COLUMBIA
BUSINESS
SCHOOL

PaineWebber

WORKING PAPER SERIES
in
MONEY, ECONOMICS
AND FINANCE

THE ECONOMICS OF INCREASING RETURNS

Geoffrey Heal
Columbia Business School

April 1998
PW-97-20

COLUMBIA BUSINESS SCHOOL
COLUMBIA UNIVERSITY
http://www.columbia.edu/cu/business/wp/

The Economics of Increasing Returns

G. M. Heal

Columbia Business School

Abstract

Increasing returns is the source of some of the most powerful metaphors and intuitions in economics. Foremost among them are Adam Smith's statement that the division of labor is limited by the extent of the market, and his discussion of the relationship between scale and economies of specialization in a pin factory. There is a weakness, strictly an error, in Adam Smith's analysis. Two phenomena that he grouped together and saw as integral to economic progress are in fact inconsistent. These are increasing returns with the consequent gains from specialization and the efficiency of the invisible hand. We now know that a society cannot have both, at least if one interprets the efficiency of the invisible hand as the Pareto efficiency of the competitive equilibrium, our only rigorous interpretation.

This paper reviews the implications of increasing returns for several areas of economics: resource allocation and welfare economics; the micro foundations of macroeconomics; product variety and imperfect competition; information and information technology; economic growth; international trade. These cover the fields in which increasing returns cause departures from the results otherwise available. These departures are rather significant. Recognizing increasing returns affects the possibility of market equilibrium, can introduce sticky prices, causes economies to lock-in to inefficient technologies and introduce path-dependence, affects the possibility of continuing growth, produces hard problems for regulators, and changes our conception of the effects of international trade. All in all, increasing returns can change quite radically our view of how the economy operates. They make the economy seem more complicated, and pose a challenge to our vision of a benign and powerful invisible hand.

This paper in the introduction to a book of readings on increasing returns to scale and their implications for a range of aspects of economics. The book will be published by Edward Elgar in the series The International Library of Critical Writings in Economics under the title The Economics of Increasing Returns. The readings to be included in the book are listed at the end of the paper.

1. Introduction

Increasing returns present economists with a paradox. It is the source of some of the most powerful metaphors and intuitions in economics. Foremost among them are Adam Smith's statement that the division of labor is limited by the extent of the market, and his discussion of the relationship between scale and economies of specialization in a pin factory. Texts that display a U-shaped cost curve and a minimum efficient scale of production also invoke increasing returns, as do those emphasizing fixed costs. Applications of microeconomics to industrial organization, public policy, and regulation stress increasing returns. A student entering our subject might therefore feel that increasing returns are a key characteristic of the economic world, which indeed they are. But here is the paradox: in spite of the evocative power of the increasing returns metaphor, and its congruence with reality, at the very deepest level we have little to say about it. The economic equivalent of Newton's Laws, the first theorem of welfare economics, is of limited value in a world of increasing returns. There need be no competitive equilibrium, so that although a competitive equilibrium is still Pareto efficient, the significance of the first welfare theorem is limited. This is serious, as the efficiency of the competitive outcome lies at the heart of the most robust policy prescriptions emerging from economics. These are prescriptions about the efficiency of competitive markets and their superiority relative to alternatives.

This points to a weakness, strictly an error, in Adam Smith's analysis. It tells us that two phenomena that he grouped together and saw as integral to economic progress are in fact inconsistent. These are increasing returns with the consequent gains from specialization and the efficiency of the invisible hand. We now know that a society cannot have both, at least if one interprets the efficiency of the invisible hand as the Pareto efficiency of the competitive equilibrium, our only rigorous interpretation[1]. Recall Smith's often quoted but still compelling description of the pin-making process:

“The ... business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations. But though they were poor, and therefore, but indifferently accommodated with the necessary machinery, they could, when they exerted themselves, make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of a middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of the forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they all had wrought separately and independently, and without any of them having been educated to this particular business, they certainly could not each of them have made twenty, perhaps not one pin in a day." (Smith 1977)

[1] Weitzman’s paper in section 2 below makes the same point.

This is a very lucid description of how scale allows specialization and consequently greater productivity. Smith clearly believed that there is a minimum efficient scale for the operation of the machinery needed to enhance productivity, so that small and isolated markets could not justify the investment needed to achieve efficiency. Smith's famous description of the invisible hand is also well worth quoting again:

“Every individual .. neither intends to promote the public interest, nor knows how much he is promoting it. He intends only his own security, his own gain. And he is in this led by an invisible hand to promote an end which was no part of his intention. By pursuing his own interest he frequently promotes that of society more effectively than when he really intends to promote it." (Smith 1977)

These are two key themes in The Wealth of Nations: scale, specialization, and economic efficiency; and the invisible hand. And as we shall see in detail below they do not go together. This is a central problem in economic analysis of increasing returns. Recognition of this conflict is not new: John Stewart Mill was aware of this in the mid nineteenth century[2].

“It is obvious, for example, how great an economy of labour would be obtained if London were supplied by a single gas or water company instead of the existing plurality. While there are even as many as two, this implies double establishments of all sorts, when one only, with a small increase, could probably perform the whole operation equally well. Were there only one establishment, it could make lower charges, consistent with making the rate of profit now realized. But would it do so? Even if it did not, the community in the aggregate would still be a gainer.

It is, however, an error to suppose that the prices are ever permanently kept down by the competition of these companies. When competitors are so few, they always end by agreeing not to compete. They may run a race of cheapness to ruin a new candidate, but as soon as he has established his footing they come to terms with him. When, therefore, a business of real public importance can only be carried on advantageously on so large a scale as to render the liberty of competition almost illusory, it is an unthrifty dispensation of the public resources that several costly sets of arrangements should be kept up for the purpose of rendering to the community this one service. It is much better to treat it at once as a public function." (Mill 1848 quoted in Quinzii 1992 .)

This is a remarkably modern-sounding argument, noting quite clearly the conflict between the technological efficiency that comes from scale, and the economic and organizational efficiency that comes from competition. This is still at the heart of one part of the policy dilemma arising from increasing returns. Today we would include with technological economies of scale the economies arising from standardization and from the use of information. They arise from increasing returns to the size of the user base for a product or system, and from the fixed cost nature of information acquisition. Natural monopolies such as telephone companies illustrate the importance of technological scale

[2] This quote is used by Martine Quinzii, whose book (Quinzii 1992) is an excellent review of the theory of increasing returns.

economies: Microsoft illustrates the power of owning a standard that underlies a technology, giving rise to economies of scale associated with the size of the user base. The market power of firms such as Standard and Poor or Dun and Bradstreet illustrate he importance of informational economies of scale. The essence of anti-trust policy is to find ways of keeping these economics of scale or of standardization while bringing in competition. Current moves to deregulate telecommunications and electric power supply in the United States are trying to do just this.

A source of our intellectual weakness in dealing with increasing returns lies in the Faustian bargain made in adopting the techniques of convex analysis. They give us clean and powerful results, but at the cost of preventing the analysis of increasing returns. Convex analysis cuts us off from some of our roots, namely Adam Smith and the intuitions of basic microeconomics. Of course, it has given us a great deal in exchange, but we are perhaps not sufficiently aware of the price paid. The price is the inability of mainstream economic theory to address questions relating to increasing returns.

It may help to make explicit some points that are implicit here. Firstly, with increasing returns in production, there need be no prices at which markets clear: at any prices, we may find excess demands for some goods and excess supplies of others. Hence reliance on the price mechanism cannot be recommended, at least not for the normal reasons. The papers in section 2 on the micro foundations of macroeconomics emphasize this point, and also set out the implications for a range of macro phenomena. Figure 1 illustrates this, showing a firm using a single input, labor, to produce a single output. The technology shows increasing then decreasing returns to scale, and the demand for labor jumps discontinuously from zero to a strictly positive number L* as the wage passes through w*. It will never demand an amount of labor between zero and L*. If the supply is in this range, then there is no price at which demand and supply can be brought into

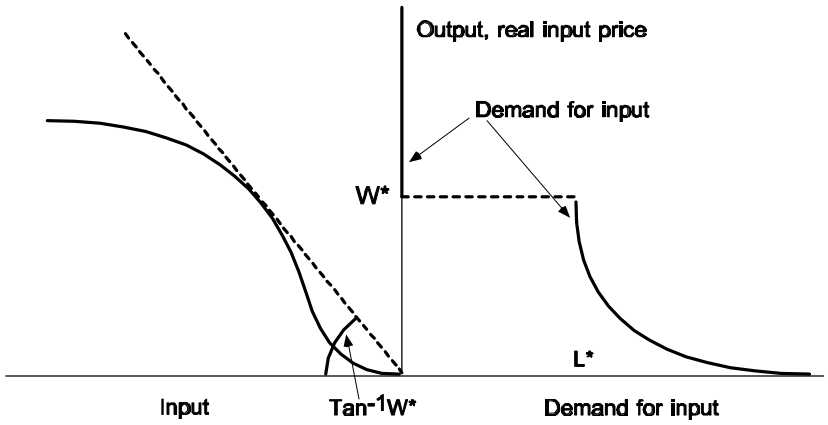

Figure 1

equality.

Secondly, convexity implies the absence of increasing returns. Most economic models assume convexity of production possibility sets, which implies non-increasing returns to scale and to proportions. Non-convexity of a production possibility set can arise from

either increasing returns to scale, or increasing returns to proportions, or from both. The paper by Masaiko Aoki in section 1 addresses an interesting world in which we have increasing returns to scale but diminishing returns once one input level is fixed, so that, for a given scale of operation, firms' production possibility sets are convex and a price mechanism can be used. Another basic microeconomic point, important in the context of regulation, is that a firm with increasing returns cannot reasonably be modeled as behaving competitively: at some prices it will want to supply an infinite amount of output, and at others none at all. For prices above minimum average cost, profits increase indefinitely with the scale of operation. To have a determinate scale of operation, the firm has to be faced with a declining demand curve or with quantitative constraints.

These and other themes are developed in the papers that follow. The papers are arranged in six main sections: resource allocation and welfare economics; micro foundations of macroeconomics; product variety and imperfect competition; information and information technology; economic growth; international trade. These cover the fields in which increasing returns cause departures from the results otherwise available. These departures are rather significant. Recognizing increasing returns affects the possibility of market equilibrium, can introduce sticky prices, causes economies to lock-in to inefficient technologies and introduce path-dependence, affects the possibility of continuing growth, produces hard problems for regulators, and changes our conception of the effects of international trade. All in all, increasing returns can change quite radically our view of how the economy operates. They make the economy seem more complicated, and pose a challenge to our vision of a benign and powerful invisible hand.

It is puzzling that although economies of scale are undoubtedly important in reality, our belief in the invisible hand, in the efficiency of competition, seems verified by observation and experience, although not supported by current theory. This suggests that our understanding of economies with increasing returns is far from complete: there may be a role for competition and markets in allocating resources in the presence of increasing returns that we have not yet understood.

2. Resource Allocation and Welfare

The papers in this section address basic theoretical questions about resource allocation and economic efficiency in the presence of increasing returns. The point of departure is the fact that we cannot rely on the market mechanism, as with increasing returns a competitive market may lack equilibrium. This particular point is documented, and its implications developed, in section 2 on the microfoundations of macroeconomics, and in particular in the papers “Rational rationing with increasing returns”, “Stable disequilibrium prices” and “Increasing returns, chaotic dynamics and the Phillips curve”. These show how an economy may fail to have an equilibrium and develop the implications of an economy being permanently away from a market-clearing equilibrium.

2.1 Alternatives to the Market

If we cannot rely on finding a Pareto efficient competitive equilibrium, what are our options? One possibility is to consider a non-market approach to resource allocation, or at least an approach in which market forces are guided or supplemented by a more visible hand. The first six papers are written in this spirit. In “Decentralization and computation in resource allocation” Kenneth Arrow and Leo Hurwicz are trying to formalize the concept of market socialism, first formulated in the 1930s by Oscar Lange (1936-37). This they do successfully, showing that in principle the issues of ownership of property and choice of resource allocation mechanism can be separated. In the process they establish that a modified version of this regime may operate with a certain degree of success even in the presence of increasing returns. They also establish a very important point, which is that a market mechanism can be viewed as an algorithm for solving constrained maximization problems. This was an important insight into the functioning of markets and the connection between market and shadow prices. In the Arrow-Hurwicz model, the market mechanism is a mechanism in which agents receive price signals from a central agency, and respond with quantity signals, namely demands and supplies. It is a mechanism that in an important sense is decentralized, and Arrow and Hurwicz were the first to consider this property in a formal framework. This is one of the few features of the market not captured by the classical economists: von Hayek (1945) set it out beautifully.

The remaining papers in this group look at different ways of formulating the visible hand. Heal's paper “Planing without prices” reverses the types of signals used in the Arrow-Hurwicz model: a central agency sends out quantity signals to firms and receives back marginal value products, which are shadow prices for goods in each firm. We have here a dual to the market mechanism, at least in terms of information flows. As firms are not profit maximizing price takers, their scale of operation is determinate and they respond continuously to the signals received. Heal shows that such a non-market approach can lead to an efficient allocation of resources, although at quite a high cost in terms of centralization of information.

Heal's second paper “Planning, prices and increasing returns” and “An investment planning process of an economy with increasing returns” by Masaiko Aoki both investigate the possibility of making some use of market forces, but supplementing them by a visible hand. Heal shows that if firms seek not to maximize profits but to follow a dynamic process in which profits are always rising, then this can lead them to optimal solutions. There is extensive use of prices and of profit incentives in this model, but in the context of a process in which agents seek a direction of improvement rather than an overall optimum. Hajime Hori's paper “The structure of equilibrium points in Heal's process” considers a set of mathematical questions of relevance to both the papers by Heal. These concern the existence of solutions to the differential equations defining the adjustment processes and the possibility of converging to stationary points that are not local optima. Aoki's paper considers the following case: technologies show increasing returns when all inputs are varied together, but there is one input, say I, such that if the amount of this is taken as fixed, then the production sets are convex. In this case, Aoki demonstrates, we can use a market for all inputs other than I, and use a non-price allocation method for I. The paper “A quantity-quantity algorithm for planning in an

economy with increasing returns" by Jacques Cremer displays a resource-allocation algorithm that has very powerful properties. It converges to a stationary point that is always a global optimum. The cost is that the amount of information exchanged is great, and the process has little connection with markets or prices. It illustrates a point to which we return: good performance with increasing returns requires the exchange of more information than in the a competitive market. Finally, Heal's paper on "The equivalence of saddle-points and optima for non-convex programs" establishes that the equivalence of saddle-points of a Lagrangean and constrained optima, which lies at the heart of the Kuhn-Tucker theorem and of the Arrow-Hurwicz paper, holds for a large class of non-convex problems. These are problems that can be transformed into convex problems: they are diffeomorphic to, or topologically similar to, convex problems. Arrow and Hurwicz had a related idea in their paper, where they discussed transforming and concavifying the Lagrangean and then using gradient processes to locate a saddlepoint of a new Lagrangean. However, Heal shows that if a problem is transformable into a convex problem, then even without the transformation, a saddle-point of the Lagrangean is an optimum.

2.2 General Equilibrium

The remaining papers in the first section deal with the general equilibrium model, although of course not in its competitive form. Given the possible non-existence of a competitive equilibrium, a natural question is: are there other equilibria that are both consistent with increasing returns and also efficient? After all, efficiency is what attracts us to the competitive model.

Policy responses have concentrated on the theory of regulation and specifically on the issue of marginal cost pricing. A central question has been whether pricing at marginal cost is appropriate in the presence of increasing returns, given that it will typically imply losses for entities practicing it. An early paper by Harold Hotelling (1938) is a classic in this field. Hotelling sets out the basic argument in favor of marginal cost pricing in a partial equilibrium framework. Another classic paper by William Baumol and David Bradford (1970) investigates how the need to break even should influence the pricing policy of a regulated firm. Prior to the publication of Roger Guesnerie's paper "Pareto optimality in a non-convex economy," these papers dominated our thinking and economists were comfortable that marginal cost pricing was the correct approach. It is a natural generalization of a competitive equilibrium to the case of non-convexities in production. Since Hotelling's paper on utility pricing (1938), we had known that marginal cost pricing is necessary for efficiency, and had also believed that given appropriate second order conditions it was sufficient. Guesnerie, and subsequently Brown and Heal, showed that this is wrong: an economy may have many marginal cost pricing equilibria, all of which are Pareto inefficient. In "Equity, efficiency and increasing returns" Donald Brown and Geoffrey Heal give a simple geometric example of an economy where for some distributions of a given total endowment between agents, all marginal cost pricing equilibria are Pareto inefficient, whereas for other distributions some are efficient and some inefficient. This suggests that there is a relationship between the distribution of endowments and the efficiency of equilibria. It also shows that the classical orthogonality

of equity and efficiency is a product of the convexity assumption. In “On marginal cost pricing with given tax-subsidy rules” Paulina Beato and Andreu Mas-Collel establish a stronger result: there may be non-convex economies in which all marginal cost pricing equilibria are not only Pareto inefficient but also production inefficient, i.e., they are not even on the aggregate production frontier. Beato and Mas-Collel also establish the existence of a general equilibrium at which firms price at marginal cost, as does “On a general existence theorem for marginal cost pricing equilibria” by Donald Brown, Geoffrey Heal, Ali Kahn and Rajiv Vohra. In “Marginal vs. average cost pricing in the presence of a public monopoly” Brown and Heal give a simplified version of an existence proof, using essentially geometric arguments.

All of this is very negative: it seems to imply that there is no market framework that offers the prospect of attaining an efficient pattern of resource use in the context of increasing returns. In “Is perfect price discrimination really efficient? Welfare and existence in general equilibrium” Aaron Edlin, Mario Epelbaum and Walter Heller go to the natural next step, and ask whether there is a market framework that will work with increasing returns. They start from the following partial equilibrium intuition. At an efficient equilibrium prices must be set at marginal costs. In this cases firms will probably not cover costs. However, Pareto efficient allocations must generate consumer surplus that is in excess of the losses incurred by marginal cost pricing. Perfect price discrimination could in principle extract this consumer surplus, and therefore could make efficient allocations profitable. The same intuition is stated in Brown and Heal (1980): there it is shown that an efficient equilibrium can be supported by individualized two-part tariffs, that is, two-part tariffs in which the fixed charge depends on the identity of the agent. This is clearly a form of price discrimination. This intuition is also clearly stated in the papers by Dixit and Stiglitz, Spence, and Heal in section 3 on product variety and imperfect competition. In fact, one of Heal's propositions is that, in a very simple model, perfect price discrimination leads to Pareto efficiency. All of these papers are about product variety in the context of fixed costs and imperfect competition. Here there is a trade-off between variety and fixed costs: the more variety, the better off are consumers, but the greater is the total of fixed costs. The intuition is that a commodity should be produced if revenues plus consumer surplus cover costs, and the optimum amount is then that at which price and marginal cost are equated. If price discrimination is possible, the consumer surplus is absorbed into profits, and profit maximization leads to the right outcome. Edlin, Epelbaum, and Heller show that under some reasonable conditions, this intuition carries over to a complete general equilibrium model. One further point about price discrimination: it rules out inefficient equilibria of the type illustrated by Guesnerie, Brown and Heal, and Beato and Mas-Collel. The cost of these strong efficiency properties is the need for information about the characteristics of each agent, something not needed by producers in a competitive equilibrium.

Finally, “Competitive equilibria with quantity-taking producers and increasing returns to scale” by Jacques Dreze and Pierre Dehez investigates what kind of equilibria one can formalize with increasing returns, if we aim to get close to a competitive rather than a regulatory framework. The equilibria that they describe are a generalization of competitive equilibria. They preserve the decentralized informationally-minimal

character of competitive equilibria even in the non-convex case. It is therefore not surprising that they are not efficient when they do not coincide with competitive equilibria.

What emerges from this review of alternatives to the competitive market for the case of increasing returns? There are some suggestions, but no conclusions. Firstly, outright price-taking profit maximization will not work. However, the planning literature suggests the effectiveness of adjustment processes in which agents seek to improve but not maximize. Maximization leads to the discontinuities in behavior that cause non-existence of equilibrium. It restricts firms to the convex hulls of non-convex production sets, and prevents them from getting inside non-convexities. Marginal cost pricing equilibria could be stable under processes that seek increases in rather than a maximum of profits.

The paper by Edlin, Epelbaum, and Heller formalizes another valuable intuition. This is that price discrimination can provide a way of reaching efficiency with increasing returns. A key aspect of price discrimination is that it requires information about individual preferences, at least about willingness-to-pay for the increasing returns goods. Given this information, Edlin and co-authors show that the intuition about consumer surplus described above can be carried through. This is fully consistent with the planning literature's conclusion that achieving efficiency with increasing returns requires more information to flow between agents than in a competitive system (see Heal 1973), and requires in particular a departure from the anonymity of agents in a competitive equilibrium. All of the planning processes that work with increasing returns are non-anonymous and information-intensive relative to the competitive system. In a sense, the Edlin, Epelbaum, and Heller results show a way of obtaining and using the extra information in a market context.

3. Microfoundations of Macroeconomics

An intriguing aspect of increasing returns is that in the context of price-taking profit maximizing behavior it seems to provide a rigorous microeconomic foundation for a number of disequilibrium phenomena. We noted in the introduction that there will be many cases in which there is no competitive equilibrium, no market-clearing price, in an economy with increasing returns. In view of this, it is perhaps not surprising that this is a productive framework for the study of disequilibrium phenomena. The paper “Stable disequilibrium prices” by Geoffrey Heal shows that in such a situation, the economy may be stuck at a price that does not clear the market: such prices may be stable, so that there is a persistent configuration at which there is excess demand or supply. In labor market terms, we could have persistence of inflation or unemployment. The paper “Chaotic price dynamics, increasing returns and the Phillips curve” by Graciela Chichilnisky, Geoffrey Heal, and Yun Lin takes this point further. It shows that in a region around a price that is stable but non-market-clearing, prices may move over time in a chaotic fashion, always remaining near the crucial price, yet oscillating irregularly around it. Both of these papers indicate that there can be rigorous microeconomic foundations for the phenomena of

price stickiness and markets that do not clear, which is an important issue in the context of macroeconomic behavior.

Martin Weitzman's paper “Increasing returns and the foundations of unemployment theory” also argues that increasing returns can provide a microeconomic explanation of the existence of equilibria with unemployment. Weitzman constructs a simple model with increasing returns in production, in which there are many equilibria, some with unemployment. There is nothing that any agents can do to move the system from an unemployment equilibrium to one with full employment: in fact, attempts to make such a move might worsen the situation.

We can also find in increasing returns an explanation for why firms may sometimes choose to ration by quantity rather than price: this is the point of “Rational rationing and increasing returns” by Heal. With a non-convex production possibility set, a firm has a discontinuous demand curve for its input. There are certain demand levels that it will never choose whatever the prices it faces. Its demand will jump over these because of the non-convexity. These may be the output levels that are most profitable for a supplier. In such a situation, the supplier can gain by quoting a price that causes the firm to pick a demand level above the discontinuity, and then forcing it back into the region of the discontinuity by quantitative rationing. This is of interest because quantity rationing is a phenomenon that characterizes unemployment equilibria.

The paper “Price-output dynamics and returns to scale” applies increasing returns directly at the macro level, showing that they can be destabilizing. They tend to give rise to patterns of cumulative causation, and to vicious and virtuous circles whereby deviations from an equilibrium or initial condition are amplified. This is really the same point that drives the results in the papers on “Increasing Returns, Networks and Standards” in section 4.

4. Product Variety and Imperfect Competition

The papers in this section study models of imperfect competition in the presence of increasing returns, checking what kind of inefficiencies emerge in the equilibria in these cases. In particular they focus on the variety of products and the degree of product differentiation at equilibrium. The motivation is the proliferation, apparently to the point of redundancy, of slightly different products in imperfectly competitive markets. The paper by Avinash Dixit and Joseph Stiglitz “Monopolistic competition and optimum product diversity” sets up a simple model of product differentiation and increasing returns, which has led to an extensive literature. The issue here, as in the Heal and Spence papers that follow, is whether the market will lead to too much too little product variety The Dixit-Stiglitz paper provides a neat and general framework for looking at these issues.

Geoffrey Heal’s paper on “Spatial structure in the retail trade: a study in product differentiation with increasing returns” looks at a simpler and more specific model. In this context it asks the same question: will a combination of imperfect competition and

increasing returns lead to too much or too little product differentiation? In this model the size of the market is a parameter, and the answer is that large markets are over-served and small ones under-served, where over- and under-served mean respectively that too many or too few product varieties are produced relative to the social optimum. Thus people with minority tastes will have less choice than is optimal, whereas those with mainstream tastes will have more. There will be too few operas and too many musicals: to much TV coverage of majority tastes and too little of minority tastes. One of Heal's results, which relates to the paper by Edlin, Epelbaum, and Heller, is that with perfect price discrimination the variety of products is optimal.

Michael Spence's paper “Product selection, fixed costs and monopolistic competition” continues the same theme. Spence also shows that under certain conditions price discrimination will lead to efficient product mixes. He also gives conditions under which the product variety is too small (when products are strongly complementary) and conditions under which the opposite is true. An issue that Spence highlights, which is also central to Edlin, Epelbaum, and Heller, is that when a new product is introduced, it changes the demand for other products (positively for complements, negatively for substitutes) and so leads to externalities to other producers. These are not taken into account by the introducer of a new product, and represent a reason for non-optimal product choices.

After reviewing the literature on equilibrium, imperfect competition and increasing returns, one is left with a clear impression that, while there are models that are appealing in specific contexts, there are none of any generality. There is no clear paradigm for thinking about how markets allocate resources in a modern imperfectly competitive economy with increasing returns in some sectors: what we have instead is models of a limited number of special cases.

5. Information and Information Technology

Two distinct but related phenomena are the subject of this section. One is the economies of scale that arise from information technology, which are mainly a recent phenomenon. The other is the economies of scale associated with the use of information itself, which is a long-standing issue that has only recently been noted. Its importance has undoubtedly increased with the growth of our ability to collect and process information.

The economies of scale associated with information technology are linked to the size of the user population. In the case of networks, the link is clear. The more people use a communications network or system, the more valuable it is to any one of them. By buying a fax machine, I make the fax machines of my friends and associates more useful to them: likewise, by joining an e-mail system I make the existing e-mail connections of my correspondents more valuable. In these cases there is a clear increase in value and willingness-to-pay linked to size. Operating systems and other standards have a similar property. The more PCs use Windows, the more attractive it is for software and hardware developers to produce products for Windows machines, whereas a little-used operating system will attract few products targeted specifically for its users. The conclusion of the

Heal paper “Spatial structure in the retail trade: a study in product differentiation with increasing returns” is relevant here. If design and manufacture have fixed costs, then there are too few products designed for networks or standards with few users, and too many for those with many users. This reinforces the winner-take-all characteristic of these markets.

Paul David's paper “Clio and the economics of QWERTY” sets out the facts of a classical example of the powers of an established user base and shows how in the case of the layout of a typewriter keyboard this lead to the adoption of an inefficient design. The paper “Competing technologies, increasing returns, and lock-in by historical events” by Brian Arthur focuses on this point in more detail, and shows in a formal model how the same phenomenon can lead to the adoption of inefficient outcomes which can then be very difficult to displace. In the Arthur paper the increasing returns are not technological: they arise from the process of adoption. As Arthur indicates, the implications of this type of increasing returns through adoption are far-reaching and becoming an important part of a world in which high technology products are a part of everyday life. They lead to dominance by a single firm or technology, as any other equilibrium is unstable. Once one technology has more users than an other, its costs fall, and there is more support for it, reinforcing its initial lead.

Geoffrey Heal’s paper “The economics of networks” looks at a similar set of issues, but in a different context. In this case the concern is with the scale economies that emerge in networks where the value to a user depends on how many other people are on that network. This is the case of so-called network externalities, an important phenomenon not just in communication networks but also more generally. The adoption of standards, particularly in computer hardware and software, has these characteristics, as do the markets for products such as fax machines whose utility depends on who else has them, and systems such as electronic mail, whose value is again enhanced by an increase in the user base. This is not a technological return to scale, but its impact on the market and on the possibility of a competitive outcome is very similar to large-scale increasing returns. The key conclusion from this paper is that ultimately we should expect one firm to dominate: equilibria in which there are several competing firms are intrinsically unstable. Once one gets ahead of the others, positive feedback reinforces this. The winner is likely to take all, and may be hard to displace. Microsoft illustrates this nicely.

The economies of scale arising from the use of information are also very simple to understand in principle. The information needed for a set of decisions or a production process is a fixed cost independent of how many times the decision is made or the process is run. The costs a portfolio manager incurs to research a security are the same whether she buys 1,000 or 100,000 units. The costs a lawyer incurs to understand a precedent are the same whether he uses it once or many times. The costs a retailer incurs in finding a supply of salable products are the same whether total sales are large or small. In all of these cases the costs referred to are costs of acquiring information, which are fixed costs with respect to the scale of the ultimate activity. So on a per activity basis they decline with the scale of the activity and represent a form of increasing returns. This is the point made by the paper “Informational economies of scale” by Robert Wilson and in

"A nonconcavity in the value of information" by Roy Radner and Joseph Stiglitz. They prove, in rather different frameworks, that the use of information is characterized by increasing returns.

The papers in this section have interesting implications for the importance of increasing returns. They imply that modern technological developments are bringing increasing returns to the forefront, both through the characteristics of the emerging technologies and through the service they provide, which is the ability to collect and process large quantities of information.

6. Economic Growth

Allyn Young's paper "Increasing returns and economic progress" returns to the point of departure of this introduction. Young looks carefully at the idea that the division of labor depends on the extent of the market, and notes that the opposite is also true. The extent of the market depends on the division of labor. Division of labor leads to specialization, greater productivity and so greater incomes. This increases the market. So the division of labor and the scale of the market must expand in concert. Young wants to understand whether this is possible. He has in his mind an interesting model of growth and specialization: he sees growth as accompanied by more roundabout production methods, which lead naturally to greater specialization and division of labor. Growth leads to more and more subdivision of the production process. This seems an interesting intuition, broadly consistent with casual empiricism, and not captured by any formal growth models. It has some semblance to evolutionary models in biology, where evolution leads to increasing complexity and longer food chains.

Intuition suggests that increasing returns should provide a stimulus to economic growth. They will permit increasing efficiency and competitiveness as an economy grows. Kenneth Arrow was the first to formalize something close to this in "The economic implications of learning by doing." His famous learning-by-doing model is driven by a dynamic version of increasing returns, with essentially the same type of cumulative causation as the later models by Heal ("Macrodynamics and returns to scale, section 2) and Arthur and Heal in section 4, although with a different underlying mechanism.

In "Optimum savings with economies of scale" Avinash Dixit, James Mirrlees, and Nicholas Stern look at the optimal growth in an economy with increasing returns, as does Graciela Chichilnisky in "Existence and characterization of optimal growth paths including models with non-convexities in utilities and technologies." Dixit, Mirrlees, and Stern characterize optimal growth paths for a particular model: Chichilnisky focuses on general conditions for the existence of optimal paths. In the model of Dixit et al. there is one sector and so one produced good, which may be consumed or invested, as usual. The unique feature is that there are economies of scale in the creation of capital goods, i.e., in investment. It therefore pays not to invest continuously but to wait until there is a reasonable amount of output put aside to invest, and then invest all of it at once, so taking advantage of scale economies. The key questions concern the size and frequency of these intermittent investments. The authors show that in some simple cases the results resemble

those of the earlier literature on the transactions demand for cash, in which agents have a fixed cost of changing securities into cash and a sequence of payments that have to be made in cash. Chichilnisky looks at a more general model. By solving some rather hard technical problems she establishes general sufficient conditions for the existence of optimal paths using a neat and direct argument. She proves that the set of feasible paths is compact and the objective to be maximized is continuous, both in the same topology, one with a natural relationship to the discount rate applied to future utilities.

Paul Romer's paper “Increasing returns and long-run growth” started an extensive literature on endogenous growth. The growth models developed in the 1950s and 1960s, based on Solow's 1956 model, all had the rather unappealing property that growth naturally comes to an end. The growth process is self-limiting, and can only be perpetuated by exogenous factors such as technical change or population growth. It was then natural to ask whether an economy could endogenously generate continuing growth, and Romer's intuition is that increasing returns can provide a positive response. He assumes knowledge to be internally generated by investment in research and development, with increasing returns to the use of knowledge, and models a dynamic competitive equilibrium. This combination of technical progress and increasing returns to knowledge, which is the integral of technical progress, can lead to paths on which output per capita can grow forever, possibly without bound. As an aside, the possibility of unbounded growth in per capita output suggests that consideration of environmental constraints be added.

7. International Trade

Conventional models of international trade all assume constant returns to scale in production. Their results depend heavily on this assumption. Extending the theory to a world of increasing returns poses a serious challenge: little of the conventional structure carries over. The choice of an equilibrium concept is probably the most important strategic choice in considering trade with increasing returns. The standard competitive equilibrium is not appropriate and the papers in this section make very different choices about how to replace it.

Paul Krugman's paper “Increasing returns, monopolistic competition and international trade” is one of the earliest and clearest in a literature that uses a variety of models of imperfect competition to provide the equilibrium concept. In these models, each of the trading countries has an imperfectly competitive domestic economy, and trade increases the scale of the market and the number of competitors. It is hard to derive clear-cut welfare or policy results in this framework: nothing equivalent to the classical theorems about gains from trade or factor-price equalization emerges (although see Wilfred Ethier (1982) for some results analogous to the classical results). Within a framework of imperfect competition, there is really little to distinguish international trade from national trade. Qualitatively they are the same, as are the economic issues at stake. At an intuitive level, the models are appealing. Their weaknesses are those of any models of imperfect competition, namely that we currently lack a really convincing model of imperfect competition in a world of increasing returns. We saw some models of this type in the

previous sections, and for specific sectors or phenomena such as network externalities, lock-in through adoption, or product differentiation, they seem insightful. But none of these models is general, although their application to international trade could lead to useful insights. A characteristic of Krugman's results, which also appears in the later papers by Graciela Chichilnisky and Geoffrey Heal and by Ralph Gomory, is the indeterminacy of the allocation of industries between countries with increasing returns. The model cannot determine fully which goods are produced in which countries. This is an intuitively appealing result – the intuition is set out geometrically in Chichilnisky and Heal – and one that generates insights into the politics of international trade. It tells us that there may be room for policy or for national strategy in resolving this analytical indeterminacy and obtaining a favorable outcome in the sense of an industry that has positive spillovers for the rest of the economy.

Graciela Chichilnisky's paper "Trade regimes and GATT: resource intensive vs. knowledge intensive growth" takes a radically different approach, one that preserves the competitive equilibrium concept. She assumes that increasing returns are external to the firm: that is, each firm on its own perceives that returns to scale are constant and costs independent of the level of output, yet when all firms expand together, productivity increases and unit costs drop. This is a familiar device for reconciling competition and increasing returns, one that dates back to Marshall. There are certainly some economic sectors in which increasing returns are external. A widely cited one is computer technology, where a regional concentration of expertise seems to enhance productivity, as for example in the case of Silicon Valley. Chichilnisky is specifically interested in these cases. She focuses not on the traditional questions about gains from trade and factor prices, but rather on issues relating to the formation of free trade areas. A traditional concern with customs unions is that the market power accruing to them by virtue of their size gives an incentive to levy optimal tariffs rather than move towards free trade. Given the proliferation of free trade areas, this could pose a threat to progress toward multilateral free trade. Chichilnisky shows that economies of scale can mitigate this tendency: there is an offsetting incentive to have access to a yet bigger market to reap the advantages of scale economies. By working with external economies of scale, she formalizes this in a competitive model: however, the intuition seems robust enough to carry through with internal economies of scale, although attempting to do this would raise the difficult question of the choice of an appropriate equilibrium concept.

In "Large-scale technologies and patterns of trade" Graciela Chichilnisky and Geoffrey Heal assume increasing returns to be internal to the firm and explore a range of different equilibrium concepts. They suggest several conclusions. One is that there may be no prices at which international trade flows will balance, so that if payments are to be in balance then countervailing capital flows will be needed. Another is that some countries may gain and some may lose from trade and that which gain and which lose is arbitrary, and depends on strategic issues such as being the first to produce in a particular industry. Finally, a regulatory approach based on pricing at marginal cost can lead to balanced trade, but not necessarily to welfare gains from all countries. All of these conclusions are based on the analysis of general equilibrium models with increasing returns, as in sections 1.2 and 2.

In “A Ricardo model with economies of scale,” Ralph Gomory presents an interesting and innovative model. The assumptions and framework are similar to those in Chichilnisky and Heal, but the method of analysis is quite distinct. Again we have increasing returns internal to the firm, and the conclusion that which industry is assigned to which country is arbitrary. What is novel about Gomory’s analysis is that rather than study a particular equilibrium, he characterizes the set of all equilibria that could possibly emerge from trade, and makes statements about the welfare implications of different regions of this set.

8. Conclusions

What overall assessment emerges from this review of increasing returns? One clear conclusion is that there are many important areas of economics in which the recognition of increasing returns makes a big difference, and changes the established wisdom significantly. Modifying our models to incorporate increasing returns is not a matter of making marginal changes: it can be of the essence. Secondly, we have not yet reached the point of diminishing returns in the study of increasing returns: there is a long way to go, and the results of the work yet to be completed will be interesting.

Important conclusions are emerging on the key issue of attaining efficiency in the face of increasing returns in production. The ability to pierce the veil of anonymity normally associated with the market, to identify agents and then to price discriminate, seems necessary for reaching efficient allocations. Further, marginal cost pricing is not a robust recommendation. These points imply little short of a revolution in regulatory thinking, which has lagged behind the development of the underlying economic theory.

The increasing returns linked to a large user base are a recent phenomenon, and still pose a serious challenge. Although clearly a form of increasing returns, they are different analytically from the classical increasing returns linked to the scale of a production process, and are not amenable to the same treatment. They have direct relevance to topical regulatory issues concerning the liberalization of telecommunications and the choice of standards in the computer hardware and software markets. Such high profile issues as public policies towards Microsoft are at stake. Closely related are the increasing returns linked to the use of information. In a world in which the collection and use of information is becoming more common, and more central to the mission of many enterprises, this source of increasing returns must become more important.

Clearly increasing returns has a key role in economic dynamics, although we have barely begun to explore this. Romer’s endogenous growth model posits increasing returns to the use of knowledge as a central assumption. This is intuitively reasonable, yet there is little formal analysis of the production, acquisition and use of knowledge or its role in production processes. In international trade we have likewise barely started to understand the impact of increasing returns. Indeterminacy of the allocation of industries between countries, and inability to reach equilibrium patterns of trade flows, seem likely

outcomes. A deeper understanding must await convincing models of competition with increasing returns.

References

Baumol, William J. and David Bradford, "Optimal departures from marginal cost pricing" American Economic Review (June 1970) 60 (3): 265-283.

Brown, Donald and Geoffrey Heal, "Two-part tariffs, marginal cost pricing and increasing returns in a general equilibrium model." *Journal of Public Economics* 13 (1980): 25-49.

Ethier, Wilfred J., "National and international returns to scale in the modern theory of international trade." *American Economic Review* June 1982 72 (3): 389-405.

Hayek, F.A. Von, "The use of knowledge in society." *American Economic Review* 1945 35 (4).

Heal, Geoffrey, *The Theory of Economic Planning*. Amsterdam: North Holland Publishing Company, 1973.

Hotelling, Harold, "The general welfare in relation to problems of taxation and of railway and utility rates." *Econometrica* (1938) 6: 242-269.

Lange, Oscar, "On the economic theory of socialism." *Review of Economic Studies* 1936-7 4 (1, 2)

Mill, John Stuart, *Principles of Political Economy*. London: J. W. Parket, 1848.

Quinzii, Martine, *Increasing Returns and Economic Efficiency*. Oxford: Oxford University Press, 1992.

Smith, Adam, *An Inquirey into the Nature and Causes of the Wealth of Nations*. Chicago: University of Chicago Press, 1977.

Readings

1. Resource Allocation and Welfare

1.1 Alternatives to the Market

Kenneth Arrow and Leonid Hurwicz, "Decentralization and computation in resource allocation." In Ralph Pfouts ed. *Essays in Economics and Econometrics in Honor of Harold Hotelling*, 34-104. Chapel Hill NC: University of North Carolina Press, 1960.

Geoffrey Heal, "Planing without prices." *Review of Economic Studies* (1969) 36 (3): 347-362.

Geoffrey Heal, "Planning, prices and increasing returns." *Review of Economic Studies* (1971) 38 (3): 281-294.

Hajime Hori, "The structure of equilibrium points in Heal's process." *Review of Economic Studies* (1975) 42 (3): 457-467.

Masaiko Aoki, "An investment planning process of an economy with increasing returns." *Review of Economic Studies* (1971) 38 (3): 273-280.

Jaques Cremer, "A quantity-quantity algorithm for planning in an economy with increasing returns." *Econometrica* (1977) 45 (6): 1303-1328.

Geoffrey Heal, "Equivalence of saddle points and optima for non-convex programs." *Advances in Applied Mathematics* (1984) 5: 398-415.

1.2 General Equilibrium

Roger Guesnerie, "Pareto optimality in a non-convex economy." *Econometrica* (1975) 43 (1): 1-30.

Donald Brown and Geoffrey Heal, "Equity, efficiency and increasing returns." *Review of Economic Studies*, (1979): 571-586.

Paulina Beato and Andreu Mas-Collel, "On marginal cost pricing with given tax-subsidy rules." *Journal of Economic theory*, (1985) 37: 356-365.

Donald Brown, Geoffrey Heal, Ali Khan and Rajiv Vohra, "On a general existence theorem for marginal cost pricing equilibria." *Journal of Economic Theory* (April 1986) 38 (2): 371-379.

Donald Brown and Geoffrey Heal, "Marginal vs. average cost pricing in the presence of a public monopoly." *American Economic Review* (May 1983) Papers and Proceedings of the ninety-fifth annual Meeting of the American Economic Association 73 (2): 189-193.

Aaron S. Edlin, Mario Epelbaum, and Walter P. Heller, “Is perfect price discrimination really efficient? Welfare and existence in general equilibrium." *Econometrica*, in press.

Jaques Dreze and Pierre Dehez, “Competitive equilibria with quantity-taking producers and increasing returns to scale." *Journal of Mathematical Economics* (1988), 17: 209-230.

2 Micro Foundations of Macroeconomics

Geoffrey Heal, “Stable disequilibrium prices." *Economics Letters* (1992) 20: 423-428.

Graciela Chichilnisky, Geoffrey Heal and Yun Lin, “Chaotic price dynamics, increasing returns and the Phillips curve." *Journal of Economic Behavior and Organization* (1995) 27: 279-291.

Martin Weitzman, “Increasing returns and the foundations of unemployment theory." *Economic Journal*, (December 1982) 92: 787-804.

Geoffrey Heal, “Rational rationing and increasing returns." *Economics Letters* (1981) 8 (1): 19-27.

Geoffrey Heal, “Macrodynamics and returns to scale." *Economic Journal* (1986) 96: 191-198.

3 Imperfect Competition

Avinash K. Dixit and Joseph E. Stiglitz, “Monopolistic competition and optimum product diversity." *American Economic Review* (June 1977) 67 (3): 297-308.

Geoffrey Heal, “Spatial structure in the retail trade: a study in product differentiation with increasing returns." *Bell Journal of Economics* (1980) 11 (2): 545-583.

Michael Spence, “Product selection, fixed costs and monopolistic competition.*" Review of Economic Studies* (June 1976) 43 (2) 134: 217-235.

4 Increasing Returns, Information and Information Technology.

Paul David, “Clio and the economics of QWERTY." *American Economic Review* (May 1985) 75 (2): 332-337.

Geoffrey Heal, “The economics of networks," in G. Chichilnisky ed. *The Challenge of Economics Theory*. New York: Cambridge University Press, in press.

Brian Arthur, “Competing technologies, increasing returns, and lock-in by historical events." *Economic Journal* (March 1989) 99 (394): 116-131.

Robert Wilson, “Informational economies of scale.” *Bell Journal of Economics and Management Science*, Spring 1975: 184-195.

Roy Radner and Joseph Stiglitz, “A nonconcavity in the value of information.” Chapter 3 (33-52) of *Bayesian Models in Economic Theory*, edited by M. Boyer and R.E. Kihlstrom, Elsevier Science Publishers B.V., 1984.

5 Economic Growth

Allyn Young, “Increasing returns and economic progress." *Economic Journal* (1928) 38: 527-542.

Kenneth Arrow, “The economic implications of learning by doing.'' *Review of Economic Studies* (1962) 29 (3): 155-173.

Avinash Dixit, James Mirrlees and Nicholas Stern, “Optimum savings with economies of scale." *Review of Economic Studies* (1975) 42 (3): 303-326.

Graciela Chichilnisky, “Existence of characterization of optimal growth paths including models with non-convexities in utilities and technologies." *Review of Economic Studies* (1981) 48 (1): 51-61.

Paul Romer, “Increasing returns and long-run growth." *Journal of Political Economy* (1986) 94 (5) 1002-1037.

6 International Trade

Paul Krugman, “Increasing returns, monopolistic competition and international trade." *Journal of International Economics* (November 1979) 9 (4) 469-479.

Graciela Chichilnisky, “Trade regimes and GATT: resource intensive vs. knowledge intensive growth." *Economic Systems merged with Journal of International and Comparative Economics* (1996) 20: 147-181.

Graciela Chichilnisky and Geoffrey Heal. “Large-scale technologies and patterns of trade.” Chapter 10 (122-135) of Graciela Chichilnisky and Geoffrey Heal, *The Evolving International Economy,* Cambridge University Press, 1986.

Ralph Gomory, “A Ricardo model with economies of scale.” *Journal of Economic Theory* Vol. 62 No. 2 (April 1994): 394-419.