—can be carried out with rudimentary tools, albeit torturously. Admittedly, many high-tech processes are impossible without heavy investments in advanced equipment. Yet even these processes afford some trade-off between investment and average variable cost. In microchip productions, for instance, more investments in process optimization would improve production yield.

Preference for varieties: In the vacation model (Section V), the travelers are interpreted collectively as a representative consumer buying many differentiated goods, so they can choose the degree of differentiation of their consumptions by deciding how far to go. Thus the utility function (12) in the vacation model can serve as an alternative to Dixit-Stiglitz utility functions, introduced by the eponyms' 1977 paper, in modeling preference for varieties.

Using the vacation model to account for preference for varieties could bring about certain advantages, especially if the research subject involves supply chains. The travel circle not only models the demand and supply in the market for final goods, it generates the markets for all intermediate goods in the supply chains. For each intermediate and final good, it endogenizes not only the number of firms, their outputs and prices, but also their investments and marginal costs. Moreover, like Salop (1979), the circle provides visualization of the results. Essentially it summarizes the entire economy in a circle. However, unlike Dixit-Stiglitz utility functions, the vacation model does not have a parameter that specifies the elasticity of substitution between the differentiated goods. In addition, since the model generates the economy from the ground up, it involves interdependence between variables that makes the effects of parameters less tractable. In particular, the effect of β on the outcomes is hard to pin down because it plays a key role in all of circular differentiation, radial differentiation, and endogenous radius.

It is commonly accepted that consumers like diversity of goods because higher diversity means more choices, whether of a particular good (e.g. car) or a combination of goods (e.g. weekly dine-outs). But the vacation model offers another perspective on the preference for product diversity. Higher circular differentiation always comes with higher radial differentiation, just as a longer circumference necessarily means a longer radius. Therefore a large variety also implies a high level of refinements. Undifferentiated goods are unrefined goods, with the most extreme case being raw materials. Differentiated goods (e.g. fashions, cuisines) are desirable not only because they offer choices, but also because they are more polished or sophisticated than the generic good (e.g. uniforms, canteen foods).

Integer profits: In the real world, when upstream firms facing few competitors earn positive profits, the profits are often taken as evidence of market power. However, the profits may be at least partly due to the integer constraint. As in Salop (1979), the zero-profit condition results in non-integers being calculated for the equilibrium numbers of firms (Lemma 1). The actual number of firms, which must be an integer, is the calculated number rounded down to the nearest integer because firms will enter only if they expect non-negative profits. To the extent that the number is rounded down, the firms can charge a higher price than that calculated and thus earn an "integer profit," the size of which depends on how significant the rounding down is. On average the amount of rounding down is 0.5 regardless of the leg. So the smaller the calculated number of firms, the more significant the rounding down is on average. Since the number of firms is smaller upstream (Theorem 2), the upstream firms are expected to earn more significant integer profits.

In contrast, the number of downstream firms is large, so they do not earn sizable integer profits.

C. Further observations

This subsection explores several broader subjects that the model sheds light on.

Integration of firms: The model does not allow integration of firms, and it does not produce any concrete predictions on mergers even if integration is allowed. However, the firms' locations provide some clues about which firms are more likely to merge. Integration tends to occur between firms that are close to each other. For both horizontal and vertical integration (i.e., circular and radial integration in the travel circle), the proximity of routes allows the merged firm to share resources (e.g. maintenance supplies) between routes efficiently. In particular, the numerous downstream firms crowded near the circumference of the circle are prone to merge. For horizontal integration, mergers between neighboring firm make even more sense because they are the closest competitors.

The tendency towards vertical integration between firms in the downstream is further exacerbated by Williamson's (1975, 1985) notion of asset specificity. As argued in the previous subsection, it is often the case that the more generic the input, the more generic the equipment that process it. Conversely, the more distinctive the output, the more specific the equipment that produces it. This implies that the suppliers in the downstream, where the outputs are distinctive, are more likely to merge with its buyer when compared to those in the upstream.

The configurations of the travel circle also has another way of hinting at potential vertical integration. Since the locations of the routes in each leg are randomly drawn (subject to equidistance), some routes in consecutive legs may happen to "connect" well, i.e., nearly line up. For instance, in Figure 3, the three consecutive routes running in the southwest direction for legs 1 to 3 nearly line up. The firms running these route are more likely to integrate vertically than others. After merger the same driver could drive multiple routes in a row without walking much from the end of one route to the start of another. Translated into production, this means that if a supplier's outputs are well-suited for the particular requirements of a buyer, then they are relatively prone to merge. Note that the integrated firm can still supply their intermediate goods to other buyers, just as merged carriers in the travel circle would still welcome passengers who take only one of their rides. For example, Samsung and Sony make image sensors for their own smartphones, but they also supply their sensors to competing phone makers; the Chinese appliance manufacturer Midea makes microwave ovens under its own brand, but it also supplies the core components of microwave to many other brands.[11]

Extent of markets: Differentiation also influences the boundary between market and non-market activities. Consider the classic example of household production: a consumer who wants a cake can buy it from a bakery, or they can buy flour, eggs, and sugar from the market and make the cake themselves (Becker 1981). For maximum self-reliance, they could even grow the ingredients themselves. This last approach is equivalent to the traveler walking to their destination without taking any carriers. The choice of how much to rely on markets depends on the peculiarity of their taste and their self-production cost relative to the market cost. If they have very peculiar taste (very isolated destination) and their cost of self-production is relatively low (low walking cost), then they tend to resort to self-production (walk).

While self-production provides an alternative to market purchases, there is a third alternative. The person may ask family members, friends, and neighbors etc. for favors, especially for tasks where personal preferences are important (e.g. baby sitting) and specific knowledge are necessary (e.g. food allergy). (Jeitschko and Lau (2017) call such reciprocal activities "soft transactions," as opposed to explicit "hard transactions" in the market.)

Varieties in the real world: Varieties and customization are costly because the higher the differentiation, the lower the economies of scale (Theorem 2). Distinc-

[11] See 2021. "Sony Dominates Smartphone Image Sensor Market in 2020." *EE Times Asia*, April 1. https://www.eetasia.com/sony-dominates-smartphone-image-sensor-market-in-2020/; and McCabe, Liam and Sullivan, Michael. 2022. "The Best Microwave." *Wirecutter*, January 8. https://www.nytimes.com/wirecutter/reviews/best-microwave.

tive goods are not only expensive to develop and manufacture, they also carry high inventory cost because their turnover is lower than the more standardized items. Jennifer Dulski, a student of economics principles class, tackled the question: "Why do brides spend thousands of dollars on wedding dresses they will never wear again, while grooms, who will have many future opportunities to wear a tuxedo, usually end up renting a cheap one?" Her explanation was essentially that wedding gowns are distinctive, while tuxedos are rather standardized. A gown rental company would have to carry a huge collection of distinctive gowns in each size. But then each gown would get rent out so infrequently such that, after taking the inventory cost into account, it would cost more to rent than buy. Tuxedos, on the other hand, are pretty standardized and therefore can be rented out cheaply (Frank (2006)). Another prediction of the "varieties are expensive" rule is that the more generic garments such as jeans are sold at a lower margin than the more distinctive items such as dresses, which have slower turnovers.

Waist belts and watch straps are adjustable for lengths; socks and hats are one-size-fits-all (at least for adult size). The adaptability of these accessories means that the manufacturers only need to produce a single size for each item. In contrast, clothes and shoes have to come in various sizes, which significantly increase their costs. Consumers who do not fit well in any standard sizes would even need to pay extra for costly individual alterations.

A restaurant can offer a set menu (or daily special) at a lower price than *à la carte* because the set menu is a more standardized product (at least to the restaurant) sold at a higher volume. Buffet restaurants take it one step further by offering a blanket product—a pool of predetermined foods from which the diners self-customize their own meals. Although buffets suffer from inefficiency caused by overeating (once paid up, diners eat until the marginal value of consumption drops to zero), the significant cost savings of pooling make them an economical way of catering, especially to a big crowd like one at the reception of a large conference for economists.

References

Balasubramanian, Sridhar. 1998. "Mail versus Mall: A Strategic Analysis of Competition between Direct Marketers and Conventional Retailers." *Marketing Science*, 17(3): 181–195.

Baumol, W. J., and W. G. Bowen. 1965. "On the Performing Arts: The Anatomy of Their Economic Problems." *The American Economic Review*, 55(1/2): 495–502.

Becker, Gary S. 1981. *A treatise on the family.* Harvard University Press.

Bouckaert, Jan. 2000. "Monopolistic competition with a mail order business." *Economics Letters*, 66(3): 303–310.

Chen, Yongmin, and Michael H. Riordan. 2007. "Price and Variety in the Spokes Model." *The Economic Journal*, 117(522): 897–921.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 67(3): 297–308.

Firgo, Matthias, Dieter Pennerstorfer, and Christoph R. Weiss. 2015. "Centrality and pricing in spatially differentiated markets: The case of gasoline." *International Journal of Industrial Organization*, 40: 81–90.

Frank, Robert H. 2006. "The Economic Naturalist Writing Assignment." *The Journal of Economic Education*, 37(1): 58–67.

Helpman, Elhanan. 1981. "International trade in the presence of product differentiation, economies of scale and monopolistic competition: A Chamberlin-Heckscher-Ohlin approach." *Journal of International Economics*, 11(3): 305–340.

Jeitschko, Thomas D., and C. Oscar Lau. 2017. "Soft transactions." *Journal of Economic Behavior and Organization*, 141: 122–134.

Krugman, Paul. 1980. “Scale Economies, Product Differentiation, and the Pattern of Trade.” *The American Economic Review*, 70(5): 950–959.

Lancaster, Kelvin. 1980. “Intra-industry trade under perfect monopolistic competition.” *Journal of International Economics*, 10(2): 151–175.

Loginova, Oksana. 2009. “Real and Virtual Competition.” *The Journal of Industrial Economics*, 57(2): 319–342.

Madden, Paul, and Mario Pezzino. 2011. “Oligopoly on a Salop Circle with Centre.” *The B.E. Journal of Economic Analysis and Policy*, 11(1).

Salop, Steven C. 1979. “Monopolistic Competition with Outside Goods.” *The Bell Journal of Economics*, 10(1): 141–156.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* The MIT Press.

Williamson, Oliver E. 1975. *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization.* New York: Free Press.

Williamson, Oliver E. 1985. *The economic institutions of capitalism : firms, markets, relational contracting.* New York: Free Press.

FOR ONLINE PUBLICATION

Mathematical Appendix

A1. Proof of Lemma 1

To a traveler, ideally their leg-l ride stops on circle l at exactly their waiting spot for their next ride, or in case of the final leg, at exactly their home. I will first show that in any leg l, the travelers' ideal ending points for their rides are uniformly distributed on circle l. Then I will follow the standard proof for Salop's circle to find the equilibrium number of firms for every leg.

Recall that the game begins with leg L. To a traveler, ideally their final ride stops on circle L at exactly their home. Since the travelers' homes are uniformly distributed on the circle, so are their ideal points.

The leg-L firms, randomly located subject to equidistance between firms, set the same price and make equal sales; so they are allocated equal waiting zones on circle $(L-1)$. Each passenger of a firm's ride is then randomly assigned a waiting spot on the firm's waiting zone such that the zone is filled up with waiting spots uniformly. Therefore the travelers' waiting spots are uniformly distributed on circle $(L-1)$, regardless of the exact locations of the firms. Repeat the same argument for leg $(L-1)$, leg$(L-2)$,..., and finally leg 1, it becomes clear that in any leg l, the travelers' ideal ending points for their rides are uniformly distributed on circle l.

I now follow the standard proof for Salop's model. Suppose n_l firms have entered leg l. One of them is firm i; it has invested f_{li} and is now setting its ticket price p_{li}. Consider a traveler whose ideal point on circle l is $x \in \left(0, \frac{R_l}{n_l}\right)$ miles away from firm i's ending point (see Figure A1).

If they take firm i's ride, after getting off they will have to walk for x miles to reach their ideal point. If they instead take the ride run by firm i's neighbor that is closest to their ideal point (see firm $(i-1)$ in the figure), they will walk for $\frac{R_l}{n_l} - x$ miles. Suppose the neighboring firm charges a price of p_l. The traveler is indifferent between the two rides if the combined cost of ticket and walking is the same in both

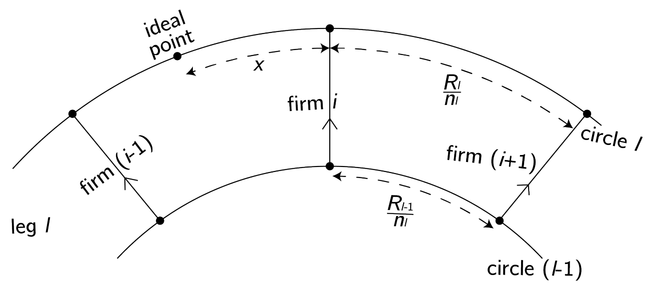


Figure A1. : Leg l

options, i.e., $p_{li} + wx = p_l + w\left(\frac{R_l}{n_l} - x\right)$. So any traveler whose ideal point is less than $x = \frac{R_l}{2n_l} + \frac{p_l - p_{li}}{2w}$ miles away from firm i's ending point will take firm i's ride.

Firm i competes with its closest neighbors on both sides. So its demand will be $\frac{2x}{R_l} = \frac{1}{n_l} + \frac{p_l - p_{li}}{wR_l}$. With an investment of f_{li}, its constant marginal cost is $\frac{\alpha}{f_{li}^{\beta}}$. Firm i chooses the optimal price to maximize its profit:

$$\max_{p_{li}} \left[\left(p_{li} - \frac{\alpha}{f_{li}^{\beta}} \cdot r_l\right)\left(\frac{1}{n_l} + \frac{p_l - p_{li}}{wR_l}\right) - f_{li}\right]$$

Solving the first-order condition with respect to p_{li}, and then setting $p_{li} = p_l$ and $f_{li} = f_l$ for symmetry, we obtain $p_l = \frac{\alpha}{f_l^{\beta}} \cdot r_l + \frac{wR_l}{n_l}$. Each entrant in leg l therefore will earn a profit of $\pi_l = \frac{wR_l}{n_l^2} - f_l$. But with free entry, the many potential entrants, all possessing the same technology, should drive the equilibrium profit to zero (up to the integer constraint). Setting the profit to zero, we obtain $n_l = \sqrt{\frac{wR_l}{f_l}}$. ■

A2. Proof of Lemma 2

The first order condition of the cost equation (1) with respect to f is $1 = \frac{\alpha\beta xm}{f^{\beta+1}}$. Solving for f leads to $f^{min}(x,m)$. The cost function follows readily. Also,

$$
\begin{aligned}
\frac{\partial f^{min}(x,m)}{\partial \beta} &= (\alpha\beta xm)^{\frac{1}{\beta+1}} \frac{\partial}{\partial \beta}\left[\frac{1}{\beta+1}\ln(\alpha\beta xm)\right] \\
&= \frac{(\alpha\beta xm)^{\frac{1}{\beta+1}}}{\beta+1}\left[\frac{1}{\beta} - \frac{\ln(\alpha\beta xm)}{\beta+1}\right] \\
&> 0 \qquad \text{iff} \qquad \ln(\alpha\beta xm) < \frac{1}{\beta} + 1.
\end{aligned}
$$

So f^{min} is increasing in β iff $xm < \frac{\exp\left(\frac{1}{\beta}+1\right)}{\alpha\beta}$. ∎

A3. Proof of Proposition 1

In equilibrium, each entrant in leg l has a sale of $1/n_l$. So the cost-minimization condition in Lemma 2 becomes $f_l = \left(\frac{\alpha\beta r_l}{n_l}\right)^{\frac{1}{\beta+1}}$. Substituting this into the zero-profit condition in Lemma 1, we get $n_l^2 = wR_l\left(\frac{n_l}{\alpha\beta r_l}\right)^{\frac{1}{\beta+1}} \Rightarrow n_l^{2\beta+1} = \frac{w^{\beta+1}R_l^{\beta+1}}{\alpha\beta r_l}$. Equation (2) follows. Then substitute equation (2) into the cost-minimization condition above to obtain (3). Equation (4) follows immediately from (2) and (3).

The firms earn zero profits, meaning that the ticket price (for one unit of passenger, which is the whole population) should cover the total investment and operating costs exactly. Recall the two components in the cost equation (1). The total investment costs for leg l are $n_l^* f_l^*$, and the total operating costs are $\frac{\alpha r_l}{(f_l^*)^\beta} = \alpha r_l\left(\frac{wR_l}{\alpha^2\beta^2 r_l^2}\right)^{\frac{\beta}{2\beta+1}} = \frac{1}{\beta}\left(\alpha\beta r_l w^\beta R_l^\beta\right)^{\frac{1}{2\beta+1}} = \frac{n_l^* f_l^*}{\beta}$. Add the two components up to obtain equation (5). The markup ratio is $\left(1+\frac{1}{\beta}\right)/\left(\frac{1}{\beta}\right) - 1 = \beta$.

Equation (6) is obtained by adding up the investment costs, operating costs, and walking costs. The investment costs and operating costs are given by equation (5). For the walking costs, recall that each traveler chooses the leg-l ride whose ending point on circle l is closest to their waiting spot for their leg-$(l+1)$ ride. Since the leg-l rides' ending points are R_l/n_l^* apart from each other circularly, upon arrival a traveler will walk at most $R_l/2n_l^*$ miles and at least zero mile away to their waiting spot. So given that the travelers' waiting spots are uniformly distributed on circle l, the average walking distance is $R_l/4n_l^*$ miles, and hence the total walking costs for leg l are $\frac{wR_l}{4n_l^*} = \frac{1}{4}\left(\alpha\beta r_l w^\beta R_l^\beta\right)^{\frac{1}{2\beta+1}} = \frac{n^* f^*}{4}$.

Remark: I now explain the reason for introducing the crashing condition in the model setup. The crashing condition forces the travelers to walk to their waiting spots, but allows them to slide from their waiting spots to their carrier for free. Since the number of leg-$(l+1)$ carriers is n_{l+1}^*, on average a traveler's waiting spot on circle l is $R_l/4n_{l+1}^*$ miles away from their leg-$(l+1)$ carrier. So had the travelers have to walk instead of slide to their carrier, they will on aggregate incur an extra walking cost of $\frac{wR_l}{4n_{l+1}^*} = \frac{1}{4}\left(\alpha\beta r_{l+1} w^\beta \frac{R_l^{2\beta+1}}{R_{l+1}^{\beta+1}}\right)^{\frac{1}{2\beta+1}}$ for leg l. This extra term is complex due to the mismatch of leg subscripts between R_l and n_{l+1}^*. Adding this term to the market total cost would bring complexity to the model without adding much value. The walking costs are meant to capture the adaptation and customization costs in between mass productions. But this is already achieved by requiring the travelers to walk to their waiting spots, considering that the size of the costs can be adjusted through w. ■

A4. Proof of Proposition 2

As explained at the end of the proof of Proposition 1, the total walking costs of all travelers in leg l is $\frac{wR_l}{4n_l}$ if there are n_l firms. Moreover, each carrier will carry $\frac{1}{n_l}$ passengers, so the operating cost of each firm is $\frac{\alpha}{f_l^\beta}\frac{r_l}{n_l}$. Therefore, to minimize the sum of firms' costs and walking costs, the social planner solves the problem:

$$\min_{n_l, f_l}\left[n_l\left(f_l + \frac{\alpha}{f_l^\beta}\frac{r_l}{n_l}\right) + \frac{wR_l}{4n_l}\right].$$

The first-order condition with respect to n_l is $n_l = \frac{1}{2}\sqrt{\frac{wR_l}{f_l}}$, which differs from the zero-profit condition in the Salop circle (Lemma 1) by the constant $1/2$. The first-order condition with respect to f_l is $f_l = \left(\frac{\alpha\beta r_l}{n_l}\right)^{\frac{1}{\beta+1}}$, which is the same as firms' cost minimization condition implied by Lemma 2. Combine the two first-order conditions to solve for n_l^{**}, f_l^{**} and $n_l^{**} f_l^{**}$ in equations (7), (8) and (9). Then substitute n_l^{**} and f_l^{**} into the social planner's cost equation to obtain C_l^{**} in equation (10). Setting these results against those in Proposition 1 yield the comparisons of the efficient outcomes with their market counterparts.

To see that $\widetilde{\beta}$ is strictly decreasing in β, consider that $\widetilde{\beta} = 2^{\frac{1}{2\beta+1}-1} \cdot \frac{8\beta+4}{5\beta+4}$, so $\frac{d \ln \widetilde{\beta}}{d\beta} = -\frac{2\ln 2}{(2\beta+1)^2} + \frac{2}{2\beta+1} - \frac{5}{5\beta+4} = -\frac{(10\ln 2-6)\beta+8\ln 2-3}{(2\beta+1)^2(5\beta+4)} \approx -\frac{0.931\beta+2.545}{(2\beta+1)^2(5\beta+4)} < 0$. Therefore, $\ln\widetilde{\beta}$ is decreasing in β. But log is an increasing function, so $\widetilde{\beta}$ is also decreasing in β. ■

A5. Proof of Proposition 3

Equations (4) and (6) imply that that $C_l^* = \left(\frac{5}{4} + \frac{1}{\beta}\right)\left(\alpha\beta w^\beta\right)^{\frac{1}{2\beta+1}}\left(r_l R_l^\beta\right)^{\frac{1}{2\beta+1}}$. So C_l^* depends on the RDP only through the expression $\left(r_l R_l^\beta\right)^{\frac{1}{2\beta+1}}$. Similarly, by equations (9) and (10), C_l^{**} has the same property. Therefore, $\sum_{l=1}^L C_l^*$ and $\sum_{l=1}^L C_l^{**}$ depend on the RDP only through the sum $\sum_{l=1}^L \left[\left(r_l R_l^\beta\right)^{\frac{1}{2\beta+1}}\right]$. An RDP that minimizes this sum also minimizes $\sum_{l=1}^L C_l^*$ and $\sum_{l=1}^L C_l^{**}$.

Therefore, the RDP under either allocation is obtained by solving the problem:

$$\text{(A1)} \qquad \min_{(r_1,r_2,\ldots,r_L)} \sum_{l=1}^{L}\left\{\left[r_l\left(\sum_{k=1}^{l} r_k\right)^{\beta}\right]^{\frac{1}{2\beta+1}}\right\} \qquad \text{subject to} \qquad \sum_{l=1}^{L} r_l = 1$$

where L can be finite or ∞. (The constraints that r_l is positive for all l are not included in the minimization problem; instead any solution with negative r_l will be rejected.) Solve the problem by the Lagrangian method. The Lagrangian function is

$$\mathcal{L}\left(r_1, r_2, ..., r_L, \lambda\right) = \sum_{l=1}^{L}\left\{\left[r_l\left(\sum_{k=1}^{l} r_k\right)^{\beta}\right]^{\frac{1}{2\beta+1}}\right\} + \lambda\left(1 - \sum_{l=1}^{L} r_l\right)$$

where λ is the Lagrange multiplier. The first-order condition with respect to r_l is

$$
\begin{aligned}
& (2\beta+1)\,\lambda \\
= \quad & \left[r_l\,(r_1+...+r_l)^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} \left[(r_1+...+r_l)^{\beta}+\beta r_l\,(r_1+...+r_l)^{\beta-1}\right] \\
& + \left[r_{l+1}\,(r_1+...+r_{l+1})^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} \beta r_{l+1}\,(r_1+...+r_{l+1})^{\beta-1} \\
& + ... \\
& + \left[r_L\,(r_1+...+r_L)^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} \beta r_L\,(r_1+...+r_L)^{\beta-1}
\end{aligned}
$$

and the first-order condition with respect to r_{l+1} is

$$
\begin{aligned}
& (2\beta+1)\,\lambda \\
= \quad & \left[r_{l+1}\,(r_1+...+r_{l+1})^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} \left[(r_1+...+r_{l+1})^{\beta}+\beta r_{l+1}\,(r_1+...+r_{l+1})^{\beta-1}\right] \\
& + \left[r_{l+2}\,(r_1+...+r_{l+2})^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} \beta r_{l+2}\,(r_1+...+r_{l+2})^{\beta-1} \\
& + ... \\
& + \left[r_L\,(r_1+...+r_L)^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} \beta r_L\,(r_1+...+r_L)^{\beta-1}.
\end{aligned}
$$

Taking the difference between the two conditions leads to

$$
\begin{aligned}
\text{(A2)} \qquad & \left[r_l\,(r_1+...+r_l)^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} (r_1+...+r_l)^{\beta}\left(1+\beta\frac{r_l}{r_1+...+r_l}\right) \\
& = \left[r_{l+1}\,(r_1+...+r_{l+1})^{\beta}\right]^{-\frac{2\beta}{2\beta+1}} (r_1+...+r_{l+1})^{\beta}.
\end{aligned}
$$

Raising both sides to the power of $(\frac{2\beta+1}{\beta})$, it follows that

$$
\text{(A3)} \qquad r_{l+1}^2\,(r_1+\ldots+r_l)\left(1+\beta\frac{r_l}{r_1+\ldots+r_l}\right)^{2+\frac{1}{\beta}} = r_l^2\,(r_1+\ldots+r_{l+1}).
$$

Dividing both sides by $r_{l+1}^2 (r_1 + \ldots + r_l)$ and applying the definition of $\widehat{r_l}$ yields

$$(1 + \beta\widehat{r_l})^{2+\frac{1}{\beta}} = \left(\frac{r_l}{r_{l+1}}\right)^2 \left[1 + \widehat{r_l}\left(\frac{r_{l+1}}{r_l}\right)\right],$$

which can be rearranged into a quadratic equation in $\frac{r_l}{r_{l+1}}$:

$$\left(\frac{r_l}{r_{l+1}}\right)^2 + \widehat{r_l}\left(\frac{r_l}{r_{l+1}}\right) - (1 + \beta\widehat{r_l})^{2+\frac{1}{\beta}} = 0.$$

Solving for the positive root of the quadratic equation, we obtain the algorithm (11).

Next, I argue that the cost-minimizing RDP must be infinite. Consider two RDPs: RDP^1 and RDP^2, which have L legs and $L+1$ legs respectively. They are both constructed by following the algorithm (11). Let $RDP^1 = (r_1^L, r_2^L, .., r_L^L)$ and $RDP^2 = (r_1^{L+1}, r_2^{L+1}, ..., r_L^{L+1}, r_{L+1}^{L+1})$. Consider another RDP with $L+1$ legs: $RDP^3 = (r_1^L, r_2^L, ..., r_L^L - \epsilon, \epsilon)$ where $0 < \epsilon < r_L^L$. When ϵ tends to zero, the total cost of the travel circle under RDP^3 and RDP^1 tend to be the same, either under market or efficient allocation. But RDP^2 costs less than RDP^3 because while they both have $L+1$ legs, only the former follows equation (11). So RDP^2 costs less than RDP^1. Therefore given any two RDPs that both follow (11), the one with more legs always costs less. It follows that the market and efficient RDPs must be infinite.

Lastly, the RDPs are unique because (11) gives unique solutions for $^{r_l}/_{r_{l+1}}$. ■

A6. Proof of Theorem 2

Recall that Proposition 3 implies the sequences $(r_1^*, r_2^*, ...)$ and $(r_1^{**}, r_2^{**}, ...)$ are identical, so any results proved for r_l^* also apply to r_l^{**}.

(i) Let $\widehat{r_l}^* := \frac{r_l^*}{\sum_{k=1}^{l} r_k^*}$. Define a function $g(\widehat{r_l}; \beta) := (1 + \beta\widehat{r_l})^{1+\frac{1}{\beta}} - (1+\beta)\widehat{r_l} - 1$. Note that $g(0; \beta) = 0$ and $\frac{\partial g}{\partial \widehat{r_l}} = (1+\beta)\left[(1+\beta\widehat{r_l})^{\frac{1}{\beta}} - 1\right] > 0$ for all β and $\widehat{r_l}$ as both are always positive. So $g(\widehat{r_l}^*; \beta) > 0$ and hence $(1 + \beta\widehat{r_l}^*)^{1+\frac{1}{\beta}} > 1 + (1+\beta)\widehat{r_l}^*$ for all β (note that $\widehat{r_l}^*$ is a function of β). Multiplying both sides of the inequality by $(1+\beta\widehat{r_l}^*)$ yields

$$(1+\beta\widehat{r_l}^*)^{2+\frac{1}{\beta}} > 1 + (1+2\beta)\,\widehat{r_l}^* + \beta\,(1+\beta)\,(\widehat{r_l}^*)^2.$$

Adding $\left(\frac{\widehat{r_l}^*}{2}\right)^2$ to both sides and then factoring the right-hand-side gives

$$\left(\frac{\widehat{r_l}^*}{2}\right)^2 + (1+\beta\widehat{r_l}^*)^{2+\frac{1}{\beta}} > \left[1 + \left(\beta + \frac{1}{2}\right)\widehat{r_l}^*\right]^2.$$

Taking square root of both sides and then subtracting $\frac{\widehat{r_l}^*}{2}$ from both sides yields

$$-\frac{\widehat{r_l}^*}{2} + \sqrt{\left(\frac{\widehat{r_l}^*}{2}\right)^2 + (1+\beta\widehat{r_l}^*)^{2+\frac{1}{\beta}}} > 1 + \beta\widehat{r_l}^*.$$

Recognizing from iteration (11) that the left-hand-side equals $\frac{r_l^*}{r_{l+1}^*}$, we have

$$\text{(A4)} \qquad \frac{r_l^*}{r_{l+1}^*} > 1 + \beta\widehat{r_l}^*.$$

As $\beta\widehat{r_l}^*$ is positive, it follows that r_l^* is strictly decreasing in l, and so is r_l^{**}. Furthermore, given that the market and efficient RDPs have infinite number of legs in a finite-sized circle, and that the length of leg is decreasing towards the circumference of the travel circle, it must be true that the lengths of legs in both RDPs approach zero towards the circumference.

(ii) Consider a function $h(x;\beta) := -\frac{x}{2} + \sqrt{\left(\frac{x}{2}\right)^2 + (1+\beta x)^{2+\frac{1}{\beta}}}$, so defined to mimic iteration (11). Taking partial derivative with respect to x,

$$\begin{aligned}\frac{\partial h}{\partial x} &= -\frac{1}{2} + \frac{1}{2}\cdot\frac{\frac{x}{2} + (1+2\beta)\,(1+\beta x)^{1+\frac{1}{\beta}}}{\sqrt{\left(\frac{x}{2}\right)^2 + (1+\beta x)^{2+\frac{1}{\beta}}}} \\ &> -\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\left(\frac{x}{2}\right)^2 + (1+2\beta)^2\,(1+\beta x)^{2+\frac{2}{\beta}}}{\left(\frac{x}{2}\right)^2 + (1+\beta x)^{2+\frac{1}{\beta}}}} \\ &> -\frac{1}{2} + \frac{1}{2}\cdot 1 = 0.\end{aligned}$$

So h is strictly increasing in x. Apply this result to iteration (11), it follows that if $\widehat{r_l}^*$ is strictly decreasing in l, then $\frac{r_l^*}{r_{l+1}^*}$ must be strictly decreasing in l as well. Meanwhile, $\widehat{r_l}^*$ is indeed strictly decreasing in l because $\widehat{r_{l+1}}^* = \frac{r_{l+1}^*}{\sum_{k=1}^{l+1} r_k^*} < \frac{r_l^*}{\sum_{k=1}^{l+1} r_k^*} < \frac{r_l^*}{\sum_{k=1}^{l} r_k^*} = \widehat{r_l}^*$, with the first inequality resulting from part (i) above. Therefore $\frac{r_l^*}{r_{l+1}^*}$ and hence $\frac{r_l^{**}}{r_{l+1}^{**}}$ are strictly decreasing in l.

Moreover, according to part (i), $\widehat{r_l}^*$ approaches 0 as l approaches ∞; but according to algorithm (11), when $\widehat{r_l}^*$ approaches 0, $\frac{r_l^*}{r_{l+1}^*}$ approaches 1. Therefore, $\frac{r_l^*}{r_{l+1}^*}$ and hence $\frac{r_l^{**}}{r_{l+1}^{**}}$ approach 1 when l approaches ∞.

(iii) The results follow immediately from part (i), equations (2) and (7), noting that R_l is increasing in l and upper-bounded by 2π.

(iv) Similarly, the results follow immediately from part (i), equations (3) and (8).

(v) Rearranging equation (A2) and applying the definition of $\widehat{r_l}^*$ yields

$$\left[\frac{r_l^*\left(r_1^* + ... + r_l^*\right)^\beta}{r_{l+1}^*\left(r_1^* + ... + r_{l+1}^*\right)^\beta}\right]^{\frac{2\beta}{2\beta+1}} = \frac{1+\beta\widehat{r_l}^*}{\left(1+\frac{r_{l+1}^*}{r_1^*+...+r_l^*}\right)^\beta}.$$

Rearranging equation (A3) and again applying the definition of $\widehat{r_l}^*$ yields

$$1+\frac{r_{l+1}^*}{r_1^*+...+r_l^*} = \left(\frac{r_{l+1}^*}{r_l^*}\right)^2 \left(1+\beta\widehat{r_l}^*\right)^{2+\frac{1}{\beta}}.$$

Substituting the second equation into the first, and raising both side of the resulting equation to the power of $1/2\beta$, it follows that

$$\text{(A5)} \qquad \left[\frac{r_l^*\left(r_1^* + ... + r_l^*\right)^\beta}{r_{l+1}^*\left(r_1^* + ... + r_{l+1}^*\right)^\beta}\right]^{\frac{1}{2\beta+1}} = \frac{r_l^*}{r_{l+1}^*}\cdot\frac{1}{1+\beta\widehat{r_l}^*},$$

which is greater than one according to inequality (A4). But by equation (4), $\frac{n_l^* f_l^*}{n_{l+1}^* f_{l+1}^*}$ equals the left-hand-side of equation (A5), so $\frac{n_l^* f_l^*}{n_{l+1}^* f_{l+1}^*} > 1$, i.e., $n_l^* f_l^*$ is strictly decreasing in l. Similarly, by equations (6), (9) and (10), C_l^*, $n_l^{**} f_l^{**}$, and C_l^{**} are also all strictly decreasing in l.

As l increases, r_l^* tends to zero, and so does $n_l^* f_l^*$ according to equation (4). Similarly, C_l^*, $n_l^{**} f_l^{**}$, and C_l^{**} also approach zero by (6), (9) and (10).

Note that although the total costs of the travel circle with the market or efficient RDP are infinite sums, there is no concern of divergence for the sums. Since $C_{l+1}^*/C_l^* < 1$ and $C_{l+1}^{**}/C_l^{**} < 1$ for all l, the infinite sums converge, courtesy of the ratio test for convergence.

(vi) The result follows immediately from part (v) and equation (5).

(vii) Equation (A5) can be rearranged into

$$\frac{r_{l+1}^*}{r_l^*} \left[\frac{r_l^* \left(r_1^* + \ldots + r_l^*\right)^\beta}{r_{l+1}^* \left(r_1^* + \ldots + r_{l+1}^*\right)^\beta} \right]^{\frac{1}{2\beta+1}} = \frac{1}{1 + \beta \widehat{r_l}^*},$$

which is obviously smaller than one. But the left-hand-side equals $\frac{p_l^*}{r_l^*} / \frac{p_{l+1}^*}{r_{l+1}^*}$ by equations (4) and (5). So $\frac{p_l^*}{r_l^*} < \frac{p_{l+1}^*}{r_{l+1}^*}$, i.e., $\frac{p_l^*}{r_l^*}$ is strictly increasing in l. Similarly, applying the inequality to equations (4) and (6), and (9) and (10) respectively yields the results that $\frac{C_l^*}{r_l^*}$ and $\frac{C_l^{**}}{r_l^*}$ are strictly increasing in l.

By equation (4), $\frac{n_l^* f_l^*}{r_l^*} = \frac{\left(\alpha \beta w^\beta R_l^\beta\right)^{\frac{1}{2\beta+1}}}{\left(r_l^*\right)^{1-\frac{1}{2\beta+1}}}$. As l increases, r_l^* tends to zero, so $\frac{n_l^* f_l^*}{r_l^*}$ tends to infinity. Similarly, $\frac{n_l^* f_l^{**}}{r_l^{**}}$ also tends to infinity. So by equations (5), (6), and (10), $\frac{p_l^*}{r_l^*}$, $\frac{C_l^*}{r_l^*}$, and $\frac{C_l^{**}}{r_l^{**}}$ all approach infinity. ■

A7. Proof of Lemma 3

Consider a travel circle of radius r miles instead of one mile. To find the market RDP, follow the same steps as in the proof of Proposition 3, except that the lengths of legs sum to r instead of 1 in the constrained minimization problem (A1). The change in the constraint constant from 1 to r has no effect on any of the steps leading up to equation (A3). (The proof of of Proposition 3 for the market RDP relies on equations (4) and (6) from Proposition 1, but Proposition 1 does not require restricting the radius to one mile.)

Therefore the algorithm (11) in Proposition 3, which follows from equation (A3), still holds when the radius is r. In other words, the ratios of the lengths of legs follow the same algorithm regardless of the radius. So if the radius is r instead of 1, then the length of leg l in the market RDP is simply r times r_l^*, which depends on β only. So by equations (4) and (6), with r_l replaced by $r \cdot r_l^*$,

$$\begin{aligned}\mathcal{C}_r^*(\alpha,\beta,w) &= \sum_{l=1}^{\infty}\left[\left(\frac{1}{\beta}+\frac{5}{4}\right)\left(\alpha\beta r \cdot r_l^*(\beta)\cdot w^{\beta}\,(2\pi)^{\beta}\, r^{\beta}\left(\sum_{k=1}^{l} r_l^*(\beta)\right)^{\beta}\right)^{\frac{1}{2\beta+1}}\right] \\ &= r^{\widehat{\beta}}\cdot \mathcal{C}_1^*(\alpha,\beta,w).\end{aligned}$$

■

A8. Proof of Proposition 4 and Corollary 1

By Lemma 3, the cost of traveling r miles is $r^{\widehat{\beta}}\mathcal{C}_1^*$. Therefore, the representative traveler maximizes $u(r,s)=r^{\theta}s^{1-\theta}$ subject to $I=r^{\widehat{\beta}}\mathcal{C}_1^*+p_s s$. Rewrite the objective function as $u(r,s(r))=r^{\theta}\left(\frac{I-r^{\widehat{\beta}}\mathcal{C}_1^*}{p_s}\right)^{1-\theta}$ without constraints. Taking derivative with respect to r leads to $\frac{\theta}{1-\theta}\cdot\left(I-r^{\widehat{\beta}}\mathcal{C}_1^*\right)=\widehat{\beta}r^{\widehat{\beta}}\mathcal{C}_1^*$. Solving for r yields the result for r^*. Corollary 1 is obvious. ■

A9. Proof of Lemma 4

As explained in the proof of Lemma 3, if the radius of the travel circle is r, then the length of leg l in the market RDP for the vacation model is simply rr_l^*. Since the radius is r^* under market allocation, the market RDP is $(r^*r_1^*, r^*r_2^*, \ldots)$. ■