

Almost efficient innovation by pricing ideas

by

Danny Quah*
LSE Economics Department
June 2002

* I thank the Andrew Mellon Foundation for financial support, and Lucien Foldes, Louise Keely, Chad Jones, Andreu Mas-Colell, Rachel Ngai, and Paul Romer for helpful suggestions.

Almost efficient innovation by pricing ideas

by

Danny Quah
LSE Economics Department
June 2002

ABSTRACT

Innovation is the creation of intellectual assets; intellectual assets are distinguished by nonconvexities. But different nonconvexities affect outcomes differently. Provided knowledge is finitely expansible, non-rivalry causes no difficulties for the existence and ex post efficiency of competitive equilibrium. However, the minimum-size indivisibility in creating intellectual assets renders competitive equilibrium socially inefficient in general, even though existence is maintained. The inefficiencies have the size of Dupuit triangles. Innovation patterns are invariant within a range of legal structures regulating intellectual asset dissemination.

Keywords: content, Dupuit triangle, finite expansibility, indivisibility intellectual asset, intellectual property rights, knowledge, lattice, nonconvexity, nonrivalry, R&D

JEL Classification: D41, D61, D90, O31

Communications to: Danny Quah, Economics Department, LSE, Houghton Street, London WC2A 2AE

Tel: +44/0 20 7955.7535, Email: dquah@econ.lse.ac.uk
(URL) <http://econ.lse.ac.uk/staff/dquah/>

Almost efficient innovation by pricing ideas
 by
 Danny Quah*
 LSE Economics Department
 June 2002

1 Introduction

Innovation is the creation of new ideas. Absent other rewards, innovation occurs when the price of a new idea exceeds its instantiation cost. This paper studies whether markets create ideas efficiently.

It has long been a commonplace that ideas or knowledge are nonrival. Having the nature of recipes or blueprints, an idea's use by one agent does not detract from its use by another. Thus, transacting in an idea only makes copies of the idea; no fixed quantity of the idea changes hands, unlike for ordinary commodities.

Economists have mostly followed Arrow (1962), Shell (1966), and Nordhaus (1969) in concluding this nonrivalry implies endemic market failure. Once an idea has been instantiated, extra copies of it can be made without expending further resources. Therefore, with competitive markets, the price of an idea in circulation is zero. But zero prices *ex post* cannot pay for positive *ex ante* instantiation costs. Thus, the reasoning goes, markets fail: equilibrium fails to exist, and innovation is insufficiently incentivized.

Intellectual property rights (IPRs) disrupt this chain of reasoning by restricting the circulation of ideas. In so doing, IPRs produce monopoly rents that can then incentivize creativity.

Boldrin and Levine (2002) broke from traditional reasoning and suggested ideas are not literally nonrival, but only nearly so. In their argument, copying rates are finite, so that even if an infinite number of copies can be made, that can happen only eventually. They prove that

* I thank the Andrew Mellon Foundation for financial support, and Lucien Foldes, Louise Keely, Chad Jones, Andreu Mas-Colell, Rachel Ngai, and Paul Romer for helpful suggestions.

with this small modification perfectly competitive markets price intellectual assets positively. Remarkably, prices remain bounded away from zero even when copying rates get arbitrarily large. Positive innovation can occur guided only by whether the market price of an idea dominates its instantiation cost.¹ What matters is having property rights only in the sense of Arrow-Debreu general equilibrium, not in the sense of IPRs.

Quah (2002a) relaxed further the restrictions on use and dissemination in the Boldrin-Levine model, and confirmed their conclusions under a wider range of dissemination schemes and technologies than originally considered. The significant exception is when economic actions can be undertaken in progressively finer timeslices: Then, when copying rates get sufficiently large, competitive markets converge to the Arrow-Shell-Nordhaus outcome.

Abstracting from this special case, both Boldrin and Levine (2002) and Quah (2002a) observed that in equilibrium the price on an idea equals the derivative of society's value function with respect to that idea's quantity (i.e., its number of copies). Given that some agent finds an idea worth instantiating in equilibrium, she instantiates the idea to its socially optimal quantity.

In other words, competitive markets create ideas efficiently, just as they do ordinary goods. However, an alternative interpretation is possible. Suppose entrepreneurs consider a range of ideas. Each idea can be instantiated only to a given, fixed quantity. The motivation for this is apparent: The first instance of an idea is, by definition, unique—there is no such thing as a fractional first copy of an idea. Instantiation quantity is not a variable that can be freely chosen below some minimum level. Does the competitive equilibrium studied in Boldrin and Levine (2002) and Quah (2002a) remain socially efficient?

This paper shows that in general the answer is No. Although competitive equilibrium still exists and still satisfies the value-derivative condition, it generates too little innovation relative to the social effi-

¹ Hellwig and Irmen (2001) provide a different approach by assuming decreasing returns to scale in production. This generates positive equilibrium profits that can then similarly incentivize innovation.

cient outcome. The reasoning differs from the standard one described above. Here, the conclusion turns on a concave value function always having marginal value below average. The economics is that of new goods and Dupuit triangles (Romer, 1994).

Boldrin and Levine (2002) too noted the minimum-size indivisibility of ideas to be the key remaining obstacle to competitive markets efficiently determining innovation. They argued, however, that both economic growth and entrepreneurs forming sufficiently large coalitions mitigate the importance of this second nonconvexity, and concluded “indivisibility need not pose a substantive problem” (Boldrin and Levine, 2002, p. 2). The results below suggest that the natural proposals along those lines will not restore efficiency.

The remainder of this paper is organized as follows. Section 2 describes some related literature. Section 3 presents the model: It characterizes social efficiency by merging the discussions from Boldrin and Levine (2002) and Quah (2002a,b). Section 4 analyzes equilibrium, comparing competitive and socially efficient outcomes. Here, innovation is a subset of an underlying abstract set of ideas, so that the collection of possible innovations forms a lattice. Equilibrium locates a fixed point of a mapping taking that lattice to itself. Fig. 1 and Theorems 4.10 and 4.12 constitute the heart of the paper, describing how minimum-size indivisibility intrudes between social efficiency and pricing ideas positively. Section 5 quantifies, in numerical examples, the social efficiency loss from indivisibility. Section 6 concludes. All proofs and auxiliary discussion are collected in section 7.

Because the paper takes up technical questions—discreteness, existence, optimality, innovation patterns as a subset of an abstract space of ideas—the presentation that follows is correspondingly abstract and formal. As a result, some economic content might be less transparent, so I provide here a roadmap to the paper’s findings. (Those who find nontechnical overviews not to their taste should look away now and go directly to either sections 2 or 3.)

Consider Fig. 1 in section 4. For simplicity, take the vertical axis there to be utility, and assume direct consumption of good s_{m0} measured on the horizontal axis. Assume, as Fig. 1 denotes, that utility begins with value 0 at the origin and is strictly concave in s_{m0} .

Also draw a ray emanating from the origin with positive slope equal to a *constant* unit cost of producing s_{m0} . (The reader might wish to relabel the different objects appearing in Fig. 1 to correspond to this description.)

In competitive markets the price that consumers will pay for the good s_{m0} equals the slope of the utility function. When s_{m0} can be continuously varied, the market clears where the slope of the utility function just equals the unit cost of production, i.e., at $s_{m0} = s_m^\dagger$ in Fig. 1. This also maximizes social surplus, measured by the vertical height between the utility function and total production costs.

Now, however, (and this is key to the paper) interchange what is fixed and what is variable: Suppose instead that s_m^\dagger on the horizontal axis is exogenously given, across a range of possible m ’s, and suppose the corresponding total cost Ψ_m is drawn from a range of possibilities, varying across m ’s. Redraw the ray from the origin so it goes through height Ψ_m at $s_{m0} = s_m^\dagger$. In general, this ray will no longer have slope equal to marginal utility at s_m^\dagger . All m ’s with Ψ_m from the value of the utility function down to where the ray with slope equal to marginal utility intersects $s_{m0} = s_m^\dagger$ will not be produced, even though they engender positive social surplus. That that range is nondegenerate follows from utility strictly concave; it is larger, the more concave is the utility function.

While the above is the central insight of this paper, many loose ends need to be cleared up. Why don’t entrepreneurs in markets see through the reasoning above and remove the social inefficiency; what market constraints do entrepreneurs work under? (Defns. 3.4 and 3.5, which distinguish ex post and ex ante efficiency; Prop. 3.6; Theorems 3.9, 3.11, 4.4, 4.9, 4.10, and 4.12.) In the model, one-time production costs are defined only at a single point of instantiation—depending on their domain of definition, those costs can be viewed as either linear or nonlinear. What then is the ray from the origin in Fig. 1; how does that relate to alternative uses for the initial endowment of the economy? (Assumptions 3.10, 4.6, and 4.8.) What does it mean to say “looking across all possible (subsets of) ideas m ”, and why does the social inefficiency argument above not apply to the usual

continuously-variable case? (Section 3.4; and Theorems 4.9, 4.10, and 4.12.) Ideas produce dynamic, not static, payoffs; consumers don't consume ideas directly; why is the payoff function in Fig. 1 smooth and strictly concave? (Technologies 3 and 4; Prop. 3.6; Assumption 3.7; and Theorem 3.9.) Fine, the payoff function in Fig. 1 is a value function, not utility, defined on s_{m0} —but why is it 0 at the origin? If it weren't 0 there but something else—more typical—then a convexifying lottery could overcome the inefficiency effects described above. (Theorems 3.11 and 4.10.) What formalizes the special features that makes m 's ideas, not just ordinary capital goods? (Section 3.3; Assumption 3.2; technologies 3 and 4; and Defns. 4.1 and 4.2.) How do different legal restrictions on the trade of ideas interact with technologies and potentially affect the reasoning in Fig. 1? (Defns. 4.1 and 4.2; Theorem 4.4.)

2 Related literature

Arrow (1962) pointed to indivisibility, inappropriability, and uncertainty as the three characteristics essential to innovation. Appropriability was considered in both Boldrin and Levine (2002) and Quah (2002a)—under their assumptions, nonrivalry raises no difficulty for appropriability. Both papers ruled out uncertainty in treating only deterministic economies.

Inappropriability focuses attention on the inability of competitive markets to incentivize the creation and dissemination of knowledge, and suggested institutions to protect IPRs. Although not always explicitly recognized, IPRs and the monopolies they sanction, or some similar device that confounds the free workings of markets, are inescapable everywhere knowledge matters—in economic geography, economic growth, industrial organization, technical progress,

and trade.²

The same market failure manifests also for other goods whose properties resemble knowledge. Such goods include computer software, console games, video content, music recordings, and Internet-distributed digital media—goods that scientists and engineers would not ordinarily take to be knowledge. Nevertheless, it is economically correct to call all these *ideas*, *intellectual assets*, or *content*. The term *intellectual property* is less apposite, as it presupposes specific social institutions, not intrinsic to the commodities themselves. Indeed, part of the goal of this paper is to see how IPR systems might be irrelevant or even harmful.

Boldrin and Levine (2002) followed the terminological convention that nonrivalry is the critical assumption from which their work departed. A related concept, due to David (1992), is that of *infinite expansibility*. A good is infinitely expansible when, after instantiation, an arbitrary number of copies can be instantaneously produced at zero marginal cost.

Quah (2002a) pointed out that where Boldrin and Levine (2002) depart importantly from traditional reasoning is not in nonrivalry but in expansibility. While close and previously not distinguished one from the other, nonrivalry and infinite expansibility turn out to have different implications for competitive equilibrium. Quah (2002a) showed that a dynamic representative-agent, single-commodity economy where the intellectual asset is nonrival but only finitely expansible has competitive equilibrium with asset prices strictly positive. Quah (2002a) showed further that keeping the same rival and finitely expansible technology but allowing agents to behave with yet greater freedom—disseminating ideas in ways that Boldrin and Levine (2002) had ruled out—continues to generate the Boldrin-Levine allocation, with no change in the price path of the intellectual asset.

Infinite expansibility and minimum-size indivisibility are distinct

² See, e.g., Dasgupta and Stiglitz (1980), Helpman (1993), Jaffee, Trajtenberg and Henderson (1993), Jones and Williams (2000), Keely (2001), O'Donoghue and Zweimüller (1998), Quah (2002b), Romer (1990), and many others.

nonconvexities. Boldrin and Levine (2002) suggested that it is the second nonconvexity, indivisibility rather than infinite expansibility, that is critical. They then observed, however, that “indivisibility need not pose a substantive problem” for two reasons. First, with economic growth, every minimum-size indivisibility eventually becomes insignificant; second, clever entrepreneurs can form sufficiently large coalitions or lotteries to overcome any minimum-size indivisibility. In other words, Boldrin and Levine claimed that the indivisibility nonconvexity becomes irrelevant in a large economy.

With economic growth, the indivisibility on a given idea certainly becomes relatively unimportant. By the late 18th-century, no industrialist would have considered making fire or the wheel to be any great obstacle before they could get on with their real work, even had they had to re-implement either from scratch. By the early 21st century, most of humanity’s important scientific ideas are given to college students in routine weekly exercises.

However, with economic growth, so too does the frontier of ideas change. The latest ideas become correspondingly intricate. Implementing an innovative engineering, pharmaceutical, or software product remains today an immense undertaking, even compared to modern economies’ considerable resources. Moreover, the work in such a project typically cannot be easily divided up across teams of researchers. In software engineering, a project taking 10 mythical people-months could well take one person 10 months to complete. But giving the project to two engineers might successfully conclude the project no sooner.³

This discussion suggests that, intuitively appealing though they might be, the empirical significance of the growth and coalition effects remains to be verified. If these effects work, they do so by making the indivisibility sufficiently small relative to the total initial endowment.

This paper shows, however, that that is not what is critical. In the model below, the indivisibility-induced social inefficiency remains even when idea-instantiation costs become trivially small relative to

³ In one view, Merton multiples (Dasgupta and Maskin, 1987; Merton, 1973) might be partly explained by this type of indivisibility.

the size of the economy. To anticipate the discussion in section 4, this is because what is critical is not the magnitude of the minimum-size indivisibility. Instead, it is the slope of the value function there, relative to its average. In many circumstances, that ratio is invariant to the overall size of the economy. Thus, neither economic growth nor coalition-formation will remove the social inefficiency.

3 The model

The model relies on analyzing value functions. It collapses consumers’ and firms’ intertemporal decisions, from period 1 on, into an initial period representation.

3.1 Commodity space

A *consumption bundle* c is a nonnegative doubly-indexed array of numbers

$$c = \{ c_{mt} \in \mathbb{R}_+ : m \in \mathbb{N}; t = 0, 1, 2, \dots \},$$

where m indexes *goods*, t indexes discrete time, and the set \mathbb{N} is the collection of non-negative integers.

The set $\mathbb{N} \setminus \{0\}$ denotes all possible intellectual assets or ideas, and individual elements in it need to be instantiated (section 3.4 below). Good 0 differs from all others in \mathbb{N} and is an ordinary commodity: It is rival; it carries zero instantiation cost; it always exists. An initial endowment of stock from which good $m = 0$ is produced will, in Assumption 3.10 below, provide the resources from which all other goods are instantiated at time $t = 0$. To emphasize the subsequent focus on $\mathbb{N} \setminus \{0\}$, hereafter denote this set by symbol \mathbb{M} . The notation is additionally useful because the special structure in the collection of positive integers—additivity, a total ordering, and so on—are irrelevant and mildly distracting when we will be concerned only with instantiation over an abstract set of ideas \mathbb{M} .

The *commodity space* \mathbb{C} is the collection of all consumption bundles c .

3.2 Consumers

The infinitely-lived representative agent has preferences at time 0 defined on \mathcal{C} :

$$\forall c \in \mathcal{C} : U(c) = \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in \mathbb{N}} U_m(c_{mt}) \right], \quad \beta \text{ in } (0, 1). \quad (1)$$

Discounting is geometric at rate given by discount factor β . That preferences (1) show additive separability across goods and time will make transparent below which m 's get instantiated.

We will restrict U_m as follows for the principal discussion.

Assumption 3.1 *Function $U_m : \mathbb{R}_+ \rightarrow \mathbb{R}$ has $U_m(0) = 0$, and is bounded, strictly increasing, strictly concave, and continuously differentiable, with $U'_m(c_m) \rightarrow \infty$ when $c_m \rightarrow 0$.* ■

Since we will study a growth model with technology eventually constant returns to scale, that Assumption 3.1 requires U_m to be bounded might seem unappealing, especially as utility functions of form

$$U_m(c_m) = \frac{c_m^{1-\sigma_m^{-1}} - 1}{1 - \sigma_m^{-1}}, \quad (2)$$

with constant $\sigma_m \leq 1$ are often used in such analyses. (Recall that when preferences (1) have U_m obeying (2) then the intertemporal elasticity of substitution is constant and equals σ_m .) However, as we will need to assess situations where U_m has argument $c_m = 0$, we will require at least boundedness from below in U_m , thereby ruling out anyway utility functions (2) when $\sigma_m \leq 1$. Nevertheless, preferences (2) will be useful for intuition in explicit examples below.

3.3 Technology

Each c_{mt} (including $m = 0$) is the flow of consumption services generated from an associated durable stock s_{mt} by $c_{mt} \in [0, s_{mt}]$. It is natural to refer to c_m as good m and s_m as asset m . While s_m is durable, c_m is not. Assume free disposal in both s_m and c_m .

Unlike for $m = 0$, not all assets $m \in \mathbb{M}$ will exist in positive quantity in equilibrium. Assume that, at time 0, the first instance of asset $m \in \mathbb{M}$ is costly to create. Should instantiation occur for such m , the fixed finite positive amount s_m^\dagger is produced. (If it aids intuition, the reader can set s_m^\dagger to its canonical value of 1. However, no insight is refined with this restriction whereas being explicit about s_m^\dagger is useful below.) Such instantiation—the creation of new ideas or new intellectual assets—is what this paper takes to be innovation. Call m 's with $s_{m0} = 0$ *inactive* at 0, or just *inactive*. Instantiation can occur only at time $t = 0$.

Unlike $m \in \mathbb{M}$, asset 0 always exists in positive quantity \mathcal{E} prior to time 0, i.e., $m = 0$ is always active. The economy-wide endowment \mathcal{E} has two rival uses: it can be used to instantiate intellectual assets $m \in \mathbb{M}$ or it can generate consumption c_{0t} . Instantiation of $m \in \mathbb{M}$ destroys positive Ψ_m units of \mathcal{E} . The implied rate of exchange

$$\psi_m \stackrel{\text{def}}{=} \Psi_m / s_m^\dagger \in (0, \infty), \quad m \in \mathbb{M},$$

is always well-defined and constant.

Innovation can be alternatively described: At time 0, a technology is available to convert \mathcal{E} into s_m , $m \in \mathbb{M}$, with input-output ratio $1/\psi_m$ and operating only at scale Ψ_m . What remains of the endowment becomes s_{00} . Call Ψ_m *instantiation cost* and, correspondingly, ψ_m *unit cost*. (Even though costs, unlike preferences and technology, aren't primitive to the Arrow-Debreu statement of the problem, this language allows greater brevity—more than saying “one over the technology input-output ratio” each time.) In the canonical case with $s_m^\dagger = 1$, we have $\psi_m = \Psi_m$, but again it will help to specify s_m^\dagger explicitly.

For reference below, summarize this in the following.

Assumption 3.2 *Asset $m \in \mathbb{M}$ costs $\Psi_m > 0$ to instantiate to the fixed quantity $s_{m0} = s_m^\dagger > 0$.* ■

From section 2, *some* version of nonrivalry or infinite expansibility must enter the discussion if \mathbb{M} is to be interpreted as ideas. To that end, assume dynamic or accumulation technologies after $t = 0$ that

vary with $m \in \mathbb{M} \cup \{0\}$. Take, from Boldrin and Levine (2002) and Quah (2002a), two possible descriptions for such accumulation technologies.

First is

$$s_{m,t+1} = \gamma_m s_{mt} + \omega_m c_{mt}, \quad \gamma_m > 1, \omega_m \geq 0. \quad (3)$$

In equation (3) if $c_{mt} = 0$, all s_{mt} is devoted to accumulation. Then, accumulation occurs at growth rate γ_m . On the other hand, putting the asset to consumption use, setting $c_{mt} > 0$, adds to growth in the asset, to a degree parameterized by ω_m .

A more typical dynamic technology would have $\omega_m = -\gamma_m$. But while specification (3) might appear unusual—having s_{mt} in use to generate c_{mt} improves its accumulation—Quah (2002a) describes how it usefully captures the notion of nonrivalry, i.e., consumption not detracting from asset use in accumulation. With positive ω_m , the more that consumers use the asset, the more dynamically productive does the asset become. Alternatively, technology (3) might be used to represent consumption network externalities, or it might capture how an item of knowledge has its bugs shaken out faster, the more that it gets applied. I will refer to technology (3) as *NR technology*.

When $\gamma_m + \omega_m < \infty$, equation (3) allows s_m to grow arbitrarily large only gradually, and thus remains consistent with the Boldrin and Levine (2002) suggestion that ideas cannot be instantaneously reproduced to satiate markets. Therefore, as emphasized in Quah (2002a), specification (3) provides a simple distinction between non-rivalry and infinite expansibility.

The second accumulation technology is

$$s_{m,t+1} = \gamma_m s_{mt} - (\gamma_m - \hat{\gamma}_m) c_{mt}, \quad \gamma_m > 1, \gamma_m - \hat{\gamma}_m > 0. \quad (4)$$

Here, as in equation (3), when the asset is not generating consumption services, it grows at rate γ_m . Now, however, setting $c_{mt} > 0$ reduces rather than increases the accumulation rate. Rewrite equation (4) as

$$s_{m,t+1} = (s_{mt} - c_{mt}) \gamma_m + c_{mt} \hat{\gamma}_m$$

to see the following interpretation: Copies of the asset left in-store grow at the high rate γ_m , whereas copies that have been rented out or used to provide consumption services grow only at the degraded rate $\hat{\gamma}_m < \gamma_m$. Rate $\hat{\gamma}_m$ can be zero; what matters is that it is dominated by γ_m .

Boldrin and Levine (2002) use equation (4) to capture how copies of an idea might grow arbitrarily numerous only gradually. As in specification (3), the larger is γ_m the closer does equation (4) describe infinite expansibility. Since, as far as I know, its application to intellectual assets originated with Boldrin and Levine (2002), I will refer to technology (4) as *BL technology*.

For asset $m = 0$, assume its stock evolves over time as (4) with $\gamma_0 - \hat{\gamma}_0 = 1$, so that it too has *BL technology*. This is the standard specification for durable goods and fits conveniently in the family of technologies (3)–(4) considered in this paper.

Together with the restriction on how consumption services relate to the underlying asset stock, the technologies in (3) and (4), for all $m \in \mathbb{N}$, then are special cases of

$$c_{mt} \in [0, s_{mt}] \quad (5)$$

$$s_{m,t+1} = g_m(s_{mt}, c_{mt}), \quad (6)$$

where accumulation function g_m is continuously differentiable in each argument, with partial derivatives $\partial g_m / \partial s_{mt} > 0$ and $\partial g_m / \partial c_{mt} \neq 0$. Everything that follows can be developed generally, using only analytical restrictions on functions g_m , rather than the specific parametric forms in (3) and (4). Treatments differ when the partial derivative $\partial g_m / \partial c_{mt}$ take different signs, exactly as when technology is either (3) or (4). This general nonlinear case (6), however, adds no further insight, so this paper will continue to assume instead only the explicit linear parametric forms.

Develop a single notation for equations (3) and (4) together with restriction (5) in the (transition) correspondence

$$s_{m,t+1} \in \Gamma_m(s_{mt}) \subset \mathbb{R}_+. \quad (7)$$

Free disposal allows setting the lower endpoint of $\Gamma_m(s_{mt})$ to zero. Summarize this construction in the following.

Assumption 3.3 Correspondence Γ_m maps \mathbb{R}_+ to subsets of \mathbb{R}_+ by

$$\forall s \in \mathbb{R}_+ : \quad \Gamma_m(s) = \begin{cases} [0, (\gamma_m + \omega_m)s] & \text{if (3);} \\ [0, \gamma_m s] & \text{if (4).} \end{cases}$$

For all $s \in \mathbb{R}_+$ the set $\Gamma_m(s)$ is nonempty and compact. Moreover, the correspondence Γ_m is continuous and monotone.⁴ Finally, Γ_m is convex in that for any $\lambda \in (0, 1)$ and any s_0, s_1 in \mathbb{R}_+ ,

$$\begin{aligned} & s'_0 \in \Gamma_m(s_0) \text{ and } s'_1 \in \Gamma_m(s_1) \\ \implies & \lambda s'_0 + [1 - \lambda]s'_1 \in \Gamma_m(\lambda s_0 + [1 - \lambda]s_1). \end{aligned}$$

Correspondence (7) makes explicit a restriction in the *NR* and *BL* technologies. Asset m accumulates independently of all other assets m' . Thus, the important issues of breadth and cumulation are not treated here.⁵ When $s_{m0} = 0$ the correspondence Γ_m in Assumption 3.3 is single-valued at 0. Therefore, whenever Assumption 3.3 holds, ideas inactive at 0 remain inactive thereafter.

3.4 Innovation

Recall $\mathbb{M} = \mathbb{N} \setminus \{0\}$ is the collection of all possible ideas. Define a *pattern of innovation* or just *innovation* to be a subset $M \subseteq \mathbb{M}$, that collects together all assets $m \neq 0$ instantiated at time 0, i.e.,

$$s_{m0} = \begin{cases} s_m^\dagger > 0 & \text{if } m \in M; \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Denote by \mathcal{M} the collection of all subsets of \mathbb{M} .

⁴ A correspondence is continuous when it is both upper and lower hemi-continuous. Correspondence Γ_m is monotone when $\Gamma_m(s_m) \subseteq \Gamma_m(s'_m)$ whenever $s_m \leq s'_m$. See, e.g., Stokey and Lucas (1989, Chs. 3–4).

⁵ Klemperer (1990) and Scotchmer (1991) provide critical results on these. See, however, section 3.4 and footnote 6 below.

That innovation $M \in \mathcal{M}$ is a subset of $\mathbb{N} \setminus \{0\}$ misleadingly suggests an overly-restrictive mathematical structure. Ideas don't add and aren't sequential. No idea necessarily dominates another—some ideas can be ranked relative to one another, but not all. Innovation M is neither more nor less than M' in general, but only when either $M \supset M'$ or $M \subset M'$.

More generally, therefore, let the collection of ideas \mathbb{M} be some abstract set, with \mathcal{M} the collection of all its subsets. Then, patterns of innovation constitute a complete lattice.⁶

3.5 Social efficiency

An *allocation* is a doubly-indexed pair

$$\{(c_{mt}, s_{mt}) : t = 0, 1, 2, \dots; m \in \mathbb{N}\}.$$

If $s \in \mathbb{R}_+$ and $M \in \mathcal{M}$, then call (s, M) *initializing*. Fix an initializing (s, M) : an allocation is *feasible given* (s, M) when it satisfies for all $t = 0, 1, 2, \dots$ and all m in M equations (5), (7), and (8) with $s_{00} = s$ and M given.

Definition 3.4 Fix an initializing (s, M) . Call a *feasible allocation* (s, M) -efficient if it maximizes preferences (1) over all feasible allocations given (s, M) . Let $W(s, M)$ denote the value in (1),

$$W(s, M) = \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in \mathbb{N}} U_m(c_{mt}) \right],$$

at an (s, M) -efficient allocation. ■

To emphasize, (s, M) -efficiency takes as given an initial condition $s_{00} = s$ and a pattern of innovation M , and says nothing about

⁶ Quah (2002b) had previously this formalism to study optimal patent durations in a dynamic representative-agent model. Structuring the space $(\mathbb{M}, \mathcal{M})$ in specific ways would allow modelling breadth and cumulation, but that is not taken up here.

endowment \mathcal{E} . One reason for studying Defn. 3.4 in an intermediate step is that it corresponds to the notion of efficiency used in Boldrin and Levine (2002) and Quah (2002a).

Given endowment \mathcal{E} call an initializing (s, M) *feasible* if

$$s + \sum_{m \in M} \Psi_m \leq \mathcal{E}. \quad (9)$$

Definition 3.5 *Innovation* $M \in \mathcal{M}$ is efficient given \mathcal{E} or socially efficient when together with $s \stackrel{\text{def}}{=} \mathcal{E} - \sum_{m \in M} \Psi_m \geq 0$ the feasible initializing (s, M) has

$$W(s, M) \geq W(s', M')$$

across all $(s', M') \in \mathbb{R}_+ \times \mathcal{M}$ satisfying equation (9). \blacksquare

Defn. 3.5 carries within it (s, M) -efficiency implicitly, from W 's definition in Defn. 3.4.

The idea in the following is that the formal analyses in Boldrin and Levine (2002) and Quah (2002a) treated efficiency in the sense of Defn. 3.4. By contrast, this paper considers that in Defn. 3.5. The two are typically indistinct. Below, we will see that the minimum-size indivisibility in innovation drives between them a wedge the size of a Dupuit triangle.

With preferences (1) and technology (5) and (7), studying the consequences of Defn. 3.4 involves, mostly, recasting the problem within the framework of Stokey and Lucas (1989, Ch. 4). However, the choice variable $(s, M) \in \mathbb{R}_+ \times \mathcal{M}$ in Defn. 3.5 contains a discrete second component, so that more work will be needed below in Theorem 3.11.

Proposition 3.6 *Suppose for each m in \mathbb{N} the mappings U_m and Γ_m satisfy Assumptions 3.1 and 3.3. Fix $s_{00} = s \geq 0$ and finite innovation $M \in \mathcal{M}$. Then (s, M) -efficiency has representation*

$$W(s, M) = V_0(s) + \sum_{m \in M} V_m(s_m^\dagger) + \sum_{m \in \mathbb{N} \setminus \{M \cup \{0\}\}} V_m(0) \quad (10)$$

with $V_m : \mathbb{R}_+ \rightarrow \mathbb{R}$, all $m \in \mathbb{N}$, bounded, continuous, strictly increasing, and strictly concave. \blacksquare

(All proofs are in section 7.) From the proof to Prop. 3.6 the representation components V_m are value functions satisfying a Bellman equation, namely functional equation (47) in section 7.

The assumptions in Prop. 3.6 imply $V_m(0)$ is 0. It is essential that this value be 0 rather than any positive or negative amount, as W would then be either undefined or degenerate. However, I will carry along $V_m(0)$ explicitly as a reminder of the economic actions being undertaken here.

Assumption 3.1 has U_m continuously differentiable, but the proof and the conclusion of Prop. 3.6 specify only continuity. Recall that continuous differentiability in U_m is used, principally, to verify continuous differentiability in V_m . But to invoke the continuous differentiability results in, e.g., Stokey and Lucas (1989), the optimum program in Defn. 3.4 must have its solution interior in the constraint set Γ_m . We will see below that an assumption sufficient for this is reasonable when technology is *BL* equation (4). However, when instead technology is *NR* equation (3), because of nonrivalry, no sensible assumptions will guarantee an interior solution then. The optimum always sets $c_{mt} = s_{mt}$, locating $s_{m,t+1}$ on the upper boundary of $\Gamma_m(s_{mt})$. We then need an alternative approach.

Assumption 3.7 *If m has NR technology (3), let γ_m , ω_m , β , and U_m jointly satisfy:*

$$\lim_{t \rightarrow \infty} \sum_{j=0}^{\infty} ([\gamma_m + \omega_m]\beta)^{t+j} \times U'_m([\gamma_m + \omega_m]^{t+j}) = 0. \quad (11)$$

If, conversely, m has BL technology (4), let γ_m , $\hat{\gamma}_m$, β , and U_m jointly satisfy

$$\gamma_m > \beta^{-1} \times \sup_{s>0} \frac{U'_m(s)}{U'_m(\hat{\gamma}_m s)}. \quad (12)$$

\blacksquare

(If the general nonlinear accumulation equation (6) were used in Assumption 3.7, rather than equations (3) and (4), the corresponding restrictions (11) and (12) would be more obscure to state and interpret.)

To understand equation (11) in Assumption 3.7, consider the sequence

$$([\gamma_m + \omega_m]\beta)^t U'_m([\gamma_m + \omega_m]^t), \quad t = 0, 1, 2, \dots$$

Each term is non-negative. If $\gamma_m + \omega_m > 1$ dominates β^{-1} , the sequence above converges to zero if and only if U'_m declines sufficiently rapidly. Equation (11) strengthens this requirement that the sequence converges to zero by requiring instead that the tail sums in the sequence do so. It is a standard property of infinite series that tail sum convergence to zero is equivalent to summability of the sequence,

$$\sum_{t=0}^{\infty} ([\gamma_m + \omega_m]\beta)^t \times U'_m([\gamma_m + \omega_m]^t) < \infty.$$

Since technology (3) always implies a boundary solution in Prop. 3.6, we verify continuous differentiability of the value function by obtaining that property directly on the supremum function in problem (46) in the Proof of Prop. 3.6 (section 7 below). Equation (11) guarantees convergence of different infinite series needed.

Inequality (12), on the other hand, bounds γ_m from below. It says that society must face a sufficiently high reward to postponing consumption for *some* resources to be set aside in accumulation.

Relations (11) and (12) in Assumption 3.7 are implied by simple restrictions when preferences have constant intertemporal elasticity, as in specification (2) of section 3.2. Since, as already observed, such preferences violate Assumption 3.1, discussion is placed in the technical appendix, and the conclusion stated only as an aside in the following.

Example 3.8 Suppose preferences (1) have U_m taking the form (2). Then (a) restriction (11) holds if $\sigma_m \leq 1$, and (b) inequality (12) holds if and only if $\gamma_m > \beta^{-1} \times (\hat{\gamma}_m)^{1/\sigma_m}$. ■

Notice in Example 3.8, Assumption 3.7 holds from restrictions only on preferences for technology (3) while restrictions on both preference and technology parameters are needed for technology (4).

From the discussion of Example 3.8 in section 7, restriction $\sigma_m \leq 1$ is sufficient but not necessary for convergence to zero in (11). Thus, log utility, $\sigma_m = 1$, lies strictly in the interior of the set of preferences allowing equation (11). Moreover, in some instances, the ratio $U'_m(s_m)/U'_m(\hat{\gamma}_m s_m)$ will have its supremum at

$$\lim_{s_m \rightarrow \infty} \frac{U'_m(s_m)}{U'_m(\hat{\gamma}_m s_m)},$$

whereupon preferences need have U_m only asymptotically approaching the constant elasticity form in equation (2).

Explicit characterization of (s, M) -efficient allocations and value function differentiability obtain as follows.

Theorem 3.9 Suppose for each m in \mathbb{N} the mappings U_m and Γ_m satisfy Assumptions 3.1 and 3.3. Fix $s_{00} = s > 0$ and finite pattern of innovation $M \in \mathcal{M}$, and suppose each $m \in M$ satisfies Assumption 3.7. Then (s, M) -efficiency has representation (10) with functions $V_m : \mathbb{R}_+ \rightarrow \mathbb{R}$ bounded, strictly increasing, strictly concave, and continuously differentiable everywhere in the interior of \mathbb{R}_+ . An (s, M) -efficient allocation $(c_{mt}^*, s_{m,t+1}^*)$, all $m \in M \cup \{0\}$, exists, and with the associated V_m satisfy:

(a) If m has NR technology then

$$c_{m0}^* = s_{m0}, \quad (13)$$

$$\begin{aligned} c_{mt}^* &= s_{mt}^* = (\gamma_m + \omega_m) s_{m,t-1}^* \\ &= (\gamma_m + \omega_m)^t s_{m0}, \quad t = 1, 2, 3, \dots, \end{aligned} \quad (14)$$

$$V_m'(s_{m0}) = \sum_{t=0}^{\infty} ([\gamma_m + \omega_m]\beta)^t \times U'_m([\gamma_m + \omega_m]^t s_{m0}). \quad (15)$$

(b) If m has BL technology then for all $t = 0, 1, 2, \dots$,

$$U'_m(c_{mt}^*) = \beta U'_m(c_{m,t+1}^*) \times \gamma_m, \quad (16)$$

$$s_{mt}^* = (\gamma_m - \widehat{\gamma}_m) \gamma_m^{-1} \sum_{j=0}^{\infty} \gamma_m^{-j} c_{m,t+j}^* \quad (17)$$

$$\text{with } s_{m0}^* = s_{m0},$$

$$V'_m(s_{m0}) = \frac{\gamma_m}{\gamma_m - \widehat{\gamma}_m} U'_m(c_{m0}^*). \quad (18)$$

Defn. 3.5 together with Theorem 3.9 (which contains Prop. 3.6) will allow the characterization of social efficiency in Theorem 3.11 below. However, just as in smooth problems first-order conditions are conveniently described when the solution is interior (even though boundary solutions are easily understandable if only at best inelegantly stated), similarly so for the problem here. The hypotheses in the Proposition below will guarantee considering only possibilities appropriately interior. To that end, it is useful to assume the following.

Assumption 3.10 Assume for $m = 0$ that Bellman equation (47) in the Proof of Prop. 3.6 in section 7 has a unique solution $V_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$ that is continuously differentiable. Assume that endowment $\mathcal{E} > \sum_{m \in \mathbb{M}} \Psi_m$ so that $s_{00} \stackrel{\text{def}}{=} \mathcal{E} - \sum_{m \in \mathbb{M}} \Psi_m > 0$ and that there exists a real-valued function $\delta(\mathcal{E}) \geq 0$ such that

$$\sup_{m \in \mathbb{M}} |V_0(s_{00}) + V'_0(s_{00}) \Psi_m - V_0(s_{00} + \Psi_m)| \leq \delta(\mathcal{E}),$$

$$\text{with } \delta(\mathcal{E}) \rightarrow 0 \text{ as } \mathcal{E} \rightarrow \infty. \quad (19)$$

Assumptions for Theorem 3.9 of course deliver continuous differentiability in all V_m , but Assumption 3.10 relies on only that for $m = 0$.

Assumption 3.10 says V_0 is effectively linear at s_{00} relative to the order of magnitude in Ψ_m . Relation (19) is satisfied if V_0 is exactly linear, for then $\delta(\mathcal{E})$ can be taken to be identically zero. It is also

satisfied if Ψ_m/s is uniformly small, i.e., that \mathcal{E} is sufficiently large relative Ψ_m .

Below, we will typically assume \mathcal{E} so large that $\delta(\mathcal{E})$ is infinitesimally small. That \mathcal{E} is taken to be large in the analysis usefully shows also that any subtleties arising from the minimum-size indivisibility come not from the economy (or any agent) simply having insufficient resources to circumvent a discrete nonconvexity, but instead reflects a deeper property to the economics in this problem.

Let \emptyset denote the empty set.

Theorem 3.11 Assume for each m in \mathbb{N} the mappings U_m and Γ_m satisfy Assumptions 3.1 and 3.3. Suppose Assumptions 3.7 and 3.10, and let endowment \mathcal{E} be fixed and sufficiently large so that in equation (19) the quantity $\delta(\mathcal{E})$ is infinitesimally small. For finite innovation $M^* \neq \emptyset$, define $s^* = \mathcal{E} - \sum_{m \in M^*} \Psi_m$. Then (s^*, M^*) is socially efficient if and only if

$$m \in M^* \iff V_m(s_m^\dagger) - V_m(0) \geq V'_0(s^*) \Psi_m, \quad (20)$$

for V_m in representation (10). ■

The inequality in (20) compares $V_m(s_m^\dagger) - V_m(0)$, the value gained from instantiating asset m , to $V'_0(s^*) \Psi_m$, the value lost from reducing the quantity of asset 0 at time 0 by m 's instantiation cost. Theorem 3.11 says that socially efficient innovation instantiates all those goods where this surplus is nonnegative.

4 Efficiency and inefficiency under perfect competition

To see when Defns. 3.4 and 3.5 imply equivalent outcomes and when they don't, we turn to competitive equilibrium under a range of alternative market and legal arrangements.

Prior to time $t = 0$ profit-seeking firms investigate whether to activate idea $m \in \mathbb{M}$. Activation entails paying the instantiation cost Ψ_m in return for actualizing the fixed quantity $s_m^\dagger > 0$.

Shares are financial claims on a firm. Adopt the convention that the number of shares equals the quantity of the asset that the firm

operates, i.e., s_{mt} at time t . Shares trade in competitive markets at price q_{mt} . Under perfect competition constant returns to scale implies that the pattern of ownership in the asset is irrelevant for production. So, hereafter, identify s_{mt} interchangeably with both the productive asset and with financial claims on the asset.

To activate idea m , the firm must find

$$\Psi_m \geq q_{m0} s_m^\dagger \iff \psi_m \geq q_{m0}. \quad (21)$$

However, since idea m is not yet active when a firm assesses relation (21), the firm needs to form expectations of the share price q_{m0} . This price can be interpreted as the IPO valuation on a share of the firm with idea m ready to be put to use.

The traditional Arrow-Nordhaus view is that competitive markets in trading copies of idea m drive q_{m0} to zero. Obviously then inequality (21) ends up activating no positive-cost ideas. Boldrin and Levine (2002) and Quah (2002a) provide conditions under which q_{m0} is positive. The question then becomes more nuanced: How does the set of m 's activated by competitive equilibrium q_{m0} 's in relation (21) compare to the set that Theorem 3.11 carves out? Answering this question is the goal of the remainder of this section.

Firms calculate q_{m0} as the Arrow-Debreu price on s_{m0} , conditional on competitive markets operating for all consumption service flows and claims on assets from period $t = 0$ onwards. To do this, firms hypothesize that a set of ideas $M \subseteq \mathbb{M}$ is active at time $t = 0$; given M , perfectly competitive markets imply a set of equilibrium prices including q_{m0} , for all $m \in M$. Those prices, in turn, imply a set of ideas $M' \subseteq \mathbb{M}$ that (21) would have activated. A rational expectations innovation equilibrium is a set $M^e \subseteq \mathbb{M}$ that is a fixed point of this induced mapping.⁷

An interesting feature uncovered below (partly also discussed in Quah (2002a)) is that different legal structures on how intellectual

⁷ Notice this procedure prices only assets m in the initially hypothesized set M ; all assets m outside M are implicitly assigned price 0 and never subsequently considered. We return to this in section 4.2 below.

assets can be used or traded all imply *identical* q_{m0} . Thus, equilibrium innovation is invariant to fine details on the dissemination of ideas.

4.1 Arrow-Debreu equilibrium conditional on M

At time $t = 0$ take endowment \mathcal{E} and innovation M exogenously given, implying $s_{00} = \mathcal{E} - \sum_{m \in M} \Psi_m > 0$. All prices are stated for time 0.

By constant returns to scale we can assume, without loss, that for every $m \in M$ a single competitive firm owns all of asset m . The firm rents consumption service flow c_{mt} at price p_{mt} . Shares on the firm—of which there are quantity s_{mt} to match the quantity of its durable asset—trade in competitive markets at price q_{mt} . The firm distributes all proceeds from its operations to its asset-holders, in proportion to the quantity of shares they own. The firm can also sell quantities of s_{mt} to competing firms. Since we have identified the number of shares of the firm with its holdings of durable asset, obviously, it is the same price q_{mt} at which the underlying asset trades. The firm takes all prices as given and selects consumption flows to maximize its value.

The representative consumer enjoys a flow of consumption services c_{mt} , at unit cost equal to the price p_{mt} , all $m \in M$. The consumer holds a portfolio comprising shares in different firms. He selects his portfolio and consumption service flows to maximize preferences, taking all prices as given.

The descriptions just given are not yet entirely precise—because I want to consider different legal structures restricting trade in ideas. Different structures will imply different formulas for firm value and different budget constraints for the consumer.

Notice that we assume the firm can sell its asset s_{mt} to other firms without restricting how those firms use s_{mt} . The property rights here are therefore just those in the Arrow-Debreu model, not intellectual property rights that only license such sales and restrict how purchasing firms can use or copy the s_{mt} that they have bought. New firms can, for instance, compete unrestrainedly with the incumbent. This is the assumption that Boldrin and Levine (2002) emphasize for how

they relaxed the restrictiveness in intellectual property rights.

Left unspecified thus far, however, is how consumers can use the service flow c_{mt} they have purchased. For traditional goods, e.g., services from physical capital, this is not an issue: the two c_{mt} and s_{mt} are completely distinct. This is how $m = 0$ will continue to be modelled below.

But what is interesting and peculiar about ideas (or software or digital music), $m \in \mathbb{M}$, is that their use and the underlying assets are not sharply distinguished. It is technically feasible for users to re jig their computer music players so that playing a piece of MP3-encoded music ends up simultaneously writing the digital bits to disk. In this stylization, sufficiently understanding an engineering blueprint or mathematical theorem to apply it means the user can re-create the blueprint or theorem.

Consider two legal structures for how ideas (goods $m \in \mathbb{M}$) can be traded:

Definition 4.1 *Ideas show unrestricted dissemination when a consumer, having rented the flow c_{mt} , is allowed to write it to durable medium and thereby create $s_{m,t+1}$ (in quantity determined by the accumulation technology) for subsequent use or resale, and thus compete with established firms. At the end of period t the consumer (a) has enjoyed $U_m(c_{mt})$ and (b) has durable asset accumulated from the flow c_{mt} over the time period.* ■

Definition 4.2 *Ideas show conventional dissemination when a consumer, having rented the flow c_{mt} , is prohibited from using it for anything but his own consumption. At the end of period t the consumer has only enjoyed $U_m(c_{mt})$, and has no durable asset left over.* ■

The second, Defn. 4.2, describes most current legal regimes. Under these, consumers attempting the actions in Defn. 4.1 would be guilty of illegal bootlegging. Certain *fair use* provisions in copyright law permit limited versions of the unrestricted dissemination actions in Defn. 4.1.

We will see below that q_{m0} 's are invariant to whether legal regimes allow consumers to behave according to Defns. 4.1 or 4.2. Recall Boldrin and Levine (2002) had weakened property rights on intellectual assets to those only in the Arrow-Debreu model. Our result says that property rights can be relaxed even further: Fair-use provisions, lax in the extreme, will not negatively affect innovation.

In what follows assume that each $m \in M$ can have its legal structure following either *unrestricted dissemination* (denoted u) or *conventional dissemination* (denoted v). By assumption from section 3.3 each $m \in M$ can also be one of two possible technologies, NR and BL . Partition ideas $m \in M$ into the four categories, combining technologies and legal structures, so that:

$$M = M_{NRu} \cup M_{NRv} \cup M_{BLu} \cup M_{BLv}. \quad (22)$$

Consumer budget constraints and firm values differ across the four subsets in equation (22). Take these in turn.

First, for all intellectual assets where the legal structure permits only conventional dissemination, the consumer begins period t holding shares s_{mt} , each valued at price q_{mt} . Out of the sale of these shares, the consumer pays for consumption $p_{mt}c_{mt}$ and next period's asset holdings $q_{m,t+1}s_{m,t+1}$. The firm takes in revenue $p_{mt}c_{mt}$ in each period t . Thus, since post-instantiation costs are zero, the firm's market value at time t is

$$q_{mt} s_{mt} = \sum_{j=0}^{\infty} p_{m,t+j} c_{m,t+j}. \quad (23)$$

Depending on whether technology is NR or BL , the firm's stock of productive asset accumulates according to either equation (3) or (4).

Next, consider intellectual assets where the legal structure permits unrestricted dissemination. In addition to the description of the previous paragraph, at the end of period t the consumer has, depending on the technology, either $\omega_m c_{mt}$ or $\hat{\gamma}_m c_{mt}$ in durable assets. He can then compete with incumbent firms or sell this quantity of durable asset back to them at unit price $q_{m,t+1}$. Notice that whether the technology is nonrival does not qualitatively affect the consumer's positive

asset accumulation, only whether the coefficient is ω_m or $\hat{\gamma}_m$, both positive. Since in all cases the technology is constant returns to scale, the quantity produced in the economy is invariant to the pattern of ownership in s_m . Thus, without loss, assume the consumer sells all the freshly-generated stock back to the firm. Now, the firm's revenue flow in period t is either $(p_{mt} - q_{m,t+1}\omega_m)c_{mt}$ or $(p_{mt} - q_{m,t+1}\hat{\gamma}_m)c_{mt}$ depending on whether technology is *NR* or *BL* respectively.

Recall good $m = 0$ has technology *BL* with $\gamma_m - \hat{\gamma}_m = 1$. From earlier reasoning its distribution follows conventional dissemination Defn. 4.2. Thus, $m = 0$ is easily considered together with M_{BLV} in partition (22).

The representative consumer begins time 0 with shareholdings equal to the market portfolio. The consumer's budget constraint contains terms from dealing in each of the four possible asset-market groups given in the partition (22), plus that for $m = 0$.

All is in place now to define and analyze competitive equilibrium from period $t = 0$ on. For convenience, I explicitly restate consumer's and firms' optimization problems. The consumer's optimization problem is:

$$\begin{aligned} & \sup_{\{c_{mt}, s_{m,t+1}\}} \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in M \cup \{0\}} U_m(c_{mt}) \right] \quad (24) \\ \text{s.t.} \quad & \sum_{m \in M_{NRu}} (p_{mt} c_{mt} + q_{m,t+1} s_{m,t+1} - q_{mt} s_{mt} - q_{m,t+1} \omega_m c_{mt}) \\ & + \sum_{m \in M_{BLu}} (p_{mt} c_{mt} + q_{m,t+1} s_{m,t+1} - q_{mt} s_{mt} - q_{m,t+1} \hat{\gamma}_m c_{mt}) \\ & + \sum_{m \in M_{NRv} \cup M_{BLv} \cup \{0\}} (p_{mt} c_{mt} + q_{m,t+1} s_{m,t+1} - q_{mt} s_{mt}) \leq 0; \\ & \text{and } s_{m0} > 0, m \in M \cup \{0\}, \text{ given.} \end{aligned}$$

Notice the terms in the consumer's budget constraint for markets whose legal structures allow only conventional dissemination are invariant to the underlying technology. The initial pre-period 0 endowment \mathcal{E} has been absorbed into the period 0 portfolio of asset holdings by $\mathcal{E} = \sum_{m \in M \cup \{0\}} s_{m0}$.

Turn to firms. For $m \in M$, following instantiation the firm has $s_{m0} = s_m^\dagger > 0$. Recall that when a firm operates in a market whose legal structure allows unrestricted dissemination, the firm's revenue includes a negative term for buying back however many copies the consumer has generated. Otherwise, revenue is just price multiplied by consumption service flow. Additionally, a firm's supply of $s_{m,t+1}$ varies with whether its technology is *NR* or *BL*. Thus, firms' optimization problems vary according to partition (22) as follows:

$$\begin{aligned} m \in M_{NRu} : \quad & \sup_{\{c_{mt}, s_{m,t+1}\}} q_{m0} s_{m0} = \sum_{t=0}^{\infty} (p_{mt} - q_{m,t+1} \omega_m) c_{mt} \quad (25) \\ \text{s.t.} \quad & s_{m,t+1} = \gamma_m s_{mt} + \omega_m c_{mt}, \quad t = 0, 1, 2, \dots, \\ & s_{m0} > 0 \text{ given;} \end{aligned}$$

$$\begin{aligned} m \in M_{NRv} : \quad & \sup_{\{c_{mt}, s_{m,t+1}\}} q_{m0} s_{m0} = \sum_{t=0}^{\infty} p_{mt} c_{mt} \quad (26) \\ \text{s.t.} \quad & s_{m,t+1} = \gamma_m s_{mt} + \omega_m c_{mt}, \quad t = 0, 1, 2, \dots, \\ & s_{m0} > 0 \text{ given;} \end{aligned}$$

$$\begin{aligned} m \in M_{BLu} : \quad & \sup_{\{c_{mt}, s_{m,t+1}\}} q_{m0} s_{m0} = \sum_{t=0}^{\infty} (p_{mt} - q_{m,t+1} \hat{\gamma}_m) c_{mt} \quad (27) \\ \text{s.t.} \quad & s_{m,t+1} = \gamma_m s_{mt} - (\gamma_m - \hat{\gamma}_m) c_{mt}, \quad t = 0, 1, 2, \dots, \\ & s_{m0} > 0 \text{ given;} \end{aligned}$$

and, lastly, including $m = 0$,

$$\begin{aligned} m \in M_{BLv} \cup \{0\} : \quad & \sup_{\{c_{mt}, s_{m,t+1}\}} q_{m0} s_{m0} = \sum_{t=0}^{\infty} p_{mt} c_{mt} \quad (28) \\ \text{s.t.} \quad & s_{m,t+1} = \gamma_m s_{mt} - (\gamma_m - \hat{\gamma}_m) c_{mt}, \quad t = 0, 1, 2, \dots, \\ & s_{m0} > 0 \text{ given.} \end{aligned}$$

Arrow-Debreu competitive equilibrium in this model can now be specified.

Definition 4.3 For initializing (s, M) , a competitive equilibrium or CE is a collection of prices and quantities

$$\{p_{mt}^e, q_{mt}^e, c_{mt}^e, s_{m,t+1}^e\}, \quad m \in M \cup \{0\}, \quad t = 0, 1, 2, \dots,$$

such that setting $s_{00} = s$ and $s_{m0} = s_m^\dagger$, all $m \in M$, in problems (24)–(28), then

- (a) quantities $(c_{mt}^e, s_{m,t+1}^e)$ solve the consumer's problem (24) at prices (p_{mt}^e, q_{mt}^e) ; and
- (b) depending on m either 0 or in an element of partition (22), quantities $(c_{mt}^e, s_{m,t+1}^e)$ solve the appropriate problem in (25)–(28) at prices (p_{mt}^e, q_{mt}^e) .

■

The principal result in this section gives CE existence and (s, M) -efficiency, and characterizes CE prices:

Theorem 4.4 Let U_m , Γ_m , and \mathcal{E} satisfy Assumptions 3.1, 3.3, 3.7, and 3.10. Take a feasible initializing (s, M) , and partition M following equation (22). Suppose the representative consumer solves problem (24) and firms $m \in M \cup \{0\}$ solve one of (25)–(28), depending on m in the appropriate element of partition (22) or $m = 0$. Then CE exists and achieves (s, M) -efficiency. (Recall that $*$ denotes (s, M) -efficient allocations from Theorem 3.9, whereas e denotes CE quantities and prices.) Defining $\mu = 1/U'_0(c_{00}^*)$, CE prices satisfy:

- (a) for $m \in M_{NRu}$,

$$q_{mt}^e = \beta^t \left[\sum_{j=0}^{\infty} [(\gamma_m + \omega_m) \beta]^j U'_m([\gamma_m + \omega_m]^{t+j} s_m^\dagger) \right] \times \mu,$$

$$p_{mt}^e = \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \times \mu + q_{m,t+1}^e \omega_m;$$

- (b) for $m \in M_{NRv}$,

$$q_{mt}^e = \beta^t \left[\sum_{j=0}^{\infty} [(\gamma_m + \omega_m) \beta]^j U'_m([\gamma_m + \omega_m]^{t+j} s_m^\dagger) \right] \times \mu,$$

$$p_{mt}^e = \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \times \mu;$$

- (c) for $m \in M_{BLu}$,

$$q_{mt}^e = p_{mt}^e = \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-t} U'_m(c_{m0}^*) \times \mu; \text{ and}$$

- (d) for $m \in M_{BLv} \cup \{0\}$,

$$q_{mt}^e = \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-t} U'_m(c_{m0}^*) \times \mu,$$

$$p_{mt}^e = \gamma_m^{-t} p_{m0}^e = \gamma_m^{-t} U'_m(c_{m0}^*) \times \mu.$$

In all cases (a)–(d), for V_m the value functions in Theorem 3.9,

$$q_{m0}^e = V'_m(s_{m0}) \times \mu, \quad \forall m \in M \cup \{0\}. \quad (29)$$

■

Two critical implications from Theorem 4.4 follow: First, competitive equilibrium is efficient under a range of legal conventions for disseminating ideas, both restrictive and unrestrictive. Second, in competitive equilibrium, the initial asset valuation, or IPO price relative to numeraire, equals the derivative of the corresponding value function.

Compare q_{mt}^e 's in Theorem 4.4 across cases (a) and (b), and across (c) and (d), respectively. Holding preferences fixed, asset prices vary only with the underlying technology: Asset prices are independent of legal structure in markets for the consumption services generated by the corresponding asset. By contrast, rental prices p_{mt}^e do vary with legal structures: they are systematically higher under unrestricted dissemination than conventional. From Theorem 4.4, for technology *NR* the difference in p_{mt}^e between unrestricted and conventional dissemination schemes is $q_{m,t+1}^e \omega_m$. For technology *BL*, a direct calculation shows the analogous difference to be

$$\left[\frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} - 1 \right] \gamma_m^{-t} U'_m(c_{m0}^e) \times \mu = \frac{\hat{\gamma}_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-t} U'_m(c_{m0}^e) \times \mu$$

$$= \hat{\gamma}_m \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-(t+1)} U'_m(c_{m0}^e) \times \mu$$

$$= q_{m,t+1}^e \hat{\gamma}_m,$$

and so is exactly a counterpart to when the technology is NR , just swapping coefficients $\hat{\gamma}_m$ and ω_m .

Two conclusions follow from this discussion. First, it is irrelevant for allocations and asset valuation whether a legal system permits unrestricted or conventional dissemination—how much property rights accorded ideas restrict consumer behavior. In competitive equilibrium the difference surfaces in the prices charged consumers. Under unrestricted dissemination consumers pay a premium equal to the market value of copies they (can) create. Nothing else changes; consumption and investment remain unaltered.

One way to understand this equivalence is that alternative legal structures—when they don't reduce efficiency—allow economies, by definition, to attain outcomes on the efficient frontier. The optimum point on the frontier and the entire frontier itself depend only on Arrow-Debreu primitives, i.e., preferences and technology. When these primitives are invariant, so too are allocation outcomes.

Second, that those legal structures are simply a veil means that equation (29) determines innovation in competitive equilibrium across a range of circumstances. This brings us back to the question that began this section.

4.2 Competitive equilibrium innovation

By now analyzing what happens before instantiation occurs, Theorem 4.4—which describes what happens after—helps determine the innovation patterns that emerge in competitive equilibrium.

Definition 4.5 A competitive innovation equilibrium or CIE is a pair (M^e, CE) , with M^e a pattern of innovation and CE a collection of prices and quantities

$$\{p_{mt}^e, q_{mt}^e, c_{mt}^e, s_{m,t+1}^e : m \in M^e \cup \{0\}, t = 0, 1, 2, \dots\}$$

that is a competitive equilibrium for initializing (s_{00}^e, M^e) with $s_{00}^e = \mathcal{E} - \sum_{m \in M^e} \Psi_m$, and (M^e, CE) satisfies

$$m \in M^e \iff q_{m0}^e \geq q_{00}^e \psi_m.$$

Call M^e competitive innovation. ■

The definition states that competitive equilibrium when M^e is the set of instantiated ideas has IPO prices that, in turn, warrant instantiating exactly that set of ideas. The inequality defining $m \in M^e$ takes into account that resources $\Psi_m = \psi_m \times s_m^\dagger$ used for instantiating sector m have alternative use in sector 0 where those resources trade at price q_{00}^e per unit.

The following artificial construction is useful to help visualize CIE. Begin by hypothesizing a pattern of innovation $M \in \mathcal{M}$. Then compute prices q_{m0}^e , all $m \in M \cup \{0\}$, using $s_{00} = \mathcal{E} - \sum_{m \in M} \Psi_m$; $s_{m0} = s_m^\dagger$, all $m \in M$; and equation (29) in Theorem 4.4. Comparing the resulting per-unit gain q_{m0}^e with the real per-unit cost $q_{00}^e \psi_m$ determines innovation demand M' ,

$$m \in M' \iff q_{m0}^e \geq q_{00}^e \psi_m. \quad (30)$$

Say idea m is *activated from* M when $m \in M'$ in (30).

Define the *CE innovation* mapping, $\mathfrak{C} : \mathcal{M} \rightarrow \mathcal{M}$ by $\mathfrak{C}(M) = M'$ according to (30), for all $M \in \mathcal{M}$. Image $\mathfrak{C}(M)$ contains all ideas activated from M . In equilibrium, innovation demand must equal the initial hypothesized innovation, i.e.,

$$M' = \mathfrak{C}(M) = M.$$

Competitive innovation M^e , therefore, is a fixed point of \mathfrak{C} mapping the complete lattice \mathcal{M} , the collection of subsets of the space of ideas, to itself.⁸

⁸ Recall that a complete lattice is a partially ordered set for which every subset has least upper and greatest lower bounds (Birkhoff, 1967). Here, the lattice is \mathcal{M} , the set of subsets under the partial ordering set-inclusion. Lattice theory has been used in economic analysis to study monotone comparative statics and supermodular games (e.g., Milgrom and Shannon, 1994; Vives, 1999), and in econometrics to study stochastic processes over set-valued indexes, i.e., in the theory of random fields (e.g., Bulinskii and Zhurbenko, 1976; Goldie

With section 4.1 in place, it suffices to verify that a fixed point exists for \mathfrak{C} to establish existence of a CIE in Defn. 4.5. However, such fixedpoint results typically rely on the mapping being isotone, i.e., $M' \subseteq M \implies \mathfrak{C}(M') \subseteq \mathfrak{C}(M)$ —see, e.g., Birkhoff (1967, Theorem 11, p. 115). In the construction above, \mathfrak{C} fails isotonicity in general.

To see this, use equation (29) in condition (30) to obtain

$$m \in M' \iff V'_m(s_{m0}^e) \geq V'_0(s_{00}^e) \psi_m.$$

Since instantiation fixes $s_{m0}^e = s_m^\dagger$, the left side of the inequality is constant while the right side varies with $s_{00}^e = \mathcal{E} - \sum_{m \in M} \Psi_m$. From Theorem 3.9, the value function is concave, so that the right side increases with dominating innovations as s_{00}^e falls.⁹ Therefore, whereas a fixed m might be activated from innovation M' , it might then become deactivated from innovation $M \supset M'$, so that $\mathfrak{C}(M) \not\subseteq \mathfrak{C}(M')$.

The economics here is straightforward. When devoting more resources to R&D raises the value of the ordinary asset $m = 0$, previously only marginal R&D projects will become unprofitable.

Independent of isotonicity, however, notice that only for $m \in M$ is the inequality in (30) considered. Therefore, $\mathfrak{C}(M) \subseteq M$ and $\mathfrak{C}(\emptyset) = \emptyset$. The same argument showing \mathfrak{C} not isotone implies also $\mathfrak{C}(\mathfrak{C}(M)) = \mathfrak{C}(M)$, i.e., every image $\mathfrak{C}(M)$ is a fixed point of \mathfrak{C} . Since $\mathfrak{C}(M) \subseteq M$, quantity s_{00}^e is never lower with $\mathfrak{C}(M)$ instead of M . But then $q_{m0}^e/q_{00}^e = V'_m(s_{m0}^e)/V'_0(s_{00}^e)$ must be at least as high under $\mathfrak{C}(M)$ as under M . Any idea m activated from M remains activated from one iteration of \mathfrak{C} .

and Greenwood, 1986). The theory's distinctive feature is that it uses only three basic concepts: partial ordering and least upper and greatest lower bounds—not any topological structure such as compactness or continuity.

⁹ A function of a single variable that is concave, increasing, and continuously differentiable has its derivative decreasing. If this isn't immediately obvious, it can also be seen from the proof of Lemma 4.7 below or, since V_0 is not in general twice differentiable, by adapting Theorem 4.4 in Rockafellar (1970, p. 26).

That the null set \emptyset is an equilibrium competitive innovation makes economic sense in this model.¹⁰ A society that doesn't contemplate innovating never sees a reason to start. Equilibrium prices are potentially positive only for that set of ideas under consideration. Interestingly, the same reasoning also says that, then, even in the best-behaved outcomes, if non-null fixed points of \mathfrak{C} exist, they will typically be multiple. Putting these observations together, if there is a unique fixed point, the null set \emptyset is it. Moreover, that can be for reasonable economic circumstances: If for all $m \in \mathbb{M}$, instantiation costs ψ_m are sufficiently high, or if society's total wealth \mathcal{E} is sufficiently low, driving up $V'_0(s_{00}^e)$, then the natural CIE is unique with $M^e = \emptyset$.

We seek to exclude the trivial outcomes just described—while not necessarily pathological, they are not particularly interesting either. CIE existence alone is not the critical difficulty; instead, it is existence of a nontrivial CIE. In what follows, two restrictions of the environment will be imposed: first, so that \mathfrak{C} is isotone; second, so that nonnull innovation emerges in equilibrium. The first obtains by having endowment \mathcal{E} sufficiently large. The second obtains by assuming some idea n has instantiation cost sufficiently low. Neither deletes null innovation from the set of equilibria—nor should they, for economically sensible reasons already described.

Assumption 4.6 *Assume for $m = 0$ that Bellman equation (47) in the Proof of Prop. 3.6 in section 7 has a unique solution $V_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$ that is continuously differentiable. Assume that endowment $\mathcal{E} > \sum_{m \in \mathbb{M}} \Psi_m$ so that $s_{00} \stackrel{\text{def}}{=} \mathcal{E} - \sum_{m \in \mathbb{M}} \Psi_m > 0$ and that there exists a*

¹⁰ To be clear, that $M^e = \emptyset$ does not mean there is no CIE. Instead, it says that in CIE, no innovation occurs. The economics corresponds appropriately to the mathematics: The lattice of all subsets of a given abstract set is complete. For that lattice, the greatest lower bound exists and equals the empty set \emptyset (e.g. Birkhoff, 1967, Ch. I, p. 7). It isn't that the greatest lower bound fails to exist if \emptyset is what that lower bound needs to be.

real-valued function $\widehat{\delta}(\mathcal{E}) \geq 0$ such that

$$|V'_0(s_{00}) - V'_0(\mathcal{E})| \leq \widehat{\delta}(\mathcal{E}) \quad \text{with} \quad \widehat{\delta}(\mathcal{E}) \rightarrow 0 \text{ as } \mathcal{E} \rightarrow \infty. \quad (31)$$

The conditions in Assumption 4.6 resemble those of the earlier Assumption 3.10. Assumption 4.6 is a statement of asymptotic relative linearity in V_0 . It says that when the endowment \mathcal{E} is sufficiently large, then even activating all ideas $m \in \mathbb{M}$ does not appreciably alter marginal valuation on the stock used for producing ordinary commodities. This asymptotic relative linearity strengthens the asymptotic effective linearity of Assumption 3.10.

Lemma 4.7 *Let $\{\Psi_m : m \in \mathbb{M}\}$ be bounded. Suppose Assumption 4.6. Then Assumption 3.10 holds, where $\delta(\mathcal{E})$ can be chosen proportional to $\widehat{\delta}(\mathcal{E})$.*

The converse, in general, does not hold, i.e., condition (19) does not imply condition (31). Condition (19) restricts variation only local to s_{00} , whereas condition (31) restricts variation between s_{00} and \mathcal{E} . Obviously, if V_0 were exactly linear, then both conditions (19) and (31) hold with $\delta = \widehat{\delta}$ identically zero.

Call innovation $M \in \mathcal{M}$ positive if $M \neq \emptyset$. To allow positive competitive innovation in equilibrium, intuition suggests that not all instantiation costs can be excessively high. The following makes this precise.

Assumption 4.8 *Suppose for some $n \in \mathbb{M}$ inducing the initializing $(s, M) = (\mathcal{E} - \Psi_n, \{n\})$ either:*

(a) *if n has NR technology then CE c_{00}^e for initializing (s, M) gives*

$$\sum_{t=0}^{\infty} [(\gamma_n + \omega_n) \beta]^t U'_n([\gamma_n + \omega_n]^t s_n^\dagger) \geq \psi_n \times U'_0(c_{00}^e); \quad (32)$$

(b) *or, if n has BL technology, then CE quantities c_{n0}^e and c_{00}^e for initializing (s, M) give*

$$\left[1 + \frac{\widehat{\gamma}_n}{\gamma_n - \widehat{\gamma}_n}\right] U'_n(c_{n0}^e) \geq \psi_n \times U'_0(c_{00}^e). \quad (33)$$

To interpret this, notice that a simpler, more transparent condition suffices simultaneously for inequalities (32) and (33) of Assumption 4.8. These inequalities have on their right the real cost, measured in good-0 utility, of instantiating good n . Their left sides, while different, have similar structure: each is $U'_n(c_{n0}^e)$ plus a positive quantity, varying with technology, so that both inequalities (32) and (33) would be implied by

$$U'_n(c_{n0}^e) \geq \psi_n \times U'_0(c_{00}^e), \quad (34)$$

at CE for initializing $(s, M) = (\mathcal{E} - \Psi_n, \{n\})$. Condition (34) says, roughly, that n 's instantiation cost is sufficiently low relative to the consumer's utility gain in consuming good n . Clearly, however, instances can arise satisfying inequalities (32) and (33) but not (34), so I keep Assumption 4.8 as stated.

Call \mathcal{M}^e the collection of all competitive innovations, i.e., all fixed points of \mathfrak{C} . Call a competitive innovation \overline{M}^e greatest if it includes all other competitive innovations,¹¹ i.e., $M \in \mathcal{M}^e \implies M \subseteq \overline{M}^e$.

Theorem 4.9 *Let U_m, Γ_m , and \mathcal{E} satisfy Assumptions 3.1, 3.3, 3.7, and 4.6. Suppose Assumptions 3.2 and 4.8. Then, as $\mathcal{E} \rightarrow \infty$, a greatest competitive innovation \overline{M}^e exists, with \overline{M}^e unique and positive,*

¹¹ It is tempting to call \overline{M}^e maximal, but that means something different in lattice theory (Birkhoff, 1967, Ch. I). Element M in \mathcal{M} is maximal when $M \subseteq M' \implies M = M'$ for $M' \in \mathcal{M}$. A greatest element of a partially ordered set is always maximal, but the converse is false in general. The two concepts are equivalent if the ordering is total. Clearly, if a greatest element of a partially ordered set exists, it is unique.

satisfying

$$\overline{M}^e = \bigcup_{M \in \mathcal{M}^e} M,$$

and

$$m \in \overline{M}^e \iff V'_m(s_m^\dagger) \geq V'_0(\mathcal{E}) \psi_m. \quad (35)$$

■

Theorem 4.9 is related to Tarski's Fixed Point Theorem, whereby an isotone mapping on a complete lattice has fixed points that themselves constitute a non-null complete lattice (the Technical Appendix section 7 contains details). Assumption 4.6 restricts the change in $V'_0(s_{00}^\dagger)$ to be sufficiently small so that \mathfrak{C} is isotone. Assumption 4.8 ensures that a fixed point of \mathfrak{C} , if one exists, can be positive. The empty set \emptyset , of course, always remains a fixed point.

Having positive innovation, however, doesn't guarantee social efficiency. To see this, compare competitive innovation in Theorem 4.9 with socially efficient innovation in Theorem 3.11.

Theorem 4.10 *Assume the conditions of Theorem 4.9. Then $\overline{M}^e \subseteq M^*$ and, necessarily, for all $m \in M^*$,*

$$V'_m(s_m^\dagger) s_m^\dagger < V_m(s_m^\dagger) - V_m(0). \quad (36)$$

Therefore, if further there exists $n \in \mathbb{M}$ with Ψ_n satisfying:

$$V'_n(s_n^\dagger) s_n^\dagger < V'_0(s_{00}) \times \Psi_n \leq V_n(s_n^\dagger) - V_n(0), \quad (37)$$

then $\overline{M}^e \subsetneq M^*$, i.e., the set inclusion is strict. ■

By the Proof of Lemma 4.7, the derivative V'_m is declining. Therefore, inequality (52) in the Proof of Theorem 4.10 gives, for positive ν , the inequality

$$V_m(s_{m0}) - V_m(0) > V'_m(s_m^\dagger) \times s_{m0}, \quad \text{all } s_{m0} \in (0, s_m^\dagger + \nu). \quad (38)$$

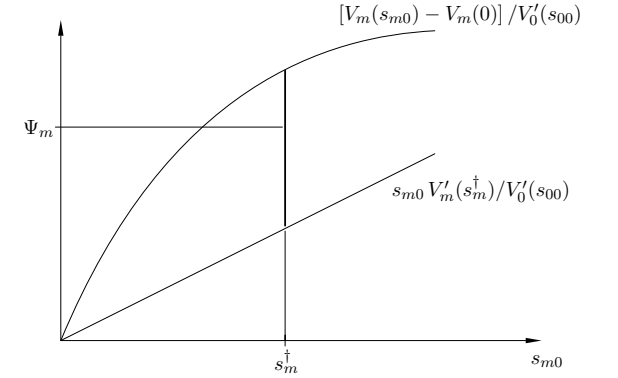


Figure 1: **Strict concavity implying inefficiency with s_m^\dagger fixed** Every idea m with cost Ψ_m below $s_{m0} V'_m(s_m^\dagger) / V'_0(s_{00})$ at $s_{m0} = s_m^\dagger$ is activated in competitive innovation equilibrium. Strict concavity and continuous differentiability in V_m implies that on $s_{m0} = s_m^\dagger$ the interval indicated by the darkened segment has positive length. If cost Ψ_m falls in that interval, then idea m is left uninstiated in CIE even though activation would be socially efficient. The signed separation between the two graphs is maximized at $s_{m0} = s_m^\dagger$.

Fig. 1 illustrates inequalities (37) and (38).¹²

Interpret the Figure as follows. Competitive markets price at the margin. With well-functioning complete dynamic markets in ideas, that margin is the value function's slope, $V'_m(s_{m0})V'_0(s_{00})^{-1}$, as s_{m0} varies in $[0, \infty)$. At $s_{m0} = s_m^\dagger$, reward to entrepreneurs, indicated in Fig. 1, is the straight line through the origin with this slope, readable directly off market prices. By contrast, social surplus $[V_m(s_{m0}) - V_m(0)] \times V'_0(s_{00})^{-1}$ is not priced, and appears explicitly in no market transaction.

Why doesn't this argument affect all goods, not just ideas? The key is instantiation to a minimum size by Assumption 3.2. We can assess this in Theorem 4.10 by asking what happens if we relax that restriction. It turns out then that an *invisible hand* property holds, where the greatest competitive equilibrium innovation is socially efficient.

In contrast to Assumption 3.2, consider instead:

Assumption 4.11 For all $m \in \mathbb{M}$, instantiation of the asset to s_{m0} costs

$$(\Psi_m/s_m^\dagger) \times s_{m0} = \psi_m \times s_{m0}, \quad \text{all } s_{m0} \in [0, s_m^\dagger].$$

Instantiating to $s_{m0} > s_m^\dagger$ is infeasible. ■

Assumption 4.11 replaces the fixed instantiation quantity restriction of Assumption 3.2 with one specifying a maximum at s_m^\dagger . This maintains the same total instantiation cost at s_m^\dagger at the maximum, but allows instantiation otherwise at constant marginal cost.

Theorem 4.12 Assume the conditions of Theorem 4.9, but replace Assumption 3.2 with Assumption 4.11. Then idea $m \in \mathbb{M}$ has

$$V_m(\bar{s}_{m0}) - V_m(0) \geq V'_0(\mathcal{E}) \psi_m \bar{s}_{m0} \quad \text{for some } \bar{s}_{m0} \in (0, s_m^\dagger] \quad (39)$$

¹² The right side of inequality (38) has units of price times quantity. Therefore, in a price-quantity diagram, both sides of inequality (37) are areas. Their positive difference is then a dynamic Dupuit triangle, corresponding precisely to area A in Fig. 4 of Romer (1994, p. 15). By construction, that Dupuit triangle area is maximized at $s_{m0} = s_m^\dagger$.

if and only if

$$V'_m(\bar{s}_{m0}) \geq V'_0(\mathcal{E}) \psi_m \quad \text{for some } \bar{s}_{m0} \in (0, s_m^\dagger]. \quad (40)$$

The Theorem states that the collection of ideas with net social benefit in (39) comprises exactly those ideas instantiated by competitive markets in (40). ■

That Assumption 4.11 and relations (39) and (40) restrict to $[0, s_m^\dagger]$ allow confining attention to constant V'_0 and simplifies the proof but is inessential to the conclusion of the Theorem. Since V_m is strictly increasing and strictly concave, if the inequalities in (39) and (40) fail on $[0, s_m^\dagger]$, then they fail everywhere.

To emphasize how Theorem 4.12 is an invisible hand result, observe that neither the construction of V_m nor the marginal condition (40) optimizes over patterns of innovations $M \in \mathcal{M}$. Markets determine only dissemination $(s_{m,t+1}, c_{mt})$ (for $t \geq 0$) and instantiation quantity s_{m0} . Theorem 4.12 asserts that such markets end up spontaneously optimizing over \mathcal{M} as well. However, this selection fails to be optimal if instantiation quantity $s_{m0} = s_m^\dagger$ is a priori fixed (Theorem 4.10).

Certain intuitively appealing proposals fail to resolve this inefficiency. First, while the problem is one of too large a minimal scale, it is not solved by having yet more initial resources to invest. Indeed, Theorem 4.10 already takes the sum total of instantiation costs Ψ_m to be trivially small relative to initial endowment \mathcal{E} . What matters is the slope of the value function at Ψ_m , not how large or small the quantity Ψ_m itself is. Forming coalitions of entrepreneurs, in an effort to deal with the minimum initial investment, does not alter that slope.

Second, randomizing to convexify the discreteness, perhaps an invention lottery, fails also to solve the problem. Any randomization will have as upper bound on investor returns the slope of $V_m(s_{m0}) - V_m(0)$ at $s_{m0} = s_m^\dagger$. Since that slope is too small in Theorem 4.10, the return from any randomized version remains again insufficient to activate a previously-deactivated idea. For this, also notice that

$V_m(s_{m0}) - V_m(0)$ vanishes at $s_{m0} = 0$ by construction, so that its graph in Fig. 1 always begins at the origin. Otherwise, lottery convexification could lift the value possibilities set and allow instantiation to beyond s_m^\dagger .

5 Quantifying social efficiency loss

From Fig. 1 and the surrounding discussion, the more concave the value function, the greater is the inefficiency in leaving dormant what would be socially beneficial ideas. This section presents some numerical examples illustrating the relative importance of dynamic Dupuit triangles.

[[GET SENSIBLE QUANTIFICATION]]

6 Conclusion

This paper has considered the problem of incentivizing innovation by competitive market pricing alone, without the social contrivance of IPRs.

Two kinds of nonconvexities potentially matter here. Boldrin and Levine (2002) and Quah (2002a) had shown that the first—nonconvexity due to nonrivalry and infinite expansibility—is, under relatively weak modification, irrelevant. Competitive equilibrium then prices ideas positively, and, conditional on a given set of ideas, is socially efficient.

The second nonconvexity—indivisibility in the instantiation of a new idea—is sometimes viewed as a fixed or sunk cost, suggesting that it too might be irrelevant, say when that cost is shared across a coalition of economic agents or when it becomes relatively small as an economy grows.

This paper has shown that this second nonconvexity leads to inefficiently low innovation. Fig. 1 and Theorems 4.10 and 4.12 characterize the inefficiency and relate it to the concept of a Dupuit triangle. They make precise how the inefficiency vanishes if the minimum-size

indivisibility becomes irrelevant. This occurs not by the indivisibility becoming small relative to the overall size of the economy, as would happen through economic growth or having entrepreneurs form sufficiently large coalitions or participating in innovation lotteries. Instead, what is needed (Theorem 4.12) is the capability to continuously adjust the level of an intellectual asset's instantiation quantity.

In developing these results, the paper has also obtained an auxiliary finding on rules for disseminating ideas—in particular for fair-use provision in copyright. Theorem 4.4 shows that different legal structures on rights to copy and disseminate ideas—Defns. 4.1 and Defns. 4.2—leave asset prices and patterns of innovation invariant.

References

- Arrow, Kenneth J. (1962) "Economic welfare and the allocation of resources for inventions," In *The Rate and Direction of Inventive Activity*, ed. Richard R. Nelson (Princeton: Princeton University Press and NBER) pp. 609–625
- Birkhoff, Garrett (1967) *Lattice Theory*, third ed., vol. 25 of *AMS Colloquium Publications* (Providence: American Mathematical Society)
- Boldrin, Michele, and David K. Levine (2002) "Perfectly competitive innovation," Staff Report 303, FRB Minneapolis Research Department, Minneapolis, March
- Bulinskii, A. V., and I. G. Zhurbenko (1976) "A central limit theorem for additive random functions," *Theory of Probability and Its Applications* 21(4), 687–697
- Dasgupta, Partha, and Eric Maskin (1987) "The simple economics of research portfolios," *Economic Journal* 97, 581–595, September
- Dasgupta, Partha, and Joseph Stiglitz (1980) "Industrial structure and the nature of innovative activity," *Economic Journal* 90, 266–293, June

- David, Paul A. (1992) “Knowledge, property, and the system dynamics of technological change,” *Proceedings of the World Bank Annual Conference on Development Economics* pp. 215–248, March
- Goldie, Charles M., and Priscilla E. Greenwood (1986) “Characterizations of set-indexed brownian motion and associated conditions for finite-dimensional convergence,” *Annals of Probability* 14(3), 802–816, July
- Hellwig, Martin, and Andreas Irmen (2001) “Endogenous technological change in a competitive economy,” *Journal of Economic Theory* 101(1), 1–39
- Helpman, Elhanan (1993) “Innovation, imitation, and intellectual property rights,” *Econometrica* 61(6), 1247–1280, November
- Jaffee, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993) “Geographic localization of knowledge spillovers as evidenced in patent citations,” *Quarterly Journal of Economics* 108(3), 577–598, August
- Jones, Charles I., and John C. Williams (2000) “Too much of a good thing? The economics of investment in R&D,” *Journal of Economic Growth* 5(1), 65–85, March
- Keely, Louise C. (2001) “Using patents in growth models,” *Economics of Innovation and New Technology* 10, 449–492
- Klemperer, Paul (1990) “How broad should the scope of patent protection be?,” *RAND Journal of Economics* 21(1), 113–130, Spring
- Merton, Robert K. (1973) *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago: University of Chicago Press)
- Milgrom, Paul, and Chris Shannon (1994) “Monotone comparative statics,” *Econometrica* 62(1), 157–180, January
- Nordhaus, William D. (1969) *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change* (Cambridge: MIT Press)

- O’Donoghue, Ted, and Josef Zweimüller (1998) “Patents in a model of endogenous growth,” Discussion Paper 1951, CEPR, London, August
- Quah, Danny (2002a) “24/7 competitive innovation,” Working Paper, Economics Department, LSE, London, May
- (2002b) “Matching demand and supply in a weightless economy: Market-driven creativity with and without IPRs,” Working Paper, Economics Department, LSE, London, March
- Rockafellar, R. Tyrrell (1970) *Convex Analysis*, vol. 28 of *Princeton Mathematical Series* (Princeton: Princeton University Press)
- Romer, Paul M. (1990) “Endogenous technological change,” *Journal of Political Economy* 98(5, part 2), S71–S102, October
- (1994) “New goods, old theory, and the welfare costs of trade restrictions,” *Journal of Development Economics* 43(1), 5–38
- Scotchmer, Suzanne (1991) “Standing on the shoulders of giants: Cumulative research and the patent law,” *Journal of Economic Perspectives* 5(1), 29–41, Winter
- Shell, Karl (1966) “Toward a theory of inventive activity and capital accumulation,” *American Economic Review* 56(1/2), 62–68, March
- Stokey, Nancy L., and Robert E. Lucas (1989) *Recursive Methods in Economic Dynamics* (Cambridge: Harvard University Press)
- Vives, Xavier (1999) *Oligopoly Pricing* (Cambridge: MIT Press)

7 Technical Appendix

This section contains all proofs to results in the text. Begin by casting the model in the framework of Stokey and Lucas (1989, Ch. 4). Do this now rather than clutter up the Proof of Prop. 3.6 below with definitions that will be used yet again in other proofs.

From preferences (1) and technology (5) and (7), define for each $m \in \mathbb{N}$ the one-period return function $F_m : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$. If technology is NR as in equation (3), then

$$F_m(s_m, s'_m) = U_m(s_m), \quad (41)$$

while if it is BL equation (4), then

$$F_m(s_m, s'_m) = U_m\left(\frac{\gamma_m s_m - s'_m}{\gamma_m - \hat{\gamma}_m}\right). \quad (42)$$

In all cases, F_m is strictly increasing and strictly concave in its first argument.

Notice that for NR technology (3), if in place of equation (41), we had simply substituted $c_m = (s'_m - \gamma_m s_m) \omega_m^{-1}$ into $U_m(c_m)$ to give

$$F_m(s_m, s'_m) = U_m\left(\frac{s'_m - \gamma_m s_m}{\omega_m}\right),$$

the result would incorrectly suggest F_m is decreasing in s_m .

If we use the accumulation equation (6) then F_m has its definition vary appropriately with $\partial g_m / \partial c_{mt} \geq 0$. Equation (41) is unchanged but equation (42) becomes more complicated, involving a partial inverse of the accumulation function g . However, the argument in the proof below remains essentially unchanged.

Prop. 3.6 obtains by simply checking conditions for applying Theorems 4.2–4.3 and 4.6–4.8 in Stokey and Lucas (1989).

Proof of Prop. 3.6 Since by Assumption 3.1 function U_m is non-

negative and bounded, we can rearrange terms:

$$\begin{aligned} U(c) &= \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in \mathbb{N}} U_m(c_{mt}) \right] \\ &= \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in M \cup \{0\}} U_m(c_{mt}) + \sum_{m \in \mathbb{N} \setminus (M \cup \{0\})} U_m(c_{mt}) \right] \\ &= \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in M \cup \{0\}} U_m(c_{mt}) \right] \\ &\quad + \sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in \mathbb{N} \setminus (M \cup \{0\})} U_m(c_{mt}) \right]. \end{aligned} \quad (43)$$

By $\Gamma_m(0) = \{0\}$ (end of section 3.3), $c_{mt} \in [0, s_{mt}]$, and $U_m(0) = 0$ in Assumption 3.1, the second summand on the right of (43) converges and can be rewritten as $\sum_{m \in \mathbb{N} \setminus (M \cup \{0\})} (1-\beta)^{-1} U_m(0)$. From M finite and U_m bounded, we can rearrange the first summand:

$$\sum_{t=0}^{\infty} \beta^t \left[\sum_{m \in M \cup \{0\}} U_m(c_{mt}) \right] = \sum_{m \in M \cup \{0\}} \left[\sum_{t=0}^{\infty} \beta^t U_m(c_{mt}) \right]. \quad (44)$$

Neither $\{U_m : m \in \mathbb{N}\}$ nor $\{\Gamma_m : m \in \mathbb{N}\}$ shows inter-dependence across m . Thus, (s, M) -efficiency solves for each $m \in M \cup \{0\}$,

$$\begin{aligned} \sup_{\{c_{mt}, s_{m,t+1}\}_{t=0}^{\infty}} & \sum_{t=0}^{\infty} \beta^t U_m(c_{mt}) \\ \text{s.t. } & c_{mt} \in [0, s_{mt}] \\ & s_{m,t+1} \in \Gamma_m(s_{mt}) \\ & s_{m0} = s_m^\dagger \text{ given.} \end{aligned} \quad (45)$$

Consider this optimization problem for all m , not just those in $M \cup \{0\}$, by replacing condition (45) with s_{m0} given at an arbitrary fixed value (i.e., not necessarily only at s_m^\dagger and, indeed, possibly at 0). For all $m \in \mathbb{N}$ define function F_m as either (41) or (42), depending on

Γ_m from either (3) or (4) respectively. Then the program above is equivalent to

$$\begin{aligned} & \sup_{\{s_{m,t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t F_m(s_{mt}, s_{m,t+1}) \\ & \text{s.t. } s_{m,t+1} \in \Gamma_m(s_{mt}) \\ & \quad s_{m0} \text{ given.} \end{aligned} \quad (46)$$

By Assumption 3.1 and the definition in either (41) or (42), function F_m is bounded, continuous, and strictly increasing in its first argument. Assumption 3.3 shows that correspondence Γ_m from either (3) or (4) is nonempty, compact-valued, continuous, monotone, and convex. Define V_m as the solution to the functional equation

$$V_m(s) = \sup_{s' \in \Gamma_m(s)} [F_m(s, s') + \beta V_m(s')], \quad \text{all } s \in \mathbb{R}_+. \quad (47)$$

Since U_m is bounded we can restrict analysis to bounded solutions of equation (47). By Theorems 4.2–4.3 in Stokey and Lucas (1989) the solution to equation (47) exists and is unique, and equals the supremum function in problem (46). If $s = 0$ then equation (47) gives $V_m(0) = (1 - \beta)^{-1} U_m(0)$. Using these conclusions in equations (43) and (44) establishes representation (10). By Theorem 4.6 in Stokey and Lucas (1989) V_m is bounded and continuous. Although in equation (41) return function F_m shows no dependence on its second argument, it satisfies the strict inequality in Assumption 4.7 in Stokey and Lucas (1989). Thus, return function F_m in equation (41) is therefore concave in the sense of Assumption 4.7 in Stokey and Lucas (1989) while equation (42) obviously has F_m similarly so. Theorems 4.7–4.8 of Stokey and Lucas (1989) then imply V_m is strictly increasing and strictly concave. *Q.E.D.*

Because technology (3) always implies optimum $s_{m,t+1}$ on the boundary of $\Gamma(s_{mt})$, we cannot apply Theorem 4.11 from Stokey and Lucas (1989) to verify differentiability in V_m solving the Bellman equation (47). Thus section 3.5 introduced Assumption 3.7 and described sufficient conditions for it in Example 3.8.

Proof of claims in Example 3.8 Equation (2) gives the derivative $U'_m(c_m) = c_m^{-1/\sigma_m}$. For (a), directly calculate

$$\begin{aligned} & \sum_{t=0}^{\infty} ([\gamma_m + \omega_m] \beta)^t \times U'_m([\gamma_m + \omega_m]^t) \\ & = \sum_{j=0}^{\infty} \left([\gamma_m + \omega_m]^{1-\sigma_m^{-1}} \times \beta \right)^t. \end{aligned}$$

This infinite series in the non-negative sequence

$$\left([\gamma_m + \omega_m]^{1-\sigma_m^{-1}} \times \beta \right)^t, \quad t = 0, 1, 2, \dots$$

converges exactly when

$$[\gamma_m + \omega_m]^{1-\sigma_m^{-1}} \times \beta < 1 \iff \beta < [\gamma_m + \omega_m]^{\sigma_m^{-1}-1}.$$

Since $\gamma_m + \omega_m > 1 > \beta > 0$ the last inequality holds if $\sigma_m \leq 1$, thereby letting the sequence be summable and its tail sums converge to zero as in equation (11). For (b) then, notice the ratio $U'_m(s_m)/U'_m(\hat{\gamma}_m s_m)$ equals $[\hat{\gamma}_m]^{1/\sigma_m}$, independent of s_m . Equation (12) then holds if and only if $\gamma_m > \beta^{-1} \times [\hat{\gamma}_m]^{1/\sigma_m}$. *Q.E.D.*

Proof of Theorem 3.9 We need to show that an (s, M) -efficient allocation exists and satisfies conditions (13), (14), (16), and (17). Then from Prop. 3.6 it remains only to verify continuous differentiability in V_m and calculate derivatives to obtain equations (15) and (18). That an (s, M) -efficient allocation exists is immediate as problem (47) in the Proof of Prop. 3.6 has return function F_m continuous and constraint set $\Gamma_m(s)$ nonempty and compact. For NR technology (3), the solution to problem (46) in the Proof of Prop. 3.6 has $c_{mt} = s_{mt}$ and $s_{m,t+1} = [\gamma_m + \omega_m] s_{mt} = [\gamma_m + \omega_m]^{t+1} s_{m0}$, verifying conditions (13) and (14). By Stokey and Lucas (1989, Theorems 4.2–4.3), the unique solution to Bellman's equation (47) coincides with the implied supremum function in problem (46). The latter is:

$$V_m(s_{m0}) = \sum_{t=0}^{\infty} \beta^t U_m([\gamma_m + \omega_m]^t s_{m0}), \quad \text{all } s_{m0} \in \mathbb{R}_+.$$

Partition the infinite sum on the right into a leading finite sum and consequent infinite tail sum. The finite sum is continuously differentiable. Similarly so is the entire right side, which is then just the limit of the first term precisely when the tail sum has derivative satisfying

$$\lim_{t \rightarrow \infty} \sum_{j=0}^{\infty} ([\gamma_m + \omega_m] \beta)^{t+j} \times U'_m([\gamma_m + \omega_m]^{t+j} s_{m0}) = 0,$$

for all $s_{m0} \in \mathbb{R}_+$. But this follows from condition (11) in Assumption 3.7. Thus, for NR technology, V_m is continuously differentiable with

$$V'_m(s_{m0}) = \sum_{t=0}^{\infty} ([\gamma_m + \omega_m] \beta)^t \times U'_m([\gamma_m + \omega_m]^t s_{m0}), \quad s_{m0} \in \mathbb{R}_+,$$

establishing equation (15). For BL technology (4), the first-order conditions for problem (46) or, equivalently, problem (47) include

$$U'_m(c_{mt}) \geq \beta U'_m(c_{m,t+1}) \gamma_m \implies \gamma_m \leq \beta^{-1} \frac{U'_m(c_{mt})}{U'_m(c_{m,t+1})}.$$

This inequality is strict at the optimum if given $s_{mt} > 0$ the last term on the right, evaluated at the upper boundary of $\Gamma_m(s_{mt})$, i.e., setting $c_{mt} = s_{mt}$ to give

$$\frac{\beta^{-1} U'_m(s_{mt})}{U'_m(c_{m,t+1})} < \frac{\beta^{-1} U'_m(s_{mt})}{U'_m(\widehat{\gamma}_m s_{mt})},$$

continues to dominate γ_m strictly. But by inequality (12) in Assumption 3.7, this is impossible. Since $U'_m(c_{mt}) \rightarrow \infty$ as $c_{mt} \rightarrow 0$, the other boundary point is also never relevant when $s_{mt} > 0$. Therefore, $0 < c_{mt} < s_{mt}$ so that optimal $s_{m,t+1}$ lies in the interior of $\Gamma_m(s_{mt})$, establishing condition (16). Next, iterate equation (4) forwards from $t = 0$,

$$\begin{aligned} s_{m0} &= (\gamma_m - \widehat{\gamma}_m) \gamma_m^{-1} c_{m0} + \gamma_m^{-1} s_1 \\ &= (\gamma_m - \widehat{\gamma}_m) \gamma_m^{-1} c_{m0} + (\gamma_m - \widehat{\gamma}_m) \gamma_m^{-2} c_{m1} + \gamma_m^{-2} s_{m2} \\ &= (\gamma_m - \widehat{\gamma}_m) \gamma_m^{-1} \sum_{t=0}^{\infty} \gamma_m^{-t} c_{mt} + \lim_{t \rightarrow \infty} \gamma_m^{-t} s_{mt}. \end{aligned}$$

Whenever $\lim_{t \rightarrow \infty} \gamma_m^{-t} s_{mt} > 0$ an allocation can be strictly improved, allowing greater c_{mt} while maintaining condition (16). Thus, at the socially efficient allocation, $\lim_{t \rightarrow \infty} \gamma_m^{-t} s_{mt} = 0$, verifying equation (17) at $t = 0$. Substituting this into equation (4), first at $t = 1$ and then recursively, delivers equation (17) for all t . Thus, by the interiority of the optimum just established for BL technology (4), Theorem 4.11 of Stokey and Lucas (1989) applies to V_m , establishing continuous differentiability in the interior of \mathbb{R}_+ . Moreover,

$$F_m(s_{mt}, s_{m,t+1}) = U_m \left(\frac{\gamma_m}{\gamma_m - \widehat{\gamma}_m} s_{mt} - \frac{1}{\gamma_m - \widehat{\gamma}_m} s_{m,t+1} \right),$$

so that again by Theorem 4.11 in Stokey and Lucas (1989),

$$V'_m(s_{m0}) = \frac{\gamma_m}{\gamma_m - \widehat{\gamma}_m} U'_m(c_{m0}^*),$$

establishing equation (18).

Q.E.D.

Proof of Theorem 3.11 From Assumption 3.10, necessarily $s^* > 0$, so that (s^*, M^*) is feasible. Take $n \in M^* \neq \emptyset$ but suppose the inequality in (20) fails. Let $s = s^* + \Psi_n > \mathcal{E} - \sum_{m \in \mathbb{M}} \Psi_m$ and $M = M^* \setminus \{n\}$. Then (s, M) is feasible and

$$\begin{aligned} W(s, M) - W(s^*, M^*) &= [V_0(s) - V_0(s^*)] - [V_n(s_n^\dagger) - V_n(0)] \\ &= [V_0(s^* + \Psi_n) - V_0(s^*) - V_0'(s^*) \Psi_n] \\ &\quad - [V_n(s_n^\dagger) - V_n(0) - V_0'(s^*) \Psi_n]. \end{aligned}$$

By hypothesis, the inequality in (20) fails, so that the last term is (together with its leading minus) positive and dominant from Assumption 3.10. Thus, $W(s, M) > W(s^*, M^*)$, so that (s^*, M^*) cannot have been socially efficient. Conversely, suppose for $n \in \mathbb{M}$ we have $n \notin M^*$ but inequality (20) holds. Let $s = s^* - \Psi_n > 0$ and $M = M^* \cup \{n\}$. Since M^* is finite, we still have $s > \mathcal{E} - \sum_{m \in \mathbb{M}} \Psi_m$. Then (s, M) is feasible and

$$\begin{aligned} W(s, M) - W(s^*, M^*) &= -[V_0(s^*) - V_0(s)] + [V_n(s_n^\dagger) - V_n(0)] \\ &= -[V_0(s^*) - V_0'(s^*) \Psi_n - V_0(s^* - \Psi_n)] \\ &\quad + [V_n(s_n^\dagger) - V_n(0) - V_0'(s^*) \Psi_n]. \end{aligned}$$

From inequality (20) the last summand is positive and, by Assumption 3.10, dominant. Therefore, $W(s, M) > W(s^*, M^*)$, so that again (s^*, M^*) cannot have been socially efficient. Summarizing, M^* implies a socially efficient (s^*, M^*) if and only relation (20) holds.

Q.E.D.

Proof of Theorem 4.4 Verify that when consumer and firms face prices

$$\{p_{mt}^e, q_{mt}^e : m \in M^e \cup \{0\}, t = 0, 1, 2, \dots\},$$

then setting

$$\{c_{mt}^e, s_{m,t+1}^e : m \in M^e \cup \{0\}, t = 0, 1, 2, \dots\}$$

to the $*$ -denoted, (s, M) -efficient values in Prop. 3.9 solves, simultaneously, consumer's problem (24) and firms' problems (25)–(28). Take first consumers. Denote by λ_t the Lagrange multiplier on the time- t budget constraint in problem (24). The consumer's FOCs then include:

$$\begin{aligned} \lambda_t q_{m,t+1} &= \beta \lambda_{t+1} q_{m,t+1}, & m \in M \cup \{0\}; & (48) \\ U'_m(c_{mt}) &= p_{mt} \lambda_t, & m \in M_{NRv} \cup M_{BLv} \cup \{0\}; \\ U'_m(c_{mt}) &= (p_{mt} - q_{m,t+1} \omega) \lambda_t, & m \in M_{NRu}; \\ U'_m(c_{mt}) &= (p_{mt} - q_{m,t+1} \hat{\gamma}) \lambda_t, & m \in M_{BLu}. \end{aligned}$$

Equation (48) gives $\lambda_t = \beta \lambda_{t+1}$, so that the remaining FOCs above become:

$$\begin{aligned} U'_m(c_{mt}) p_{mt}^{-1} &= \beta U'_m(c_{m,t+1}) p_{m,t+1}^{-1}, & m \in M_{NRv} \cup M_{BLv} \cup \{0\}; \\ U'_m(c_{mt}) (p_{mt} - q_{m,t+1} \omega)^{-1} &= \beta U'_m(c_{m,t+1}) (p_{m,t+1} - q_{m,t+2} \omega)^{-1}, & m \in M_{NRu}; \\ U'_m(c_{mt}) (p_{mt} - q_{m,t+1} \hat{\gamma})^{-1} &= \beta U'_m(c_{m,t+1}) (p_{m,t+1} - q_{m,t+2} \hat{\gamma})^{-1}, & m \in M_{BLu}. \end{aligned}$$

At the hypothesized (p^e, q^e) , statements (a)–(d) give:

$$\begin{aligned} m \in M_{NRu} : p_{mt}^e - q_{m,t+1}^e \omega &= \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \times \mu \\ \implies c_{mt}^e &= [\gamma_m + \omega_m]^t s_m^\dagger = c_{mt}^* \end{aligned}$$

and

$$\begin{aligned} m \in M_{NRv} : p_{mt}^e &= \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \times \mu \\ \implies c_{mt}^e &= [\gamma_m + \omega_m]^t s_m^\dagger = c_{mt}^* \end{aligned}$$

by equations (13)–(14);

$$\begin{aligned} m \in M_{BLu} : \\ p_{mt}^e - q_{m,t+1}^e \hat{\gamma} &= \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-t} U'_m(c_{m0}^*) [1 - \gamma_m^{-1} \hat{\gamma}_m] \times \mu \\ \implies U'_m(c_{mt}^e) &= \beta U'_m(c_{m,t+1}^e) \gamma_m, \\ \implies c_{mt}^e &= c_{mt}^*, \quad \text{from equation (16)} \end{aligned}$$

and finally,

$$\begin{aligned} m \in M_{BLv} \cup \{0\} : p_{mt}^e &= \gamma_m^{-t} U'_m(c_{m0}^*) \times \mu \\ \implies U'_m(c_{mt}^e) &= \beta U'_m(c_{m,t+1}^e) \gamma_m, \\ \implies c_{mt}^e &= c_{mt}^*, \quad \text{again from equation (16)}. \end{aligned}$$

At CE (p^e, q^e) the consumer is indifferent across alternative portfolio holdings, so it remains only to ensure that (c^e, s^e) satisfies the budget constraint. For $m \in M_{NRu}$,

$$\begin{aligned} q_{m,t+1}^e s_{m,t+1}^e &= q_{mt}^e s_{mt}^e - (p_{mt}^e - q_{m,t+1}^e \omega_m) c_{mt}^* \\ &= q_{mt}^e s_{mt}^e - \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \mu \times [\gamma_m + \omega_m]^t s_m^\dagger. \end{aligned}$$

Given (a) this is satisfied at

$$s_{mt}^e = [\gamma_m + \omega_m]^t s_m^\dagger = s_{mt}^*,$$

from equation (14) in Theorem 3.9. Next, for $m \in M_{NRV}$,

$$\begin{aligned} q_{m,t+1}^e s_{m,t+1}^e &= q_{mt}^e s_{mt}^e - p_{mt}^e c_{mt}^* \\ &= q_{mt}^e s_{mt}^e - \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \mu \times [\gamma_m + \omega_m]^t s_m^\dagger. \end{aligned}$$

By (b) this is satisfied at

$$s_{mt}^e = [\gamma_m + \omega_m]^t s_m^\dagger = s_{mt}^*,$$

again using equation (14) in Theorem 3.9. For $m \in M_{BLu}$,

$$\begin{aligned} q_{m,t+1}^e s_{m,t+1}^e &= q_{mt}^e s_{mt}^e - (p_{mt}^e - q_{m,t+1}^e \hat{\gamma}_m) c_{mt}^* \\ &= q_{mt}^e s_{mt}^e - \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-t} U'_m(c_{m0}^*) [1 - \gamma_m^{-1} \hat{\gamma}_m] \mu c_{mt}^*. \end{aligned}$$

From (c) this becomes

$$\gamma_m^{-1} s_{m,t+1}^e = s_{mt}^e - [1 - \gamma_m^{-1} \hat{\gamma}_m] c_{mt}^*,$$

and is satisfied at

$$s_{mt}^e = (\gamma_m - \hat{\gamma}_m) \gamma_m^{-1} \sum_{j=0}^{\infty} \gamma_m^{-j} c_{m,t+j}^* = s_{mt}^*$$

by equation (17). Finally, for $m \in M_{BLV} \cup \{0\}$,

$$q_{m,t+1}^e s_{m,t+1}^e = q_{mt}^e s_{mt}^e - p_{mt}^e c_{mt}^*.$$

Using (d) this is

$$\frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} \gamma_m^{-1} s_{m,t+1}^e = \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} s_{mt}^e - c_{mt}^*,$$

or,

$$\gamma_m^{-1} s_{m,t+1}^e = s_{mt}^e - [1 - \gamma_m^{-1} \hat{\gamma}_m] c_{mt}^*,$$

and is satisfied at

$$s_{mt}^e = (\gamma_m - \hat{\gamma}_m) \gamma_m^{-1} \sum_{j=0}^{\infty} \gamma_m^{-j} c_{m,t+j}^* = s_{mt}^*,$$

again, by equation (17). Next, turn to firms. First, for $m \in M_{NRu} \cup M_{NRV}$, the firm only expands its production set by increasing the supply of c_m . Given (a), per unit revenue for $m \in M_{NRu}$ is

$$p_{mt}^e - q_{m,t+1}^e \omega_m = \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \mu > 0;$$

given (b), per unit revenue for $m \in M_{NRV}$ is

$$p_{mt}^e = \beta^t U'_m([\gamma_m + \omega_m]^t s_m^\dagger) \mu > 0.$$

Since revenue is positive in each case, firms $m \in M_{NRu} \cup M_{NRV}$ optimally set

$$c_{mt}^e = s_{mt}^e = [\gamma_m + \omega_m]^t s_m^\dagger = s_{mt}^*.$$

Next, for $m \in M_{BLu}$, using (c) the revenue stream present value is

$$\begin{aligned} &\sum_{t=0}^{\infty} (p_{mt}^e - q_{m,t+1}^e \hat{\gamma}_m) c_{mt} \\ &= \frac{\gamma_m}{\gamma_m - \hat{\gamma}_m} U'_m(c_{m0}^*) [1 - \gamma_m^{-1} \hat{\gamma}_m] \mu \sum_{t=0}^{\infty} \gamma_m^{-t} c_{mt}, \quad (49) \end{aligned}$$

while the consumption stream feasible from s_m^\dagger is restricted by

$$s_m^\dagger \geq \frac{\gamma_m - \hat{\gamma}_m}{\gamma_m} \sum_{t=0}^{\infty} \gamma_m^{-t} c_{mt}, \quad (50)$$

iterating on the constraint in problem (27). Take a specific sequence c_m satisfying restriction (50) with equality, thereby yielding a particular revenue stream present value (49). Every alternative sequence c_m that provides larger value (49) is infeasible in restriction (50). Conversely, if a choice for sequence c_m satisfies restriction (50) with strict inequality, then it fails to maximize value (49). Consequently, optimization problem (27) has a solution at $c_{mt}^e = c_{mt}^*$ and $s_{mt}^e = s_{mt}^*$. Now, for $m \in M_{BLV} \cup \{0\}$, using (d) the revenue stream present value is

$$\sum_{t=0}^{\infty} p_{mt}^e c_{mt} = U'_m(c_{m0}^*) \mu \sum_{t=0}^{\infty} \gamma_m^{-t} c_{mt},$$

where again the consumption stream feasible from s_m^\dagger is restricted by inequality (50). Then the argument following restriction (50) above applied here establishes that optimization problem (28) has solution at $c_{mt}^e = c_{mt}^*$ and $s_{mt}^e = s_{mt}^*$. Finally, the value function derivative relation (29) is immediate from equations (15) and (18) in Theorem 3.9 and (a)–(d). Q.E.D.

Proof of Lemma 4.7 By Prop. 3.6 and the hypotheses in the Lemma, function V_0 is increasing, concave, and continuously differentiable. [Recall V_0 need not be twice differentiable; otherwise, we could just apply Theorem 4.4 in Rockafellar (1970, p. 26)]. Necessarily $s_{00} < \mathcal{E}$. For any fixed \underline{s}, \bar{s} with $s_{00} \leq \underline{s} \leq \bar{s} \leq \mathcal{E}$, and $\lambda \in [0, 1]$, define $s_{(\lambda)} \stackrel{\text{def}}{=} \lambda \underline{s} + (1 - \lambda) \bar{s}$. Since

$$\begin{aligned} V_0(s_{(\lambda)}) &= V_0(\underline{s}) + \int_{\underline{s}}^{s_{(\lambda)}} V_0'(s) ds = V_0(\bar{s}) - \int_{s_{(\lambda)}}^{\bar{s}} V_0'(s) ds \\ \implies V_0(s_{(\lambda)}) &= \lambda V_0(\underline{s}) + (1 - \lambda) V_0(\bar{s}) \\ &\quad + \lambda \int_{\underline{s}}^{s_{(\lambda)}} V_0'(s) ds - (1 - \lambda) \int_{s_{(\lambda)}}^{\bar{s}} V_0'(s) ds, \end{aligned}$$

concavity implies

$$\lambda \int_{\underline{s}}^{s_{(\lambda)}} V_0'(s) ds \geq (1 - \lambda) \int_{s_{(\lambda)}}^{\bar{s}} V_0'(s) ds.$$

By Mean Value Theorem, the above can be rewritten

$$\lambda \left[V_0'(\underline{s}_{(\lambda)}) \times (\bar{s} - \underline{s})(1 - \lambda) \right] \geq (1 - \lambda) \left[V_0'(\bar{s}_{(\lambda)}) \times (\bar{s} - \underline{s})\lambda \right]$$

for some $\underline{s}_{(\lambda)} \in [\underline{s}, s_{(\lambda)}]$ and $\bar{s}_{(\lambda)} \in [s_{(\lambda)}, \bar{s}]$, which is just

$$V_0'(\underline{s}_{(\lambda)}) \geq V_0'(\bar{s}_{(\lambda)}).$$

Since $\underline{s} \leq \bar{s}$ but are otherwise arbitrary in $[s_{00}, \mathcal{E}]$, we then have

$$V_0'(s_{00}) \geq V_0'(s) \geq V_0'(\mathcal{E}) \geq 0 \quad \forall s \in [s_{00}, \mathcal{E}]. \quad (51)$$

Next, again by Mean Value Theorem, some $s_{(m)} \in [s_{00}, s_{00} + \Psi_m]$ exists such that

$$V_0(s_{00} + \Psi_m) = V_0(s_{00}) + V_0'(s_{(m)}) \times \Psi_m.$$

Therefore,

$$\begin{aligned} V_0(s_{00}) + V_0'(s_{(m)}) \Psi_m - V_0(s_{00} + \Psi_m) \\ &= [V_0'(s_{00}) - V_0'(s_{(m)})] \times \Psi_m \\ \implies |V_0(s_{00}) + V_0'(s_{(m)}) \Psi_m - V_0(s_{00} + \Psi_m)| \\ &= |V_0'(s_{00}) - V_0'(s_{(m)})| \times \Psi_m. \end{aligned}$$

By relation (51),

$$|V_0'(s_{00}) - V_0'(\mathcal{E})| \geq \sup_{m \in \mathbb{M}} |V_0'(s_{00}) - V_0'(s_{(m)})|,$$

so that

$$\begin{aligned} \sup_{m \in \mathbb{M}} |V_0(s_{00}) + V_0'(s_{(m)}) \Psi_m - V_0(s_{00} + \Psi_m)| \\ \leq \left[|V_0'(s_{00}) - V_0'(\mathcal{E})| \times \sup_{m \in \mathbb{M}} \Psi_m \right] \rightarrow 0 \quad \text{as } \mathcal{E} \rightarrow \infty. \end{aligned}$$

Finally, to complete the proof, take $\delta(\mathcal{E}) = \widehat{\delta}(\mathcal{E}) \times \sup_{m \in \mathbb{M}} \Psi_m$. Q.E.D.

Proof of Theorem 4.9 Define \mathcal{M}_n to comprise all elements of \mathcal{M} containing n . Then \mathcal{M}_n is a complete lattice with $\{n\} = \inf_{\mathcal{M}} \mathcal{M}_n \in \mathcal{M}_n$. For \mathcal{E} sufficiently large, Assumption 4.6 gives $V_0'(s_{00}) = V_0'(\mathcal{E})$ constant, so that equations (29)–(30) and Assumption 4.8 imply that $n \in \mathfrak{C}(M)$ for all $M \in \mathcal{M}_n$, i.e., \mathfrak{C} takes \mathcal{M}_n to \mathcal{M}_n , and that \mathfrak{C} is isotone on \mathcal{M}_n . In particular, $\{n\}$ is a fixed point of \mathfrak{C} on \mathcal{M}_n . Clearly, it remains that $\mathfrak{C}(M) \subseteq M$ for all $M \in \mathcal{M}_n$. The set $\mathcal{S} \stackrel{\text{def}}{=} \{M \in \mathcal{M}_n : M \subseteq \mathfrak{C}(M)\}$ contains $\{n\}$ and is, therefore, non-empty.

Moreover, $\{n\} = \inf_{\mathcal{M}_n} \mathcal{S} = \inf \mathcal{S}$. Let $\overline{M} \stackrel{\text{def}}{=} \sup_{\mathcal{M}_n} \mathcal{S}$. Since \mathfrak{C} is isotone and $M \subseteq \overline{M}$ for all $M \in \mathcal{S}$, necessarily $M \subseteq \mathfrak{C}(M) \subseteq \mathfrak{C}(\overline{M})$ for all $M \in \mathcal{S}$. Thus, $\mathfrak{C}(\overline{M})$ is an upper bound for \mathcal{S} which implies $\overline{M} = \sup_{\mathcal{M}_n} \mathcal{S} \subseteq \mathfrak{C}(\overline{M})$, since \overline{M} is the least upper bound. But then, $\overline{M} \in \mathcal{S}$ so that $\overline{M} = \sup_{\mathcal{M}_n} \mathcal{S} = \sup \mathcal{S}$. By construction, \mathfrak{C} has $\mathfrak{C}(M) \subseteq M$ for all $M \in \mathcal{M}_n$, so that \mathcal{S} is then also the collection of fixed points containing n , i.e.,

$$\mathcal{S} = \mathcal{M}_n^e \stackrel{\text{def}}{=} \{M \in \mathcal{M}_n : M = \mathfrak{C}(M)\} \subset \mathcal{M}^e,$$

and we have just shown that $\{n\}$ and \overline{M} are \mathcal{M}_n^e 's least and greatest elements, respectively. Repeating the argument above with (\mathcal{M}, \emptyset) replacing $(\mathcal{M}_n, \{n\})$ gives \emptyset and some $\sup \mathcal{M}^e$ as, respectively, the least and greatest elements of \mathcal{M}^e , with $\sup_{\mathcal{M}} \mathcal{M}^e \in \mathcal{M}^e$. Assumption 4.8 ensures $\{n\} \in \mathcal{M}^e$. Assumptions 4.6 and 4.8 guarantee that for any $M \in \mathcal{M}^e$ the augmented $M_n = M \cup \{n\} \supseteq M$ satisfies $M_n = \mathfrak{C}(M_n)$. Therefore, even though $\mathcal{M}^e \supsetneq \mathcal{M}_n^e$, necessarily $\sup \mathcal{M}^e \in \mathcal{M}_n$, giving $\sup \mathcal{M}^e = \sup \mathcal{M}_n^e = \overline{M}$. Therefore, $\overline{M}^e = \overline{M} = \mathfrak{C}(\overline{M}) \supset \{n\}$ and is consequently unique and positive, and by the sup definition has,

$$\overline{M}^e = \bigcup_{M \in \mathcal{M}^e} M.$$

Finally, substituting equation (29) into condition (30) and using Assumption 4.6 establishes condition (35). Q.E.D.

Theorem 4.9 contains parts of Tarski's Fixed-Point Theorem for isotone mappings on complete lattices [e.g., Birkhoff (1967, Theorem 11, p. 115 and Ex. 7, p. 116) or Vives (1999, Theorem 2.2, p. 20)]. Because \mathfrak{C} is not only isotone but has every argument containing the image, the relevant reasoning in the Proof above is marginally simpler than for the more general result. On the other hand, that Theorem 4.9 has a distinguished element n makes for reasoning marginally more elaborate than usual.

The Proof above does not establish that the collections of fixed points \mathcal{M}^e and \mathcal{M}_n^e each constitutes a complete lattice. That result

is not needed for the principal analysis in the text, but the argument to verify it is easily adapted from Vives (1999, Theorem 2.2, p. 20).

More generally, that the collection of fixed points containing n is a complete lattice guarantees that the union of all such fixed points is such a fixed point as well. This doesn't draw on whether those fixed points constitute a sub-lattice in \mathcal{M} ; in general, they won't. Instead, the result uses that in the lattice comprising all fixed points containing n , the set-union of all such fixed points is their least upper bound; and that, by Tarski's Fixed Point Theorem, that lattice is complete.

Proof of Theorem 4.10 By the maintained hypotheses, Theorem 3.9 gives V_m strictly concave and continuously differentiable. Then, for fixed $s_{m0} > 0$, defining

$$s_{(\lambda)} \stackrel{\text{def}}{=} \lambda \cdot 0 + (1 - \lambda) \cdot s_{m0}, \quad \text{all } \lambda \in (0, 1),$$

has

$$\begin{aligned} V_m(s_{(\lambda)}) &> \lambda V_m(0) + (1 - \lambda)V_m(s_{m0}) \\ \implies V_m(s_{m0}) - V_m(0) &> [V_m(s_{m0}) - V_m(s_{(\lambda)})] \lambda^{-1}. \end{aligned}$$

Taking $\lambda \rightarrow 0$ on the right side above, by continuous differentiability, gives

$$V_m(s_{m0}) - V_m(0) > V'_m(s_{m0}) \times s_{m0}. \quad (52)$$

At $s_{m0} = s_m^\dagger$, this yields:

$$V'_m(s_m^\dagger) \geq V'_0(s_{00}) \psi_m \implies V_m(s_m^\dagger) - V_m(0) > V'_0(s_{00}) \Psi_m,$$

so that $\overline{M}^e \subseteq M^*$, by equation (20) in Theorem 3.11 and equation (35) in Theorem 4.9. Inequality (52) at $s_{m0} = s_m^\dagger$ gives inequality (36) in the Theorem. If n satisfies condition (37), then $n \in M^*$ but $n \notin \overline{M}^e$. Q.E.D.

Proof of Theorem 4.12 (\Leftarrow) Inequality (52) in the Proof of Theorem 4.10 evaluated at $s_{m0} = \tilde{s}_{m0}$ implies condition (39) at $\tilde{s}_{m0} =$

$\tilde{s}_{m0} \in (0, s_m^\dagger]$ whenever condition (40) holds. (\Rightarrow) By Mean Value Theorem, there is $\tilde{s}_{m0} \in (0, \bar{s}_{m0})$ such that

$$V_m(\bar{s}_{m0}) - V_m(0) = \int_0^{\bar{s}_{m0}} V'_m(s) ds = V'_m(\tilde{s}_{m0}) \times \bar{s}_{m0},$$

so that condition (39) implies condition (40).

Q.E.D.