Massachusetts Institute of Technology
Department of Economics
Working Paper Series

Knife-Edge or Plateau: When Do Market Models Tip?

Glenn Ellison
Drew Fudenberg

Working Paper 03-07
February 2003

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

KNIFE-EDGE OR PLATEAU: WHEN DO MARKET MODELS TIP?*

GLENN ELLISON
MIT

DREW FUDENBERG
Harvard University

First draft: July 10, 2002
This draft: February 5, 2003

This paper studies whether agents must agglomerate at a single location in a class of models of two-sided interaction. In these models there is an increasing returns effect that favors agglomeration, but also a crowding or market-impact effect that makes agents prefer to be in a market with fewer agents of their own type. We show that such models do not tip in the way the term is commonly used. Instead, they have a broad plateau of equilibria with two active markets, and tipping occurs only when one market is below a critical size threshold. Our assumptions are fairly weak, and are satisfied in Krugman's [1991b] model of labor market pooling, a heterogeneous-agent version of Pagano's [1989] asset market model, and Ellison, Fudenberg and Möbius's [2002] model of competing auctions.

* The authors thank the NSF for financial support, Bob Anderson, Abhijit Banerjee, Ed Glaeser, and Dimitri Vayanos for helpful comments, and Dan Hojman for careful proofreading.

1. Introduction

Many economic activities are agglomerated: people are crowded into a small fraction of the Earth's land mass; individual industries are geographically concentrated; trading is concentrated in a few marketplaces. The standard way to account for concentration (and the arbitrariness of where activity concentrates) has been to propose "tipping models" with three equilibria: one with most activity at location A; one with most activity at B; and an unstable "knife-edge" equilibrium with exactly half of the activity in each market. At the core of most models of agglomeration is some type of increasing returns or "scale effect" that favors the emergence of a single dominant site. In some of these models, all agents are ex-ante identical, and they all prefer to be part of the larger market. In such cases, it is clear that an equilibrium where all markets are exactly the same size is an unstable knife-edge; every agent would rather be in a market with 51% of the agents than a market with 49%, so any departure from exactly equal sizes leads to "tipping" to a single site. In other models, there are differences between the agents, and while all agents prefer larger markets, they also prefer markets where the other agents are less like themselves. For example, firms like markets with an excess of workers, upstream firms like markets with many downstream firms, sellers of financial assets like markets with many buyers, and men prefer a dating site that has many women. In this two-sided case, there is a potential "market impact" or "competitive" effect that may discourage agents from switching markets. We find that this can turn the "knife-edge" of exactly equal shares into a "plateau" of many stable equilibria with unequal market sizes, thus generalizing an observation that we made in the context of a model of competing auctions in Ellison, Fudenberg, and Möbius [2002].

For example, in Krugman's [1991b] labor pooling model, a firm that switches into a market raises the average wage and so lowers the utility of all firms. Because of this market impact effect, the equal-sizes configuration is not only an equilibrium, but a strict equilibrium: If a buyer or seller were to switch to the other market he or she would find that there were now more participants on his or her side of the market and no more on the other, which would make it strictly less attractive. It is true that in Krugman's model, as in many others, the market impact effect vanishes as the market becomes large. However,

the scale effect that favors large markets also vanishes as the market becomes large, so it is misleading to retain one of these effects and ignore the other unless one knows more about the rates at which the two effects vanish.

To investigate the importance of these effects, we study a model with two kinds of agents, who we will call "buyers" and "sellers." At the start of the period, buyers and sellers simultaneously choose between two possible locations or markets; their payoffs are determined by the numbers of each type of agent who chose the same location. Our assumptions are consistent with models where trade is voluntary and a market with only one agent provides no opportunity for trade; and indeed this is a property of the examples we study; in such cases there are always equilibria in which all agents concentrate in either location. When the numbers of buyers and sellers are even, there is also an equilibrium where the two markets are exactly the same size.

Our main point is that the equal-sizes configuration need not be a knife-edge. We provide sufficient conditions for there to be a wide range (a "plateau") of size ratios for the two markets at which all of the incentive constraints for equilibrium are satisfied. Roughly speaking, these conditions are that as the number of agents increases, the payoff functions converge to well-defined limits that are continuous and differentiable, that the derivatives of these limit payoff functions with respect to the ratio of various types of agents be non-zero, and that the convergence to the limit occurs at rate at least $1/N$, where N is the total number of participants.

Throughout the paper, we simplify by ignoring the restriction that the numbers of each type of agent in each market should be integers. Anderson, Ellison, and Fudenberg [2003] studies the additional complications caused by the restriction to integer values. They show that for "typical" economies the conditions of this paper suffice for a plateau of equilibria, but that there are examples where the integer restriction is inconsistent with any split-market equilibrium at all.

Section II of the paper states and discusses our general conditions, and gives our theorem on the lower bound on the width of the plateau. We then show that the conditions are satisfied in a series of examples. Specifically, Section III analyzes the Krugman labor-pooling model mentioned above, Section IV analyzes a two-population

version of Pagano [1989]'s model of competing financial markets. Section V gives examples of how the assumptions can fail.

We are agnostic as to whether the models we use as examples account for actual agglomeration, but we do believe that models with an equilibrium plateau are needed to account for some of the stylized facts. Consider, for example, patterns of industry agglomeration. Even in the most concentrated industries one rarely finds that most activity has tipped to one rather than to several locations. The upholstered furniture industry, for example, is famous for its concentration in North Carolina, but only 74 of the industry's 219 large plants are located there. Another 52 are in Mississippi and 27 are in California.[1] It does not seem reasonable to claim the 52 furniture plants in Mississippi are consistent with a tipping model by arguing that they are serving local demand; nor can one reasonably argue that tipping toward North Carolina has reached an upper bound due to congestion, etc. – the upholstered furniture industry employs less than 1% of North Carolina's workforce and the density of large furniture plants is about one per 658 square miles of land.[2] While "tipping" does not seem to be occurring between industry centers, there do appear to be threshold effects at the bottom end: Nineteen states have exactly zero large plants making upholstered furniture.[3] A model with an equilibrium plateau could account for the coexistence of multiple centers and for many locations having tipped to having almost no activity.

The reader should note that, under our assumptions, per-capita utility is about the same in the "split markets" equilibria as when the economy has tipped to a single market. Nevertheless, the aggregate welfare loss need not be negligible, and it is this aggregate which influences the rents that might be earned by consolidating the markets. Moreover, our focus is not on welfare *per se*, but rather on understanding the positive question of whether we should expect agglomeration economies to lead to tipping.

---

[1] By "large plant" we mean a plant with at least 100 employees. The plant counts are taken from County Business Patterns for 2000. The upholstered furniture industry is approximately at the 90th percentile in Ellison and Glaeser's [1997] tabulation of industry agglomeration.

[2] It would also be difficult to argue that the secondary industry centers are a disequilibrium feature of a market that is in the process of tipping. Dumais, Ellison, and Glaeser [2002] find that in the typical agglomerated industry, there is mean reversion in the state-industry employment shares.

[3] Similar threshold effects appear in many less concentrated industries. For example, the pharmaceutical manufacturing industry has about the mean level of agglomeration in Ellison and Glaeser's [1997]

2. The Model and Result

We examine a simple two-stage model of location choice. In the first stage S sellers and B buyers simultaneously choose whether to attend market 1 or market 2. In the second stage, they play some game with the other players who have chosen to attend the same market, e.g. they may trade at prices set by competing market makers or play some wage-setting game. Rather than specifying the market game, we simply assume that if S_i sellers and B_i buyers attend market i, then the market game gives the sellers in market i an expected payoff of $u_s(S_i, B_i)$ and the buyers an expected payoff of $u_b(S_i, B_i)$. We treat the utility functions $u_s(S,B)$ and $u_b(S,B)$ as the primitives of our analysis.[4]

A pure strategy Nash equilibrium of this model is a profile of location choices such that each of the buyers and sellers receives at least as high a payoff in the market they attend as they would if they had deviated and instead gone to the other market. There will be two types of pure-strategy equilibria: equilibria where the market has tipped and all buyers and sellers attend a single market, and equilibria in which both markets are active. The goal of this section is to characterize the range of market sizes that are possible in equilibria with two active markets. Our first result is immediate.

PROPOSITION 1: Let S_1, S_2, B_1, and B_2 be positive integers with $S_1 + S_2 = S$ and $B_1 + B_2 = B$. There is a pure strategy Nash equilibrium with S_i sellers and B_i buyers choosing market i if and only if the following four constraints hold.

(B1) $u_b(S_1, B_1) \geq u_b(S_2, B_2 + 1)$

(B2) $u_b(S_2, B_2) \geq u_b(S_1, B_1 + 1)$

(S1) $u_s(S_1, B_1) \geq u_s(S_2 + 1, B_2)$

(S2) $u_s(S_2, B_2) \geq u_s(S_1 + 1, B_1)$

---

tabulation with 37 of the 231 large plants in New Jersey, 31 in New York, 21 in California, 17 in North Carolina, etc. Still, twenty-two states have one or fewer large plant.

[4] Assuming that these are well defined implicitly means either that the market game has a unique equilibrium or that we have chosen a fixed equilibrium selection.

As we remarked earlier, we will usually ignore the requirement that S_1, S_2, B_1, and B_2 be integers. We will say that S_1, S_2, B_1, and B_2 are a *quasi-equilibrium* for the model with S sellers and B buyers if $S_1 + S_2 = S$, $B_1 + B_2 = B$, and the four constraints above are satisfied.

Our first step in analyzing the implications of these constraints is to note that they can be rewritten as:

(B1′) $u_b(S_2, B_2) - u_b(S_2, B_2 + 1) \geq u_b(S_2, B_2) - u_b(S_1, B_1)$

(S1′) $u_s(S_2, B_2) - u_s(S_2 + 1, B_2) \geq u_s(S_2, B_2) - u_s(S_1, B_1)$

(B2′) $u_b(S_1, B_1) - u_b(S_1, B_1 + 1) \geq u_b(S_1, B_1) - u_b(S_2, B_2)$

(S2′) $u_s(S_1, B_1) - u_s(S_1 + 1, B_1) \geq u_s(S_1, B_1) - u_s(S_2, B_2)$

The left-hand sides of the two "stay-in-market-1" conditions (B1′) and (S1′) measure the detrimental "market impact" that the agents have when they move to market 2. In most applications one would expect that the expressions will be positive, e.g. adding one extra firm to a labor market raises the equilibrium wage and thereby reduces profits or adding another seller to a financial market reduces the price that all sellers receive. The right-hand sides measure the degree to which market 2 is more attractive given the current division of buyers and sellers. Agglomeration models typically employ assumptions that make larger markets are more efficient. As a result, one or both of the right-hand sides of (B1′) and (S1′) will usually be positive when market 2 is larger than market 1, but this is not necessary for our results. The rewritten equilibrium conditions show that whether a particular split of buyers and sellers is an equilibrium depends on whether the market impact effect is sufficiently large so as to outweigh the current differences between the utilities in the two markets.

The main assumption of our general theorem concerns the behavior of the market as the number of agents increases holding the seller-buyer ratio $\gamma \equiv S / B$ fixed. One can think of it as imposing three restrictions. The first is that the sellers' and buyers' utility functions in the large finite economies converge to well-defined limits as the number of agents increases. The requirement that payoffs converge rules out some potential

applications. In Krugman's [1991a] model of agglomeration due to increasing returns in product variety, for example, each worker's payoff increase without bound as the number of agents grows.[5] The second restriction is that in the continuum-of-players limit, each buyer's utility is strictly increasing in the proportion of agents who are sellers and vice versa. While we view this requirement as not being very demanding, there are several ways in which it can fail. Section V gives examples where strict monotonicity fails in all finite markets, and an example where the monotonicity condition is satisfied in each finite economy but fails in the limit. The third restriction is that the utility functions must converge to the large population limit at a rate of at least $1/B$; faster rates like $1/B^2$ are allowed but slower rates such as $1/\sqrt{B}$ are not. This third restriction holds for all of the models we have analyzed.

ASSUMPTION A1: There is a non-empty interval $\Gamma = [\underline{\gamma}, \overline{\gamma}] \subset (0, \infty)$ and twice continuously differentiable functions F_s, F_b, G_s, and G_b on Γ with $dF_s/d\gamma < 0$ and $dF_b/d\gamma > 0$ such that the approximations

$$u_s(\gamma B, B) = F_s(\gamma) - G_s(\gamma)/B + o(1/B)$$

and

$$u_b(\gamma B, B) = F_b(\gamma) - G_b(\gamma)/B + o(1/B)$$

hold uniformly in γ when B is large.[6]

One might expect G_s and G_b to be positive, at least for large B, so that larger markets provide higher payoffs, but even if larger markers are more efficient this need

---

[5] In the equilibrium of Krugman's model, the number of "varieties" of manufactured goods that are produced increases linearly with the number of workers. The workers' utility is assumed to be a CES aggregate of their consumption of each variety. It increases without bound as workers are able to divide their consumption into smaller and smaller portions of a larger and larger number of goods. One could, of course, write down many other reasonable specifications of a preference for product variety in which utility would remain bounded even as the number of varieties produced grew without bound. For example, this would happen in a model where each consumer has an ideal point in a fixed product space and varieties just fill up the product space as their number becomes large.

[6] More formally, the assumption on the seller's utility function is that there exists a function $m_s(B)$ with $\lim_{B\to\infty} m_s(B) = 0$ such that $|B(u_s(\gamma B, B) - F_s(\gamma)) - G_s(B)| < m_s(B)$ for all $\gamma \in \Gamma$ and all integers B.

not be true, as the size of the market can influence the division of the gains from trade.[7] For this reason, we should emphasize that our results do not require any restrictions on the signs of G_s and G_b .[8] In particular, in cases where convergence to the limit is at rate faster than $1/B$, the functions G_s and G_b will both be identically equal to 0.

Our next proposition explores the implications of Assumption A1 at configurations where the buyer-seller ratio is the same in both markets. In this case the market-impact effects, which are the left-hand sides of the inequalities in part (a) of the proposition, are proportional to $1/B$ for large B. The scale effect is the difference in utility in two markets of different sizes but the same seller-buyer ratios; Part (b) of the proposition shows that the size of this effect declines at rate $1/B$ as the population size increases, and also goes to zero as the fraction α of agents in market 1 approaches 1/2.

PROPOSITION 2: Assume A1. Let $\underline{\alpha}$ be any positive constant. Suppose that $\gamma \in \Gamma$ and that $\alpha_1, \alpha_2 \in [\underline{\alpha}, 1-\underline{\alpha}]$ with $\alpha_1 + \alpha_2 = 1$. Let $B_1 = \alpha_1 B$, $B_2 = \alpha_2 B$, $S_1 = \gamma B_1$, and $S_2 = \gamma B_2$. Then,

(a) The market impact effects can be approximated uniformly in γ, α_1, and α_2 by

$$u_s(S_i, B_i) - u_s(S_i + 1, B_i) = -F_s'(\gamma)/\alpha_i B + o(1/B)$$

and

$$u_b(S_i, B_i) - u_b(S_i, B_i + 1) = \gamma F_b'(\gamma)/\alpha_i B + o(1/B).$$

(b) The scale effects can be approximated uniformly in γ, α_1, and α_2 by

$$u_s(\gamma B_2, B_2) - u_s(\gamma B_1, B_1) = G_s(\gamma)\frac{(\alpha_2 - \alpha_1)}{\alpha_1 \alpha_2}\frac{1}{B} + o(1/B)$$

and

$$u_b(\gamma B_2, B_2) - u_b(\gamma B_1, B_1) = G_b(\gamma)\frac{(\alpha_2 - \alpha_1)}{\alpha_1 \alpha_2}\frac{1}{B} + o(1/B)$$

---

[7] For example, Schwartz and Ungo (2002) generalize an example in Ellison, Fudenberg and Möbius (2002) to show that when the sellers in a market conduct an $S+1^{st}$ price auction and buyers values are independently drawn from a distribution F with the monotone hazard rate property, then buyers in at least one market are made worse off by the combination of two formerly separate markets.

Proof: The first approximation in part (a) follows immediately from

$$u_s(\gamma B_i, B_i) - u_s(\gamma B_i + 1, B_i) = u_s(\gamma B_i, B_i) - u_s((\gamma + 1/B_i)B_i, B_i)$$
$$= -F_s(\gamma + 1/B_i) + F_s(\gamma) - (G_s(\gamma) - G_s(\gamma + 1/B_i))(1/B_i) + o(1/B_i)$$
$$= -F_s'(\gamma)(1/\alpha_i B) + o(1/B).$$

The second follows from

$$u_b(\gamma B_i, B_i) - u_b(\gamma B_i, B_i + 1) = u_b(\gamma B_i, B_i) - u_b(\gamma B_i(B_i + 1)^{-1}(B_i + 1), B_i + 1)$$
$$= \gamma F_b'(\gamma)(B_i + 1)^{-1} - G_b(\gamma)((B_i + 1)^{-1} - B_i^{-1}) + G_b'(\gamma)\gamma(B_i + 1)^{-2} + o(1/B_i)$$
$$= \gamma F_b'(\gamma)/\alpha_i B + o(1/B).$$

The argument for the first approximation in part (b) is

$$u_s(\gamma B_2, B_2) - u_s(\gamma B_1, B_1) = -G_s(\gamma)[1/B_2 - 1/B_1] + o(1/B)$$
$$= (B_2 - B_1)(B_1 B_2)^{-1} G_s(\gamma) + o(1/B) = G_s(\gamma)(\alpha_2 - \alpha_1)/\alpha_1\alpha_2 B + o(1/B).$$

The argument for the second is the identical but with different subscripts. QED

One way of analyzing agglomeration in a model like this is to ignore the market impact effect and define a split-market equilibrium as an allocation of buyers and sellers that makes them exactly indifferent between the two markets. With this definition, any symmetric model will have an equilibrium where exactly half of the agents are in each market, but this will typically be the only split-market equilibrium, and one can argue that the model would tip away from this unstable knife-edge. This sort of argument implicitly supposes that the market impact effect is small enough to be ignored, at least when the number of agents is large. Part (a) of Proposition 2 shows that the market impact effect is small in the sense of being order $1/B$, but part (b) of the proposition shows that it cannot be ignored since it is no smaller than the scale effect.

Our main theorem is a demonstration that the 50-50 equilibrium is not a "knife-edge." Instead, a model satisfying A1 has a "plateau" of quasi-equilibria where all of the incentive constraints are satisfied, and the size of this plateau (as a fraction of the space

---

[8] Note that when both G_s and G_b are negative, larger markets are less efficient, which tends to favor

of possible market divisions) does not go to zero as the number of agents goes to infinity. For example, it may be that no matter how large B is there is always a quasi-equilibrium with one-third of the buyers in market 1 and two-thirds of the buyers in market 2.

THEOREM 1: Assume A1. Then for any $\varepsilon > 0$ there exists a $\underline{B}$ such that for any integer $B > \underline{B}$ and any integer S with $S/B \in \Gamma$ the model with B buyers and S sellers has a quasi-equilibrium with B_1 buyers in market 1 for every B_1 with

$B_1/B \in [\alpha^*(S/B)+\varepsilon, 1-\alpha^*(S/B)-\varepsilon]$, where $\alpha^*(\gamma) = \max\{0, \frac{1}{2} - \frac{1}{2r^*(\gamma)}\}$ for

$r^*(\gamma) = \max\left(\left|\frac{2G_s(\gamma)}{-F_s'(\gamma)}+1\right|, \left|\frac{2G_b(\gamma)}{\gamma F_b'(\gamma)}+1\right|\right)$.

Proof: We want to show that there exists a $\underline{B}$ such that for any integer $B > \underline{B}$ and any $\gamma \in \Gamma$, the four constraints (S1'), (S2'), (B1') and (B2') are satisfied at the allocation $B_t = \alpha_t B$ and $S_t = \alpha_t \gamma B$ for every $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1-\alpha^*(\gamma)-\varepsilon]$. By symmetry it suffices to show this for all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$.

Our first step is to note that it suffices to show that four simpler constraints obtained by applying first-order approximations to both sides of (S1'), (S2'), (B1'), and (B2') are satisfied for all α_1, α_2, and γ with $\gamma \in \Gamma$ and $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$. The four simpler constraints are:

(AS1) $\quad -F_s'(\gamma) > G_s(\gamma)(\alpha_2 - \alpha_1)/\alpha_1$

(AS2) $\quad -F_s'(\gamma) > G_s(\gamma)(\alpha_1 - \alpha_2)/\alpha_2$

(AB1) $\quad \gamma F_b'(\gamma) > G_b(\gamma)(\alpha_2 - \alpha_1)/\alpha_1$

(AB2) $\quad \gamma F_b'(\gamma) > G_b(\gamma)(\alpha_1 - \alpha_2)/\alpha_2$

The sufficiency of these four constraints is a straightforward consequence of Proposition 2 and our various continuity and differentiability assumptions; Since the arguments for each of the constraints are similar we show this only for (AS1). Suppose

---

equilibria with two active markets.

(AS1) holds for all $\gamma \in \Gamma$ and all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$. Because both sides of (AS1) are continuous in α and γ, and these parameters lie in a compact set, there is a $\delta > 0$ for which

(AS1′) $$-F_s'(\gamma) > G_s(\gamma)(\alpha_2 - \alpha_1)/\alpha_1 + \delta$$

for all $\gamma \in \Gamma$ and all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$. Dividing by $\alpha_2 B$ we get that

(AS1″) $$-F_s'(\gamma)/\alpha_2 B > G_s(\gamma)\frac{\alpha_2 - \alpha_1}{\alpha_1\alpha_2}\frac{1}{B} + \frac{\delta}{\alpha_2 B}$$

for all $\gamma \in \Gamma$ and all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$. Proposition 2 implies that there are functions $m_1(B)$ and $m_2(B)$ that are independent of γ with $\lim_{B\to\infty} B\,|\,m_i(B)\,| = 0$ such that the LHS of (AS1″) is at least $u_s(S_2, B_2) - u_s(S_2+1, B_2) - m_1(B)$ and the first term on the RHS is at most $u_s(S_2, B_2) - u_s(S_1, B_1) + m_2(B)$. Choose $\underline{B}$ so that $\delta > B(m_1(B)+m_2(B))$ for all $B > \underline{B}$. Then, since

$$u_s(S_2, B_2) - u_s(S_2+1, B_2) > u_s(S_2, B_2) - u_s(S_1, B_1) + \frac{\delta}{\alpha_2 B} + m_1(B) + m_2(B)$$

for all $\gamma \in \Gamma$ and all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$, (S1′) is satisfied at $B_i = \alpha_i B$ and $S_i = \alpha_i \gamma B$ whenever $B > \underline{B}$, $\gamma \in \Gamma$, and $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$.

We now show that the four constraints (AS1), (AS2), (AB1) and (AB2) are satisfied for all $\gamma \in \Gamma$ and all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$. We begin with the seller constraints, showing they are satisfied by considering three cases.

Case 1: Suppose $G_s(\gamma) \geq 0$. Since Assumption A1 implies that $-F_s'(\gamma)$ is positive, (AS2) is satisfied for all $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$. (AS1) is equivalent to $\frac{G_s(\gamma)}{-F_s'(\gamma)} < \frac{\alpha_1}{\alpha_2 - \alpha_1}$. Defining $r_s = \left|2\frac{G_s(\gamma)}{-F_s'(\gamma)} + 1\right| = 2\frac{G_s(\gamma)}{-F_s'(\gamma)} + 1$, this is equivalent to $r_s < 2\frac{\alpha_1}{\alpha_2 - \alpha_1} + 1 = \frac{1}{1-2\alpha_1}$ or $\alpha_1 > \frac{1}{2} - \frac{1}{2r_s}$. We have chosen $r^*(\gamma)$ so that $r^*(\gamma) \geq r_s$, so α_1 is bigger than $\frac{1}{2} - \frac{1}{2r_s}$ and (AS1) is satisfied whenever $\alpha_1 \in [\alpha^*(\gamma)+\varepsilon, 1/2]$.

Case 2: Suppose $F_s'(\gamma) < G_s(\gamma) < 0$. In this case (AS1) is obviously satisfied. Moreover, since $-F_s'(\gamma) > -G_s(\gamma)$, $\frac{G_s(\gamma)}{F_s'(\gamma)} < 1$, while $\frac{\alpha_2}{\alpha_2 - \alpha_1} > 1$ all $\alpha_1 \in [\alpha^*(\gamma) + \varepsilon, 1/2]$. Hence, (AS2) is satisfied.

Case 3: Finally, suppose $G_s(\gamma) \leq F_s'(\gamma) < 0$. As in the second case, (AS1) is obviously satisfied. (AS2) is equivalent to $\frac{G_s(\gamma)}{F_s'(\gamma)} < \frac{\alpha_2}{\alpha_2 - \alpha_1}$. In this case, $r_s = \left| 2\frac{G_s(\gamma)}{-F_s'(\gamma)} + 1 \right| = 2\frac{G_s(\gamma)}{F_s'(\gamma)} - 1$, so (AS2) becomes $r_s < 2\frac{\alpha_2}{\alpha_2 - \alpha_1} - 1 = \frac{1}{1 - 2\alpha_1}$. As in the first case, this is equivalent to $\alpha_1 > \frac{1}{2} - \frac{1}{2r_s}$ and we have chosen $r^*(\gamma)$ so that this is true for $\alpha_1 \in [\alpha^*(\gamma) + \varepsilon, 1/2]$.

A nearly identical argument shows that (AB1) and (AB2) are satisfied if $\alpha_1 > \frac{1}{2} - \frac{1}{2r_b}$, for $r_b = \left| 2\frac{G_b(\gamma)}{\gamma F_b'(\gamma)} + 1 \right|$. Again, $r^*(\gamma)$ was chosen so that $\alpha^*(\gamma) = \frac{1}{2} - \frac{1}{2r^*(\gamma)} \geq \frac{1}{2} - \frac{1}{2r_b}$.

QED

Remarks:

1. The size of the quasi-equilibrium plateau identified in theorem 1 is decreasing in the ratios $\left| -2G_s(\gamma)/F_s'(\gamma) + 1 \right|$ and $\left| 2G_b(\gamma)/\gamma F_b'(\gamma) + 1 \right|$. (More precisely, it is decreasing in whichever of the two is the binding constraint provided that the equilibrium plateau is not the entire space.) Intuitively, the scale effect, which is proportional to G, favors tipping, while the market impact effect is proportional to the derivative of F and makes the plateau larger. Inspection of the formula in the theorem reveals that the quasi-equilibrium plateau converges to the full interval as $\left| G_s(\gamma)/F_s'(\gamma) \right| \to 0$ and

$\left|G_b(\gamma)/F_b(\gamma)\right| \to 0$; here the scale effect is absent. The equilibrium plateau shrinks to zero in the limit as either ratio goes to infinity.

2. The theorem provides sufficient but not necessary conditions for the existence of quasi-equilibria, and the proof considers only allocations with the same seller/buyer ratio in each market. We can establish a partial converse: If $\alpha \notin [\alpha^*(\gamma), 1-\alpha^*(\gamma)]$ then for B large the model with B buyers and S sellers does not have an equilibrium with exactly equal seller/buyer ratios in the two markets.[9] In general, though, only one of the four constraints is binding at $\alpha^*(\gamma)$, and the range of splits for which there is a quasi-equilibrium is somewhat larger when the seller/buyer ratios in the two markets are allowed to differ. In Ellison, Fudenberg, and Möbius [2002] we give an exact characterization for the quasi-equilibrium set for the case of competing auctions when buyers' valuations have the uniform distribution. We show that the actual lower bound on the size of the smaller market is strictly smaller than the lower bound established here, but that the two lower bounds converge as the seller-to-buyer ratio increases towards 1.

3. The theorem applies to quasi-equilibria as opposed to equilibria because it ignores the constraint that the numbers of sellers in each market be integer-valued. One might expect that this integer constraint is not very demanding in large economies, but its implications turn out to be fairly complex. Ellison, Fudenberg and Möbius (2002) show that in the case of competing auctions, when the number of buyers is large and $S = \gamma B$, then for any point in the quasi-equilibrium set there is a γ' close to γ for which the model with B buyers and $\gamma' B$ sellers has an equilibrium that is close to the given point. This, however, leaves open the possibility that there might not be equilibria for many or even most values of B and S. Anderson, Ellison, and Fudenberg [2003] show that there is indeed a broad plateau of equilibria for B sufficiently large whenever B/S approaches almost any value of γ, but that there are nongeneric sequences (S^n, B^n) for which the models with S^n sellers and B^n buyers have no equilibria with two active markets. We conclude that the results we derive in this paper about the sizes of the quasi-equilibrium plateaus do reflect what would we would find in a full equilibrium analysis in generic economies.

---

[9] To see this, observe that when $\alpha^*(\gamma) > 0$, at any allocation with the same seller-buyer ratio in both markets, at least one of the buyer or seller constraints is violated for sufficiently large B when $\alpha < \alpha^*(\gamma)$.

3. Krugman's Labor Market Pooling Model

Our first example is Krugman's [1991b] labor market pooling model of industry agglomeration. In this model, the advantage of industry agglomeration is that each firm can better adjust its employment level in response to idiosyncratic productivity shocks.[10]

We treat the model as a two-stage game. In the first stage, F firms and L workers choose between two possible locations. In the second stage each firm i receives an independent productivity shock, ε_i, which is assumed to have mean zero and variance σ^2. Firms observe these shocks and then hire workers from the pool at their location. If firm i hires L_i workers at a wage of w, then its profits are $\pi_i = \alpha + (\beta + \varepsilon_i)L_i - (\delta/2)L_i^2 - wL_i$.[11] Workers are risk-neutral and supply one unit of labor inelastically; their utility is their money wage.

Suppose that F firms and L workers are in a given market. Following Krugman, we suppose that firms are price takers in the labor market. Each firm's labor demand is easily derived from the first-order condition for profit maximization, yielding $L_i^*(w) = (\beta + \varepsilon_i - w)/\delta$.[12] The market-clearing wage is $w^* = \beta - \delta(L/F) + \sum_{j=1}^{F} \varepsilon_j / F$.

At the start of the second stage, before the productivity shocks are revealed, we see that a worker's expected utility is

$$u_w(L,F) = E\left(w^*\right) = \beta - \delta(L/F)$$

Firm i's labor demand at the market-clearing wage is $L_i^*(w^*) = \left(\varepsilon_i + \delta(L/F) - \sum_{j=1}^{F} \varepsilon_j / F\right)/\delta$. Substituting this into firm i's profit function and taking expectations we find after some algebra that

---

[10] Krugman cites Marshall's [1920] discussion (which notes both that larger markets may provide workers with insurance and that firms may benefit from access to skilled labor) as the inspiration for his model.

[11] Our notation departs slightly from Krugman's. We write δ for the parameter he called γ so that we can use γ for the ratio L/F.

[12] Note that some firms may employ a negative number of workers; we follow Krugman in not worrying about this and thereby keeping the model tractable. Note also that the labor demands need not be integers; this can be interpreted as workers splitting their time between several jobs.

$$u_f(L,F) = E\left(\pi_i(L_i^*(w^*), w^*)\right) = \alpha + \frac{\delta}{2}\frac{L^2}{F^2} + \frac{\sigma^2}{2\delta}\left(1 - \frac{1}{F}\right).$$ [13]

Writing γ for L/F (which is analogous to the seller-buyer ratio in the labor market), these utility functions have the form

$$u_w(L,F) = F_w(\gamma) - G_w(\gamma)/F$$
$$u_f(L,F) = F_f(\gamma) - G_f(\gamma)/F$$

for $F_w(\gamma) = \beta - \delta\gamma$, $G_w(\gamma) = 0$, $F_f(\gamma) = \alpha + \frac{\sigma^2}{2\delta} + \frac{\delta}{2}\gamma^2$, and $G_f(\gamma) = \frac{\sigma^2}{2\delta}$. Note that $F_w'(\gamma) = -\delta < 0$ and $F_f'(\gamma) = \delta\gamma > 0$. Hence, Assumption 1 is satisfied for any γ.[14]

Now we analyze the choice of location in the first stage. As in our general model, we suppose that agents take their market impact into account when choosing a location; this contrasts with Krugman's analysis, which directly defines equilibrium to mean that expected profits and wages are equal in the two markets.[15] Although our solution concept does not require that wages be equalized, expected wages are the same in both markets in the equilibria we analyze.[16]

Theorem 1 now implies that this model has a plateau of equilibria with two active markets, and gives a characterization of how unequal in size the two markets can be.

COROLLARY 1. Fix γ and let $\alpha^* = \frac{\sigma^2}{2\delta^2\gamma^2 + 2\sigma^2}$. Then, for any $\varepsilon > 0$ there exists an $\underline{F}$ such that for all $F > \underline{F}$, the quasi-equilibrium set of the model with F firms and

---

[13] This formula corrects an error in Krugman's equation (C.10).

[14] The example satisfies Assumption A1 even though the firm profit function need not be monotone in F.

[15] This is equivalent to assuming that both firms and workers ignore their market impact. An alternate way to justify Krugman's solution would be to consider a game where firms choose locations first, workers then choose locations, and there is a continuum of workers so each worker has no market impact. A problematic aspect of this justification is the contrast between assuming that workers migrate instantly across the two markets to arbitrage away any expected wage differences due to unexpected plant location choices, but are immobile when productivity shocks are realized and wages become different. Our analysis corresponds to the case where workers are stuck in the locations they choose and can neither relocate in response to expected wage differences if they are surprised by firms' location decisions, nor in response to realized wage differences; our qualitative findings should carry over to models with anything less than perfect mobility.

[16] Wages are equal in equilibrium, but not following a deviation from the equilibrium.

$\lfloor \gamma F \rfloor$ workers includes splits with F_1 firms in market 1 for every F_1 with $F_1 / F \in [\alpha^* + \varepsilon, 1 - \alpha^* - \varepsilon]$.

Proof: Since $G_w(\gamma) = 0$, and $G_f(\gamma) > 0$, the width of the plateau is determined by the firms' incentive constraint, and we find that $r^*(\gamma) = 2G_f(\gamma)/F_f'(\gamma)\gamma + 1 = \sigma^2/\delta^2\gamma^2 + 1$, and $\alpha^* = \sigma^2/2\delta^2\gamma^2/\left(1 + \sigma^2/\delta^2\gamma^2\right) = \dfrac{\sigma^2}{2\delta^2\gamma^2 + 2\sigma^2}$.

Remarks:

1. The equilibrium plateau can cover almost the whole space, that is, a market with a tiny fraction of the firms can coexist with a large one. This occurs when productivity shocks are very small ($\sigma^2 \approx 0$), when there are strong decreasing returns at the firm-level ($\delta \to \infty$), or when the worker-firm ratio is large ($\gamma \to \infty$). When the parameters are at the opposite extremes, the equilibrium plateau shrinks down to a point. Intuition for each result can be obtained by comparing the market impact of a firm that moves to the larger market, thus bidding up wages, to the advantage of larger markets, namely that a firm has a smaller wage impact when raising or lowering labor demand in response to its productivity shock. For example, when there are strong decreasing returns (δ large), there is little advantage to being in a large market because a firm won't want to increase or decrease its labor demand very much in response to a productivity shock, and a firm would bid up wages substantially if it moved to the larger market because the other firms' labor demands are inelastic.

2. Although we would be hesitant to put much stock in any calibration of this model, one way to try to do so would be to note that the ratio of parameters that determines the size of the equilibrium plateau also determines the variability of firm-level employment. Specifically, $Var\left(L_i^* / E(L_i^*)\right) = ((F-1)/F)^2\sigma^2/\delta^2\gamma^2$. If one assumed that $Var\left(L_i^* / E(L_i^*)\right)$ was 1 when $F = 10$, then $\alpha^* \approx 0.28$; this says that the smaller market needs to be about three-eighths of the size of the larger market to be viable ($0.28/0.72 \approx 0.38$).

3. Corollary 1 does not include an assumption that the number of firms be sufficiently large. Theorem 1 required an assumption about the number of buyers being large only

because we assumed just that the utility functions were approximately equal to $F(\gamma)-G(\gamma)/B$ when B was large. In this application, the expressions for the utility functions are exact, and hence the formula for the bounds on the equilibrium plateau holds for all F.[17]

4. In the quasi-equilibria described in Corollary 1 the workers receive exactly equal utility in both markets. Thus, they would still be quasi-equilibria if workers did not consider their market impact. This would be the appropriate assumption if one modeled the workers as a continuum of agents of mass L.

5. In the Krugman model with a continuum of workers of mass L, only the number of firms in each market must be an integer, and hence all of the quasi-equilibria described in the corollary are equilibria. An alternate way to state the corollary would be to say that the model with F firms and a mass L of workers has an equilibrium with F_1 workers in market 1 for any integer F_1 with $F_1/F \in [\alpha^* + \varepsilon, 1-\alpha^* - \varepsilon]$.

6. The generalization of the result to a model with N locations (and a continuum of workers) would be that there exists an equilibrium with F_1 firms in market 1, F_2 firms in market 2, ..., and F_N firms in market N provided that every market has either zero firms or at least $\alpha^*/(1-\alpha^*)$ times as many firms as the largest market.

4. Pagano's Security Market Model with Heterogeneous Traders

Our next example is a two-population version of Pagano's [1989] model of competing securities markets. In Pagano's model markets serve to diversify endowment risk. Specifically, he considered a two-stage game. In the first stage, N agents simultaneously choose which of two markets to attend. In the second stage, each agent i receives a random endowment $K_{0i} = K_0 + e_i$ of a single risky asset, where the e_i are i.i.d. draws from a symmetric distribution with mean 0 and variance σ_e^2. Agents then trade by simultaneously submitting demand curves to a market maker who executes trades at the market-clearing price. The asset pays a random dividend d, so an agent who

[17] The bounds are exact bounds for the existence of an equilibrium with equal worker-firm ratios in the two markets. There are equilibria with unequal worker-firm ratios (with a higher ratio in the smaller market) for smaller values of α.

keeps K_i shares has random final wealth $w_i = dK_i + Rp(K_{0i} - K_i)$, where p is the market price of the asset and R is the risk-free rate of return. Both p and asset demands are allowed to be negative, and there is not free disposal.

We modify Pagano's model by replacing the assumption that K_0 is the same for all N agents with the assumption that S "sellers" have $K_0 = 1$ and B "buyers" have $K_0 = -1$; this sort of ex-ante asymmetry in expected purchases seems reasonable for some applications. (As will be clear shortly, the names "buyer" and "seller" are only suggestive, as a "seller" with a negative endowment shock may end up purchasing shares, and a "buyer" may sell them.) To simplify the algebra, we will specialize to the case where the dividend has a symmetric distribution with mean 0 and variance σ^2.[18] This means that the asset is a "bad," but preserves the key aspect of the model, which is that larger markets are more efficient because there is less undiversifiable social risk; allowing for a positive dividend would only complicate the algebra without altering the nature of our conclusions.

As in Pagano, we suppose that agents' preferences over distributions of wealth are described by a mean-variance utility function defined directly on the space of wealth distributions $V(w_i) = E(w_i) - (b/2)Var(w_i)$. Further following the original, we look for a symmetric equilibrium in which the agents simultaneously submit linear demand functions $D_i(p)$.[19] In the equilibrium we find firm i submits the demand curve $D_i(p) = K_{0i}/(N-1) - R(N-2)p/(N-1)b\sigma^2$.[20] Equilibrium asset holdings are $K_i^* = K_{0i}/(N-1) + (N-2)Z/(N-1)N$, where $Z = S - B + \sum_{i=1}^{N} e_i$ is the realized

---

[18] We have already simplified by assuming that the endowment of cash is 0.

[19] $D_i(p)$ is the agent's desired asset holding, e.g. if $D_i(p) = K_{0i}$ then agent i makes no trades. Note that $D_i(p)$ will depend on K_{0i}. To replicate Pagano's results we assume that agents maximize their mean-variance utility in the first stage, but not in the second. In the second stage, they instead maximize the expectation over all possible realizations of the vector of endowment shocks of their expected utility conditional on the shocks. This would be equivalent to expected utility maximization if mean-variance preferences were an expectation of a utility function defined on realized wealth, but they are not.

[20] The details of this and subsequent calculations are in the Appendix. The calculations show that this is the unique equilibrium in which agents' demand curves are linear and downward sloping. We do not know if there are other equilibria. Klemperer and Meyer [1989] discuss conditions sufficient for equilibrium uniqueness in a loosely related game where agents submit supply functions.

net supply of the asset. Thus, equilibrium allocations are better diversified when N is larger; agents offset their endowment shocks more completely because the sum of the other $N-1$ agents' demand curves is steeper and the aggregate endowment risk they share is also smaller. This is the source of the larger market's efficiency advantage. To simplify notation, let $q(\gamma) = (\gamma - 1)/(\gamma + 1)$. A series of calculations shows that the equilibrium payoffs are approximated by

$$u_s(\gamma B, B) = b\sigma^2\left[\frac{q(\gamma)^2}{2} - q(\gamma)\right] - \frac{b\sigma^2}{2(\gamma+1)}\left[\sigma_e^2 + 4q(\gamma)^2 b^2\sigma^2\sigma_e^2 + b^2\sigma^2(\sigma_e^2 + \sigma_e^4) - 2q(\gamma)\right]\frac{1}{B} + o\left(\frac{1}{B}\right)$$

$$u_b(\gamma B, B) = b\sigma^2\left[\frac{q(\gamma)^2}{2} + q(\gamma)\right] - \frac{b\sigma^2}{2(\gamma+1)}\left[\sigma_e^2 + 4q(\gamma)^2 b^2\sigma^2\sigma_e^2 + b^2\sigma^2(\sigma_e^2 + \sigma_e^4) + 2q(\gamma)\right]\frac{1}{B} + o\left(\frac{1}{B}\right)$$

Thus, both the seller's and buyer's utility functions have the form $F(\gamma) - G(\gamma)/B + o(1/B)$ assumed in assumption A1. Specifically,

$$F_s(\gamma) = b\sigma^2\left[q(\gamma)^2/2 - q(\gamma)\right], \quad F_b(\gamma) = b\sigma^2\left[q(\gamma)^2/2 + q(\gamma)\right],$$

$$G_s(\gamma) = b\sigma^2\left[\sigma_e^2 + 4q(\gamma)^2 b^2\sigma^2\sigma_e^2 + b^2\sigma^2(\sigma_e^2 + \sigma_e^4) - 2q(\gamma)\right]/2(\gamma+1),$$

and

$$G_b(\gamma) = b\sigma^2\left[\sigma_e^2 + 4q(\gamma)^2 b^2\sigma^2\sigma_e^2 + b^2\sigma^2(\sigma_e^2 + \sigma_e^4) + 2q(\gamma)\right]/2(\gamma+1).$$

Thus $F_s'(\gamma) = -4b\sigma^2(\gamma+1)^{-3} < 0$, and $F_b'(\gamma) = 4\gamma b\sigma^2(\gamma+1)^{-3} > 0$, so assumption A1 is satisfied for any γ. The seller's market impact effect reflects that sellers expect to sell shares, and that adding a seller to a market lowers the expected price. Note that as B goes to infinity, the total expected welfare in the population (summing over buyers and sellers) converges to $-b\sigma^2 q(\gamma)^2/2$, which is the utility of holding the average endowment.

We can now immediately apply Theorem 1 to show that our two-sided version of Pagano's model has a plateau of equilibria in which different-sized markets coexist. The algebra is simplest if there are equal numbers of buyers and sellers (i.e. $\gamma = 1$). We then have $G_s(\gamma)/-F_s'(\gamma) = G_b(\gamma)/\gamma F_b'(\gamma) = \left(\sigma_e^2 + b^2\sigma^2(\sigma_e^2 + \sigma_e^4)\right)/2$. This gives

COROLLARY 2. Consider our Pagano model with equal numbers of buyers and sellers.

Let $r^* = 1 + \sigma_e^2 + b^2\sigma^2(\sigma_e^2 + \sigma_e^4)$, and let $\alpha^* = 1/2 - 1/2r^*$. Then, for any $\varepsilon > 0$ there exists an $\underline{N}$ such that for all $N > \underline{N}$ the equilibrium set of the model with N sellers and N buyers includes the splits with N_1 buyers and sellers in market 1 for every N_1 with $N_1/N \in [\alpha^* + \varepsilon, 1 - \alpha^* - \varepsilon]$.

Remarks:

1. The size of the equilibrium plateau is inversely related to the size of the endowment shocks. When the endowment shocks are trivial ($\sigma_e^2 \approx 0$), efficiency differences are unimportant and a market need only have a tiny fraction of the traders to be viable. When endowment shocks become extremely large ($\sigma_e^2 \to \infty$), or dividend shocks become extremely large ($\sigma^2 \to \infty$) or agents become extremely risk-averse ($b \to \infty$) markets can coexist only if they are almost equal in size. To interpret the scale of the variances, note that we have assumed that the mean endowment of a seller is 1 and the mean endowment of buyer is –1. If $\sigma_e^2 = 1$ and $b^2\sigma^2 = 0.1$ then the interval in the corollary is approximately (0.27, 0.73), so a market would need to have about one-quarter of the traders to be viable.[21]

2. The second remark after Theorem 1 implies that these bounds are tight. If α is strictly outside the specified interval, then for sufficiently large B there will not be an equilibrium with αB buyers in market 1.

5. Some Models that Don't Fit the Assumptions

In this section we discuss a few models of agglomeration that don't fit into our framework, explain why they don't fit, and suggest ways in which they might be modified so that an analysis like ours would apply.

A. A Matching Model

[21] Calibrating a mean-variance utility function can be problematic. For example, even with $b = 0.1$ a decision-maker would prefer getting 0 for sure to getting a lottery ticket that pays 0 with probability $1 - p$ and 100 with probability p for any $p < 0.8$.

Consider a matching model with two types of agents, men (type m) and women (type w). Each participant has exogenous probability $q > 0$ of becoming unable to participate, and these chances are independent across agents. To simplify the algebra, we suppose that each type of agent gets utility $1/(1-q)$ if matched with an agent of the other type, and utility 0 otherwise. Thus in the continuum limit, an agent who is sure to be matched if he participates has expected payoff of $(1-q)/(1-q) = 1$, and in general, if there are M agents of type m and W agents of type w, the utilities of the men and women are $\min(1, 1/\gamma)$ and $\min(1, \gamma)$, respectively, where $\gamma = M/W$.

In the finite markets, let M' and W' be the realized numbers of agents of each type who are able to participate, and let $x = M' - W'$ be the realized excess supply of men. Then the payoffs are $u_m(\gamma) = \Pr(x \leq 0) + \Pr(x > 0) E\left[\frac{W'}{M'} \mid x > 0\right]$ and $u_w(\gamma) = \Pr(x \geq 0) + \Pr(x < 0) E\left[\frac{M'}{W'} \mid x < 0\right]$. The excess supply is the difference of two binomials with the same success probability but different sample sizes, so holding the ratio γ to be fixed and greater than 1, and sending M and W to infinity, $\lim_{M\to\infty} \Pr(x \leq 0) = o(1/M)$.[22] Thus for the case $\gamma > 1$, $u_m(M, W) = 1/\gamma + o(1/M)$ and $u_w(M, W) = 1 - o(1/M)$, and the utility of the agents on the "short side" of the market does not satisfy Assumption A1, as it is insensitive to γ.

Two aspects of the model are noteworthy. First, convergence to the continuum limit is at a faster rate than the $1/N$ rate required by Assumption A1. This is consistent with A1, as it corresponds to the case where the functions G_s and G_b are both 0. Second, although both players care about the "buyer-seller" ratio γ when there are a finite number of agents, this is not true in the continuum limit. The limit of the women's utility, $F_w(\gamma)$, is one for any $\gamma > 1$. This is not consistent with A1's requirement that the functions F_s and F_b have non-zero derivatives. One way to modify the model to make it compatible

[22] In fact, Chernoff's theorem (Billingsley 1995, p. 151) shows that the $\lim_{M\to\infty} \Pr(x \leq 0) \leq \rho^M$ for some constant $\rho < 1$, so that the convergence is (at least) exponentially fast.

with A1 might be to assume that a woman's expected payoff from a match is increasing in M/W because the women are able to select from the available men.

B. Preference for Variety

Consider the following greatly simplified version of Gehrig [1998]. In the first stage, S firms and B consumers choose one of two locations. The firms at each location are then placed uniformly around a Hotelling circle of possible products and the consumers draw ideal points from a uniform distribution. Firms play a standard competition-on-a-circle pricing game.

Assume that the transportation cost parameter is such that the equilibrium price in a market with S firms is $p^* = c + 1/S$. Each firm in expectation sells to $1/S$ of the consumers, so the firm's profit function is $u_s(S,B) = B/S^2$. The consumer's expected utility would be his or value, minus the price, minus an expected mismatch cost, $u_b(S,B) = v - (c + 1/S) - 1/(2S)$.

Neither of these functions satisfies A1. On the firm side, profit is converging to zero as $S \to \infty$ with B/S fixed so $F_s(\gamma) = 0$. This prevents the application of theorem 1, but a direct analysis of the seller's incentive constraints (S1′) and (S2′) shows that they are satisfied with equal buyer-seller ratios in the two markets over an interval of sizes, just as in the proof of Theorem 1 .

On the consumer side, utility is converging to $F_b(\gamma) = v - c$, which is independent of γ. More fundamentally, even in the finite case buyers don't care about how many other buyers are in the market, and the (B1′) and (B2′) constraints are satisfied only if $S_1 = S_2$. Substituting this into the sellers' constraints yields $1 + 2/S_1 + 1/S_1^2 \geq B_1/B_2 \geq 1 - 2/S_1 - 1/S_1^2$, so that as S goes to infinity, the set of ratios $\alpha = B_1/B_2$ that are consistent with equilibrium collapses to the knife-edge $\alpha = 1$. To modify this model so that it had an equilibrium plateau, a buyer market-impact effect would be necessary. For example, adding a buyer to a market would reduce the utility of the other buyers if there were congestion costs or if firms had increasing marginal costs.

C. Preference for Variety II

Fujita [1988] develops a general equilibrium model of spatial agglomeration (with a continuum of locations) in which a preference for being able to buy a variety of products locally is offset by higher land prices in the city center. A greatly simplified two-location version of this model with a sufficient amount of land per agent would tip to having all consumers and firms in one location.

It would fail to satisfy our assumptions for a couple reasons. First, Fujita assumes that the preference for the variety is such that a consumer who purchases a small quantity x_i of each good i at a price of p_i receives utility $u_b = \sum_{i=1}^{S} -x_i \log(x_i) - x_i p_i$. This utility function increases like $\log(S)$ as the agent is able to divide his or her consumption among a larger number of goods (at a constant price). When utility increases without bound, the set of interior equilibrium will typically collapse to a knife-edge. For example, if the seller's utility has the form $u_s(S,B) = K(B)\left(F_s(\gamma) - G_s(\gamma)/B\right)$, then when $B_2 = (1/2+\varepsilon)B$ and the market impact effect is

$u_s(S_2,B_2) - u_s(S_2+1,B_2) \approx 2K(B_2)F_s'(\gamma)/B(1+2\varepsilon) \approx 2(K(B/2)+\varepsilon K'(B/2))F_s'(\gamma)/B(1+2\varepsilon)$

and the scale effect is

$u_s(S_2,B_2) - u_s(S_1,B_1) \approx \left(K(B_2) - K(B_1)\right)F_s(\gamma) \approx 2\varepsilon BK'(B/2)F_s(\gamma)$. Holding ε fixed, the market impact effect will be smaller than the scale effect for large B if $B^2K'(B)/K(B)$ diverges. This ratio is $B/\log(B)$ if $K(B) = \log(B)$, and is aB if $K(B) = B^a$.

Second, the primary market impact that a firm or consumer has when moving to a market is that he increases the land rent. If land is sufficiently plentiful, then the equilibrium rent would always be the value of land in its alternate use, agriculture, which is assumed constant. Hence, there would be no market impact effect. To create a market impact effect, one could modify the model so that consumers and firms had to outbid a heterogeneous population of farmers for land.

6. Conclusion

This paper shows that Assumption A1 is sufficient for market models to have a plateau of quasi-equilibria. The assumption also yields an easily computed formula that provides a lower bound on the "width" of this plateau. In our opinion, Assumption A1 is

not very restrictive; it applies to many if not most market models, including those of Krugman [1991b] and a two-type version of Pagano [1989]. The leading cases where it fails are when in the continuum limit, some agents are indifferent about the ratio of "buyers" to "sellers," and when per agent payoffs converge to zero or infinity as the number of agents goes to infinity.

That said, the generality of Assumption A1 and the simplicity of Theorem 1 have been obtained by leaving our results incomplete in a couple ways. First of all, the sufficient conditions in Theorem 1 apply to quasi-equilibria with exactly the same buyer/seller ratio in each market; this leaves open the question of how much broader the quasi-equilibrium plateau is when one also considers the possibility of quasi-equilibria in which the buyer-seller ratios differ. Ellison, Fudenberg, and Möbius [2002] provides a detailed analysis of the case of competing uniform-price auctions, in which each buyer purchases a single unit, and the price in a market with k goods for sale is the (k+1)st highest buyer value.

Secondly, Theorem 1 concludes that the incentive constraints are satisfied, but does not reveal when these constraints can be satisfied along with the constraints that the numbers of each type of agent in each market should be integers; these constraints are addressed in Anderson, Ellison and Fudenberg [2003].

We have written this paper to emphasize that even when larger markets are more efficient there will typically still be a plateau of equilibria with two active markets. Our assumptions also encompass models in which there are no scale effects, and even some models in which larger markets give lower payoffs to both buyers and sellers, e.g. because of crowding effects. One implication of our theorem is that as long as the crowding effects are no stronger than is allowed under assumption A1 there will be an equilibrium plateau that includes splits in which some markets are substantially larger than others. Hence, the observation that an activity is concentrated in a small number of locations need not imply that there are increasing rather than decreasing returns to agglomeration.

Appendix

We follow Pagano in assuming agents have mean-variance preferences over wealth distributions and maximize their utility when choosing between markets in the first stage. Also following Pagano, we suppose that in the second stage, agents choose the demand function $D(p)$ that maximizes the expected value of what their conditional preferences would be if they knew the full vector of endowments but did not know the random dividend.[23]

Write $Z_{-i} \equiv \sum_{j \neq i} K_{0j}$ for the aggregate endowment of all agents other than i. Consider the possibility of an equilibrium in which each agent j submits the linear demand curve $D_j(p) = AK_{0j} - mp$.[24] The sum of the demand curves submitted by agent i's opponents is then $AZ_{-i} - (N-1)mp$. Thus, the ability to submit a demand curve gives agent i the power to choose the quantity of the risky asset $K_i(Z_{-i})$ that he will receive conditional on every realization of Z_{-i}.[25] If agent i receives allocation K_i in state Z_{-i} then the market-clearing condition $AZ_{-i} - (N-1)mp + K_i = Z_{-i} + K_{oi}$ implies that the price must be $P(K_i, Z_{-i}) = -((1-A)Z_{-i} + K_{0i} - K_i)/(N-1)m$. Hence, the expectation over the endowment shocks of agent i's mean-variance utility conditional on the endowment shock when he receives allocation $K_i(Z_{-i})$ is

$$E_{Z_{-i}}\left[E\left(dK_i(Z_{-i}) + R(K_{0i} - K_i(Z_{-i}))\hat{P}(Z_{-i}) \mid Z_{-i}\right) - \frac{b}{2}Var\left(dK_i(Z_{-i}) + R(K_{0i} - K_i(Z_{-i}))\hat{P}(Z_{-i}) \mid Z_{-i}\right)\right]$$

where we have written $\hat{P}(Z_{-i})$ for $P(K_i(Z_{-i}), Z_{-i})$.

---

[23] Since the equilibrium prices are fully revealing, this would be utility maximizing behavior for agents with standard expected-utility preferences. Mean-variance utility, however, does not have the property that $V(w) = E_z(V(w \mid Z))$, so the assumed behavior is not utility-maximizing; we regard it as a "behavioral" assumption.

[24] We will see that there is always an equilibrium of this form.

[25] This is true provided that $K_i(Z_{-i})$ and $(1-A)Z_{-i} - K_i(Z_{-i})$ are monotone increasing. To obtain allocation $K_i(Z_{-i})$ agent i can submit the downward-sloping demand curve

To maximize this expectation over all functions $K_i(Z_{-i})$, one can maximize the expression within the brackets, i.e. maximize over K_i for each Z_{-i}. For a fixed Z_{-i} the expression for the conditional mean-variance utility inside the brackets simplifies to $V(w \mid K_i, Z_{-i}) = R(K_{0i} - K_i)P(K_i, Z_{-i}) - \frac{b}{2}K_i^2\sigma^2$. The first-order condition for maximizing this expression is $RP(K_i, Z_{-i}) - R(K_{0i} - K_i)\frac{\partial P}{\partial K_i} - b\sigma^2 K_i\Big|_{K_i^*(Z_{-i})} = 0$. Substituting $-\frac{(1-A)Z_{-i} + K_{0i} - K_i}{(N-1)m}$ for $P(K_i, Z_{-i})$ and $\frac{1}{(N-1)m}$ for $\frac{\partial P}{\partial K_i}$ gives

$$K_i^*(Z_{-i}) = \frac{2K_{0i} + (1-A)Z_{-i}}{2 + (N-1)mb\sigma^2/R}.$$

To complete the derivation we need to find the demand curve that produces this allocation and use the equilibrium condition $D_i^*(p) = AK_{0i} - mp$ to solve for A and m. Writing X for the denominator in the equation for $K_i(Z_{-i})$, the price in state Z_{-i} that corresponds with this allocation is

$$\hat{P}(Z_{-i}) = -\frac{(1-A)Z_{-i} + K_{0i} - (2K_{0i} + (1-A)Z_{-i})/X}{(N-1)m} = \frac{2-X}{X(N-1)m}K_{0i} + \frac{(1-A)(1-X)}{X(N-1)m}Z_{-i}.$$

The inverse of this function is

$$\hat{P}^{-1}(p) = \frac{(X-2)}{(1-A)(1-X)}K_{0i} + \frac{X(N-1)m}{(1-A)(1-X)}p.$$

From the expression in footnote 22, the demand curve that implements $K_i^*(Z_{-i})$ is

$$D_i^*(p) = (1-A)P^{-1}(p) + K_{0i} + (N-1)mp = \left(\frac{X-2}{1-X} + 1\right)K_{0i} + \left(\frac{X}{1-X} + 1\right)(N-1)mp.$$

Note that this is indeed linear and it satisfies the equilibrium condition if and only if $-1/(1-X) = A$ and $1/(1-X) = -1/(N-1)$. The solution to these equations is $A = 1/(N-1)$ and $X = N$. The latter implies $m = ((N-2)/(N-1))R/b\sigma^2$.

$D(p) = \hat{P}^{-1}(p)(1-A) + K_{0i} + (N-1)mp$, where $\hat{P}^{-1}$ is the inverse of the market-clearing price consistent with $K_i(Z_{-i})$, $\hat{P}(Z_{-i}) = -((1-A)Z_{-i} + K_{0i} - K_i(Z_{-i}))/(N-1)m$.

Agent i's equilibrium asset holding as a function of the endowment vector are

$$K_i^*(K_0) = \frac{2K_{0i} + (1-A)Z_{-i}}{X} = \frac{(1+A)K_{0i} + (1-A)Z}{X} = \frac{K_{0i}}{N-1} + \frac{N-2}{N-1}\frac{Z}{N}.$$

The equilibrium price is $p^*(Z) = -\frac{b\sigma^2}{R}\frac{Z}{N}$.

We now compute the mean-variance utility that each agent maximizes in the first stage. Agent i's wealth as a function of the dividend and endowment shocks is

$$\begin{aligned} w_i &= RpK_{0i} + K_i(d - Rp) = -b\sigma^2\frac{Z}{N}K_{0i} + \left(\frac{N-2}{N-1}\frac{Z}{N} + \frac{K_{0i}}{N-1}\right)(d + b\sigma^2\frac{Z}{N}) \\ &= \left(\frac{N-2}{N-1}\right)\frac{Z}{N}d + \left(\frac{(N-2)b\sigma^2}{N-1}\right)\frac{Z^2}{N^2} + \left(\frac{N}{N-1}\right)\frac{K_{0i}d}{N} - \left(\frac{(N-2)b\sigma^2}{N-1}\right)\frac{Z}{N}K_{0i}. \end{aligned} \quad (1)$$

We now compute the mean and variance of this expression.

Since the expected dividend is 0 and dividends and endowments are independent,

$$E(w_i) = \left(\frac{(N-2)b\sigma^2}{(N-1)}\right)\frac{E(Z^2)}{N^2} - \left(\frac{(N-2)b\sigma^2}{N-1}\right)E\left(\frac{Z}{N}K_{0i}\right).$$

Note that $\frac{S-B}{N} = \frac{\gamma-1}{\gamma+1}$, so $\frac{Z}{N} = \frac{S-B+\sum_{i=1}^{N}e_i}{N} = \frac{\gamma-1}{\gamma+1} + \frac{\sum_{i=1}^{N}e_i}{N}$ and

$$\frac{E(Z^2)}{N^2} = \left(\frac{\gamma-1}{\gamma+1}\right)^2 + \frac{\sigma_e^2}{N},$$

where σ_e^2 is the variance of e_i. We also have

$$E\left(\frac{Z}{N}K_{0i}\right) = E\left(\left(\frac{\gamma-1}{\gamma+1} + \sum_{i=1}^{N}e_i/N\right)(e_i \pm 1)\right) = \frac{\sigma_e^2}{N} \pm \frac{\gamma-1}{\gamma+1},$$

where the "+" corresponds to sellers. Hence,

$$\begin{aligned} E(w_i) &= \left(\frac{(N-2)b\sigma^2}{N-1}\right)\left(\left(\frac{\gamma-1}{\gamma+1}\right)^2 + \frac{\sigma_e^2}{N}\right) - b\sigma^2\frac{N-2}{N-1}\left(\frac{\sigma_e^2}{N} \pm \frac{\gamma-1}{\gamma+1}\right) \\ &= \left(\frac{(N-2)b\sigma^2}{N-1}\right)\left(\left(\frac{\gamma-1}{\gamma+1}\right)^2 \pm \frac{\gamma-1}{\gamma+1}\right) \end{aligned}$$

where the "+" now corresponds to buyers.

Now we need to compute the variance of w_i. The expression (1) for w_i is a sum of four terms. Z and d are independent, and d has mean 0, so the variance of the first term is $Var\left(\frac{Z}{N}d\right) = E\left(\frac{Z^2}{N^2}d^2\right) - E\left(\frac{Z}{N}d\right)^2 = E(d^2)E\left(\frac{Z^2}{N^2}\right) = \frac{\sigma^2\sigma_e^2}{N} + \sigma^2\left(\frac{\gamma-1}{\gamma+1}\right)^2$.

Next,

$$Var\left(\frac{Z^2}{N^2}\right) = Var\left(\left(\frac{\gamma-1}{\gamma+1}\right)^2 + 2\frac{\gamma-1}{\gamma+1}\frac{\sum_{i=1}^N e_i}{N} + \frac{\left(\sum_{i=1}^N e_i\right)^2}{N^2}\right)$$

$$= 4\left(\frac{\gamma-1}{\gamma+1}\right)^2\frac{\sigma_e^2}{N} + o\left(\frac{1}{N}\right).$$

The third random variable has a smaller variance,

$$Var\left(\frac{K_{0i}d}{N}\right) = \frac{\sigma^2\sigma_e^2}{N^2} = o\left(\frac{1}{N}\right).$$

The fourth random variable has

$$Var\left(\frac{Z}{N}K_{0i}\right) = \frac{1}{N^2}Var\left((\pm 1 + e_i)\sum_{j=1}^N e_j\right) = \frac{1}{N^2}Var\left(\pm\sum_{j=1}^N e_j + \sum_{j=1}^N e_i e_j\right)$$

$$= \frac{1}{N^2}Var\left(\sum_{j=1}^N e_j\right) + \frac{1}{N^2}Var\left(\sum_{j=1}^N e_i e_j\right)$$

$$= \frac{\sigma_e^2}{N} + \frac{\sum_{j\neq i}E(e_i^2e_j^2)}{N} + \frac{Var(e_i^2)}{N^2} = \frac{\sigma_e^2}{N} + \frac{\sigma_e^4}{N} + o\left(\frac{1}{N}\right).$$

The first random variable is uncorrelated with the second and fourth, the covariance of the first and third is $\sigma^2\sigma_e^2/N^2$, the covariance of the second and fourth is $o(1/N)$ because

$$Cov\left(\frac{Z^2}{N^2}, \frac{Z}{N}K_{0i}\right) = \pm Cov\left(\frac{Z^2}{N^2}, \frac{Z}{N}\right) + Cov\left(\frac{Z^2}{N^2}, \frac{Z}{N}e_i\right) = 0 + Cov\left(\frac{Z^2}{N^2}, \frac{Z_{-i}+e_i}{N}e_i\right)$$

$$= Cov\left(\frac{Z^2}{N^2}, \frac{Z_{-i}}{N}e_i\right) + Cov\left(\frac{Z^2}{N^2}, \frac{e_i^2}{N}\right) = \frac{1}{N}Cov\left(\frac{Z^2}{N^2}, e_i^2\right) = o\left(\frac{1}{N}\right),$$

and the covariances of the second and fourth terms with the third term are $o(1/N)$ because the products of the standard deviations are $o(1/N)$. Hence, the variance of w_i is approximated to order $1/N$ by the sum of the variances of each term:

$$Var(w_i) = \left(\frac{N-2}{N-1}\right)^2 Var\left(\frac{Z}{N}d\right) + \left(\frac{(N-2)b\sigma^2}{N-1}\right)^2 Var\left(\frac{Z^2}{N^2}\right) + \left(\frac{(N-2)b\sigma^2}{N-1}\right)^2 Var\left(\frac{Z}{N}K_{0i}\right) + o\left(\frac{1}{N}\right)$$

Approximating the first constant to order $1/N$, the other constants to order 1, and the variances to order $1/N$ gives

$$Var(w_i) = \left(1-\frac{2}{N}\right)\left(\frac{\sigma^2\sigma_e^2}{N} + \sigma^2\left(\frac{\gamma-1}{\gamma+1}\right)^2\right) + b^2\sigma^4 4\left(\frac{\gamma-1}{\gamma+1}\right)^2\frac{\sigma_e^2}{N} + b^2\sigma^4\frac{\sigma_e^2+\sigma_e^4}{N} + o\left(\frac{1}{N}\right)$$

$$= \left(\frac{\gamma-1}{\gamma+1}\right)^2\sigma^2 + \left[-2\left(\frac{\gamma-1}{\gamma+1}\right)^2\sigma^2 + \sigma^2\sigma_e^2 + 4\left(\frac{\gamma-1}{\gamma+1}\right)^2 b^2\sigma^4\sigma_e^2 + b^2\sigma^4(\sigma_e^2+\sigma_e^4)\right]\frac{1}{N} + o\left(\frac{1}{N}\right).$$

Agent *i*'s mean-variance utility before the endowment shocks are realized is thus

$$V(w_i) = E(w_i) - \frac{b}{2}Var(w_i) = \left(\frac{(N-2)b\sigma^2}{N-1}\right)\left(\left(\frac{\gamma-1}{\gamma+1}\right)^2 \pm \frac{\gamma-1}{\gamma+1}\right)$$

$$-\frac{b}{2}\left(\frac{\gamma-1}{\gamma+1}\right)^2\sigma^2 - \frac{b}{2}\left[-2\left(\frac{\gamma-1}{\gamma+1}\right)^2\sigma^2 + \sigma^2\sigma_e^2 + 4\left(\frac{\gamma-1}{\gamma+1}\right)^2 b^2\sigma^4\sigma_e^2 + b^2\sigma^4(\sigma_e^2+\sigma_e^4)\right]\frac{1}{N} + o\left(\frac{1}{N}\right)$$

$$= \left[\frac{1}{2}\left(\frac{\gamma-1}{\gamma+1}\right)^2 \pm \frac{\gamma-1}{\gamma+1}\right]b\sigma^2 - \left[\pm 2\frac{\gamma-1}{\gamma+1} + \sigma_e^2 + 4\left(\frac{\gamma-1}{\gamma+1}\right)^2 b^2\sigma^2\sigma_e^2 + b^2\sigma^2(\sigma_e^2+\sigma_e^4)\right]\frac{b\sigma^2}{2N} + o\left(\frac{1}{N}\right)$$

where the + term corresponds to buyers.

References

Anderson, Robert, Glenn Ellison, and Drew Fudenberg, "A Non-Tipping Theorem: Integer Constraints and Equilibria in Two-Sided Markets," in preparation, 2003.

Billingsley, Patrick, *Probability and Measure* (New York, NY: Wiley, 1995).

Dumais, Guy, Ellison, Glenn and Edward L. Glaeser, "Geographic Concentration as a Dynamic Process," *The Review of Economics and Statistics*, LXXXIV (2002), 193-204.

Ellison, Glenn, Drew Fudenberg, and Markus Möbius, "Competing Auctions," unpublished, MIT and Harvard University, 2002.

Ellison, Glenn and Edward L. Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, CV (1997), 889-927.

Fujita, Masahisa, "A Monopolistic Competition Model of Spatial Agglomeration – Differentiated Product Approach," *Regional Science and Urban Economics*, XVIII (1988), 87-124.

Gehrig, Thomas, "Competing Markets," *European Economic Review*, XLII (1998), 277-310.

Henderson, J. Vernon, "The Sizes and Types of Cities," *American Economic Review*, LXIV (1974), 640-656.

Klemperer, Paul D. and Margaret A. Meyer, "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica*, LVII (1989), 1243-1277.

Krugman, Paul, "Increasing Returns and Economic Geography," *Journal of Political Economy*, XCIX (1991a), 483-499.

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

Kyle, Albert S., "Informed Speculation with Imperfect Competition," *Review of Economic Studies*, LVI (1989), 317-356.

Pagano, Marco, "Trading Volume and Asset Liquidity," *Quarterly Journal of Economics*, CIV (1989), 255-274.

Schwartz, Jesse A. and Ricardo Ungo, "Merging Auction Houses," unpublished, Vanderbilt University, 2002.

Vayanos, Dimitri, "Strategic Trading and Welfare in a Dynamic Market," *Review of Economic Studies*, LXVI (1999), 219-254.