2010

Optimal Antitrust Enforcement, Dynamic Competition, and Changing Economic Conditions

Keith Hylton
Boston Univeristy School of Law

BOSTON
UNIVERSITY

OPTIMAL ANTITRUST ENFORCEMENT, DYNAMIC COMPETITION, AND CHANGING ECONOMIC CONDITIONS

Keith N. Hylton
Haizhen Lin*

The recent financial crisis and ensuing recession have touched off a round of introspection on antitrust policy. Federal antitrust enforcement officials referred to the crisis as a justification for a more aggressive stance, and suggested that the crisis and current recession are due in part to lax enforcement in previous administrations.[1] The financial sector itself has become more concentrated as a result of the crisis, as regulators have pushed and approved mergers in order to rescue distressed banks.[2] The rising unemployment rate, now standing just under 10 percent, provides another perspective on the argument for rethinking antitrust priorities.[3]

---

[1] This view was presented in a speech announcing the antitrust policies of the current administration. For the text of the speech, see Christine A. Varney, Ass't Att'y Gen., Antitrust Div., U.S. Dep't of Justice, Vigorous Antitrust Enforcement in This Challenging Era (May 12, 2009), *available at* http://www.justice.gov/atr/public/speeches/245777.htm.

[2] *See, e.g.*, Ian Guider & Louisa Fahy, *Roubini Says Bank Mergers May Create "Bigger Monster,"* Bloomberg.com, Nov. 5, 2009, http://www.bloomberg.com/apps/news?pid=206010 87&sid=ae.OQGynRAX4&pos=4; Andrea Gomes da Silva & Mark Sansom, *Antitrust Implications of the Financial Crisis: A UK and EU View*, Antitrust, Spring 2009, at 24.

[3] Specifically, the rising unemployment rate introduces a wealth distribution perspective, which has generally not been a major feature of antitrust policy. The traditional economic argument against monopolies is that they reduce social welfare by supplying fewer goods and services than would be observed in competitive markets. *See, e.g.*, Richard A. Posner, Antitrust Law 13–14 (2d ed. 2001). The distributional effect of monopolization has not been stressed in the economic literature. For an exception, see Robert H.

The questions to be addressed in any major rethink of antitrust policy are obvious. Should enforcement be extended in new directions? Should courts reconsider the tradeoff between false convictions and false acquittals reflected in existing law? Should the crisis be viewed as a justification for a temporary change in antitrust policy, or an opportunity to reconsider policy for all seasons?

Given the potentially infinite number of questions, we will not attempt a broad survey of the crisis's implications for antitrust policy. There are too many issues, and their implications too varied. Instead we will start with a narrow focus on enforcement principles and branch out to an examination of their implications for antitrust policy.

We will take this as an opportunity to reexamine the dynamic versus static cost tradeoff in antitrust policy[4] and its implications for varying enforcement with economic conditions. The dynamic versus static cost tradeoff is at the foundation of policy questions in antitrust, especially in monopolization law. But modern policy debates have given surprisingly little attention to the dynamic efficiency problem in antitrust—that is, the problem of designing antitrust law to encourage, or at least to avoid discouraging, incentives to invest in the creation and expansion of markets. Of course, this has always been recognized as a legitimate concern,[5] but it has also been waved off as a question too complicated and ambiguous to play a serious role in antitrust policy. As a result, dynamic incentives are mentioned, if at all, as an afterthought in most discussions of antitrust policy.

This is a deep and longstanding issue in antitrust theory. The Chicago School revolution took static efficiency from being a marginal concern and afterthought to a central issue in antitrust analysis. The same sort of intellectual revolution has yet to happen with respect to dynamic efficiency.[6] Our goal is to advance the case a bit further for incorporating

---

Lande, *Wealth Transfers as the Original and Primary Concern of Antitrust: The Efficiency Interpretation Challenged,* 34 HASTINGS L.J. 65 (1982).

[4] David S. Evans & Keith N. Hylton, *The Lawful Acquisition and Exercise of Monopoly Power and Its Implications for the Objectives of Antitrust,* COMPETITION POL'Y INT'L, Autumn 2008, at 203; J. Gregory Sidak & David J. Teece, *Dynamic Competition in Antitrust Law,* 5 J. COMPETITION L. & ECON. 581, 584 (2009); Michael L. Katz & Howard A. Shelanski, *"Schumpeterian" Competition and Antitrust Policy in High-Tech Markets,* COMPETITION, Fall/Winter 2005, at 47 (2005); Keith N. Hylton & Haizhen Lin, *Innovation and Optimal Punishment, with Antitrust Applications* (B.U. School of Law Working Paper No. 08-33, 2008), *available at* http://ssrn.com/abstract=1305147.

[5] The dynamic efficiency concern is typically attributed to Schumpeter. *See* JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM AND DEMOCRACY (1975).

[6] Sidak & Teece, *supra* note 4, at 584.

dynamic efficiency in antitrust analysis. Along the way we will study the lessons for examining antitrust policy in poor economic conditions.

The "dynamic enforcement model" examined here implies that antitrust enforcers should put a greater weight on the dynamic cost of enforcement during recessions. On the theory that a crisis opens the door to reassessing everything,[7] we examine implications for antitrust policy in all conditions as well. We argue below that taking dynamic efficiency into account leads to a positive theory of Section 2 doctrine and to a better understanding of the tradeoff between false conviction and false acquittal costs in monopolization and merger cases.

I. ECONOMIC PRINCIPLES OF ANTITRUST ENFORCEMENT

The economic principles of antitrust enforcement widely accepted today are largely those emphasized in the Chicago School approach to antitrust[8] and formalized in Gary Becker's examination of optimal law enforcement.[9] In Figure 1, we present a diagram that illustrates the key policy concerns in the antitrust setting.[10] For simplicity, we will assume that the entity of concern is a monopolizing firm, though the analysis that follows could also be applied to a cartel.

A. STATIC ENFORCEMENT MODEL

The firm has a choice whether to take a "monopolizing act." The act could be a decision to enter into an exclusivity contract or to tie one product to another. The monopolizing act allows the firm to increase its price, leading to a transfer (T in Figure 1) of consumer surplus to the firm. The price increase also leads to a reduction in output below the competitive level and associated loss in consumer welfare (D), which we will also refer to below as "destroyed surplus" or as "deadweight loss."

---

[7] Rahm Emanuel's comment that "You never want a serious crisis to go to waste," has become perhaps the most memorable quote of the Great Recession. For the quote and its context, see Gerald F. Seib, *In Crisis, Opportunity for Obama*, WALL ST. J., Nov. 21, 2008, at A2. This article reflects the spirit of opportunism reflected in Emanuel's remark. Crises invite specific short-term solutions, but they also invite total reassessments of past practices.

[8] For the original presentation of the Chicago School point of view, see Aaron Director & Edward H. Levi, *Law and the Future: Trade Regulation*, 51 Nw. U. L. REV. 281 (1956). *See also* Richard A. Posner, *The Chicago School of Antitrust Analysis*, 127 U. PA. L. REV. 925 (1979).

[9] Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

[10] For a discussion of the implications of the economics of enforcement for antitrust, see William M. Landes, *Optimal Sanctions for Antitrust Violations*, 50 U. CHI. L. REV. 652 (1983). *See also* Herbert Hovenkamp, *Antitrust's Protected Classes*, 88 MICH. L. REV. 1, 28 (1989).

After the monopolizing act, consumers are left with the residual surplus (*W* in the diagram).

The firm's monopolizing act may have efficiency consequences. For example, an exclusive dealing contract with a key input supplier could have a monopolizing effect by excluding rival firms from access to the supplier, but it could also enhance efficiency by reducing supply costs. The efficiency gain is shown in Figure 1 by *E*. The diagram assumes that the efficiency gain appears in the form of reduction of average cost from c_0 to c_1. The new cost curve c_1 is shown with a dotted line because we are assuming that the efficiency gain is a random event that may or may not materialize.

The Chicago School is given credit for having made efficiency a central concern of antitrust analysis. Its most important lesson is that because of the potential efficiencies associated with many acts that might have a monopolizing effect, society should not set out to prohibit such acts. To do so might reduce society's wealth, both in the short run and in the long run. Specifically, the firm's monopolizing act may generate gains (to firm owners and consumers) that exceed the short-run consumer harm caused by supracompetitive pricing. In Figure 1, the efficiency gain *E* could be greater than the deadweight loss *D*. If the efficiency gain is greater than the deadweight loss, then the firm's monopolizing act would enhance social welfare, which is the sum of welfare going to consumers and to producers.

The distribution of the wealth generated by the market is not taken into account in this argument. An efficiency gain in the form of a cost reduction is a benefit to the owners of the monopolizing firm, and not to consumers, at least in the short run. An approach to antitrust that equates social welfare exclusively with short-run consumer welfare would hold that an act that reduces consumer welfare should be condemned, even if it leads to a greater gain in producer welfare. We will not enter this debate.[11] We will assume, consistent with the Chicago School approach, that total welfare is the appropriate objective function for antitrust.

Given that the goal of antitrust enforcement should *not* be to prohibit, or completely deter, all monopolizing acts, what is the appropriate goal? The alternative to prohibition is a policy of internalizing to the mono-

---

[11] On the debate, see Lande, *supra* note 3; Joseph Farrell & Michael L. Katz, *The Economics of Welfare Standards in Antitrust*, COMPETITION POL'Y INT'L, Autumn 2006, at 3; Ken Heyer, *Welfare Standards and Merger Analysis: Why Not the Best?*, COMPETITION POL'Y INT'L, Autumn 2006, at 29; Russell Pittman, *Consumer Surplus as the Appropriate Standard for Antitrust Enforcement*, COMPETITION POL'Y INT'L, Autumn 2007, at 205.
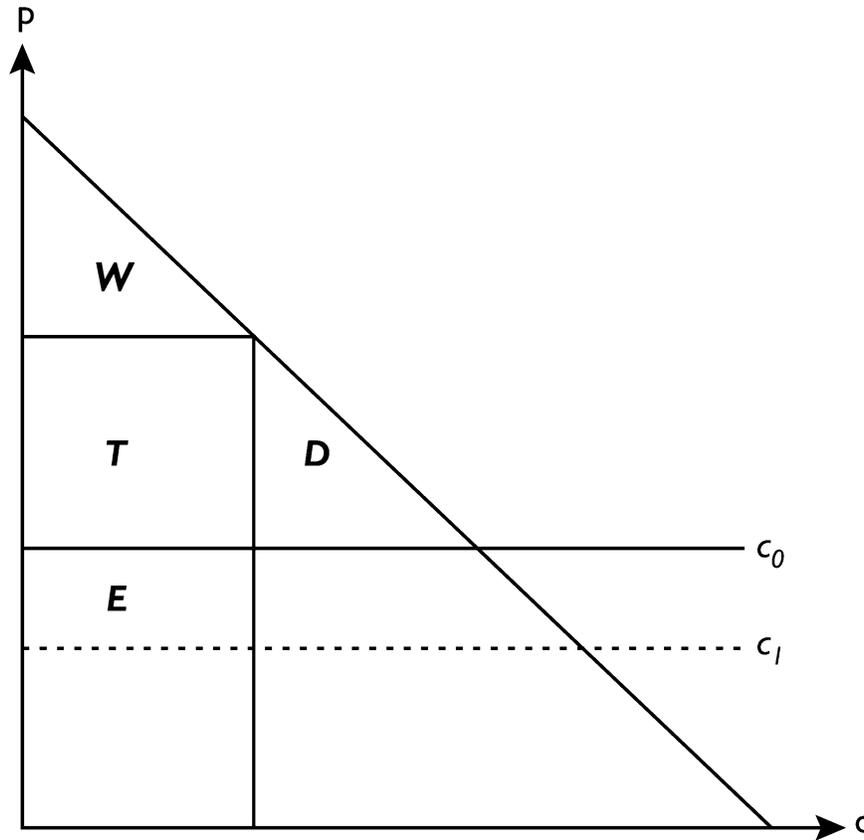
FIGURE 1: WELFARE CONSEQUENCES OF MONOPOLIZING ACT
THAT ALSO REDUCES COSTS

polizing firm the social costs of its conduct.[12] Under the internalization
approach, the firm would choose to take a monopolizing act when and
only when the gain to the firm exceeds the loss to consumers. Efficient
conduct would not be prohibited. The internalization rule generates a
simple recommendation for the optimal monetary penalty: if enforce-

---

[12] Becker, *supra* note 9. Internalization results in "first-best" deterrence—in the sense
that the monopolization decision will be made when and only when it increases social
welfare. We are equating first-best deterrence with optimal deterrence, but the two can be
distinguished in some settings. For example, if enforcement agents have discretion over
whether to bring an action in court or in an administrative proceeding, an optimal en-
forcement regime might discourage costly types of litigation or weak claims. For a discus-
sion of the difference between first-best and optimal deterrence in the private
enforcement (litigation) setting, see Keith N. Hylton & Thomas J. Miceli, *Should Tort Dam-
ages be Multiplied?*, 21 J.L. ECON & ORG. 388, 410 (2005).

ment is perfect and costless, the penalty should be set equal to the sum of the transfer from consumers and the forgone consumer surplus (deadweight loss). Referring to Figure 1, the optimal punishment scheme will require the imposition of a fine equal to $T+D$ on the monopolizing firm. Faced with such a fine, the firm will take the monopolizing act only when its gain is greater than the consumer harm.

If the enforcement authority is unlikely to detect and bring an enforcement action in every instance of a monopolizing act, the optimal penalty will include a multiplier. This is based on the simple principle that if a bad actor gets away with offenses half of the time, his punishment will have to be doubled in the times that he is caught in order to preserve a given level of deterrence. In addition, if enforcement is costly, the enforcement cost should be internalized to the monopolizing firm, since enforcement is a natural byproduct of its conduct.[13] If the probability of enforcement is P, and the enforcement cost is C, the optimal antitrust penalty is then[14]

$$Static\ Penalty = \frac{T + D}{P} + C.$$

We refer to this as the optimal static penalty because our description of the enforcement problem does not incorporate any consideration of the long-term incentive effects of antitrust enforcement.

Another variation on the optimal static penalty is the case of a firm that takes a monopolizing act that also enhances the quality of the firm's product. For example, the firm could introduce a new product that sets an industry standard that has an exclusionary effect on rivals.[15] The firm's act reduces consumer surplus by leading to a surplus transfer and the imposition of deadweight loss. However, it also increases surplus by enhancing the value of the firm's product, or all products that meet the new standard. In this case, the optimal punishment scheme will reduce

---

[13] The assumption that enforcement is a natural byproduct of an offense simplifies matters, but it not necessarily valid. Suppose the enforcement agency decides each case by comparing the gain from enforcement to its cost. In this case, an optimal scheme might shift the enforcement cost to the agency in order to generate efficient enforcement decisions.

[14] *See* Landes, *supra* note 10; KEITH N. HYLTON, ANTITRUST LAW: ECONOMIC THEORY AND COMMON LAW EVOLUTION 43–52 (2003) [hereinafter ANTITRUST LAW].

[15] *See* Steven C. Salop, *Exclusionary Conduct, Effect on Consumers, and the Flawed Profit-Sacrifice Standard,* 73 ANTITRUST L.J. 311, 345–46 (2006).

the fine by the amount of the new surplus generated by the product's enhancement.[16]

The static model just considered has been criticized for its softening effect on antitrust enforcement.[17] When efficiency is incorporated into the analysis, the arguments for limiting the size of penalties and for discretionary enforcement become stronger, especially in the area of monopolization. Many of the acts that have been held anticompetitive in the monopolization case law involve conduct that might introduce efficiencies or might be efficiency based. In *United States v. Aluminum Co. of America*,[18] Judge Learned Hand found Alcoa in violation of Section 2 because it preemptively expanded capacity,[19] occupying new markets before rivals had a chance to gain a significant presence. But the expansion of capacity in anticipation of new markets is conduct that many business managers would view as concrete evidence of intelligent foresight. Anticipatory capacity expansion may exclude rivals by deterring entry, but it may also be necessary to efficiently meet consumer needs.

The role of the enforcement probability also has implications for the quality of antitrust enforcement. Price fixing is difficult to detect, as participants try to hide their agreements from the public. Monopolization, however, generally is not difficult to detect. The most common monopolizing acts (e.g., exclusive dealing, tying, mergers, cutting price below cost)[20] are carried out in broad daylight and even trumpeted to the public. Even when the monopolizing efforts are concealed, the eventual result if successful is obvious to everyone and will invite a challenge from its victims. The optimal enforcement model suggests that a trebling of damages is probably too small a multiplier for collusion cases, and probably excessive for the typical monopolization case. Since the probability of detection and enforcement are high in the monopolization setting, the case for multiplying monopolization damages as a general rule is unclear.

To be sure, not every case of monopolization will lead to an enforcement action. Monopolizing acts may be observed and best understood

----

[16] For a discussion of this point in the context of optimal enforcement in the patent-antitrust setting, see Keith N. Hylton & Sungjoon Cho, *The Economics of Injunctive and Reverse Settlements*, 12 AM. L. & ECON. REV. 181 (2010).

[17] *See, e.g.*, Hovenkamp, *supra* note 10.

[18] 148 F.2d 416 (2d Cir. 1945).

[19] *Id.* at 431.

[20] We do not mean to suggest these acts serve only a monopolizing purpose—of course, they are often normal, competitive business practices. This short list of practices is based on the types of conduct repeatedly challenged in antitrust litigation.

by rival firms and contracting parties.[21] These may be constrained from petitioning enforcement agencies or bringing lawsuits in several ways. Complaining to an enforcement agency or bringing a lawsuit may damage long-term business relationships. Alternatively, the cost of litigation or even of participating as a third party in government enforcement action, in terms of financial and managerial resources, may be too high in light of the private return. These are all possible constraints that might reduce the likelihood of enforcement in the monopolization area.

This model is for the most part suggestive because it treats enforcement as an exclusively public sector activity. This is also true of Becker's analysis of the economics of law enforcement. When private actions are modeled carefully, one finds a close link between the probability of a private action and the profitability of a lawsuit.[22] If the multiplier is set at a level that induces all victims to bring suit, the probability of an enforcement action will be 100 percent. But once the probability of a private enforcement action reaches 100 percent, there will no longer be a need to multiply damages. It follows that the optimal multiplier for private lawsuits efficiently balances the supply of lawsuits with the "demand" required by the optimal deterrence goal.[23]

In a more detailed or disaggregated model the probability of enforcement would be broken down into the product of three components: the probability of detection, the probability of an enforcement action, and the probability of liability. The simple model discussed here assumes that these are all assessed on the assumption that a violation has occurred. Courts are assumed to be perfectly accurate and baseless or frivolous enforcement actions are not incorporated. If baseless or frivolous actions are incorporated, along with the assumption of judicial error, it may not be desirable on optimal deterrence grounds to determine the penalty by multiplying the total consumer harm by the reciprocal of the probability of enforcement.[24]

B. Dynamic Enforcement Model

Although controversial for its enforcement implications, the Chicago School antitrust policy fails to incorporate dynamic considerations. By dynamic, we refer to the effects of enforcement on incentives in previ-

---

[21] Joe Brodley made this point in the context of his assessment of merger disputes. *See* Joseph F. Brodley, *Antitrust Standing in Private Merger Cases: Reconciling Private Incentives and Public Enforcement Goals*, 94 Mich. L. Rev. 1 (1996).

[22] Hylton & Miceli, *supra* note 12, at 410.

[23] *Id.*

[24] *See, e.g.*, Hylton, Antitrust Law, *supra* note 14, at 66–67.

ous time periods to invest in activities. Return to the simple model shown in Figure 1. Suppose there are two time periods. In the first, the firm decides whether to invest in an activity that generates the market. In the second, the firm decides whether to take the monopolizing act.

Suppose the firm designs and produces a new, superior type of artificial tooth in the first period.[25] Suppose also that the firm cannot get a patent on the design and that the tooth is easily replicable. Facing the risk of immediate competition from firms that copy its design, the firm may decide to take an action that excludes rivals for some period of time. Suppose, for example, the firm enters into exclusivity contracts with the most important downstream sellers of dental products. The exclusivity contracts will not exclude competition forever, nor for a period as long as a patent, but the contracts will exclude competition long enough for the innovating firm to earn a supracompetitive return on its investment in the artificial tooth design.

In this dynamic story, what happens to consumers when the firm creates the market and then takes an act that excludes competition? As in the static story, some surplus is transferred to the firm (T) and some surplus is destroyed (D). However, unlike the static story, the firm's conduct has also rewarded consumers with the residual consumer surplus that remains after the monopolizing conduct (W). If not for the firm's first-period investment, which was undertaken because of anticipation of profits generated from second-period exclusionary conduct, consumers never would have received the residual surplus.

The optimal antitrust penalty has to be designed to reconcile conflicting welfare concerns. There is the static welfare concern addressed earlier: the monopolizing firm should be forced to regurgitate the transfer and to pay for the destroyed surplus D in order to optimally regulate (in the first-best sense) its incentive to monopolize. However, the penalty will also affect ex ante investment incentives. In order to optimally regulate investment incentives, the ideal penalty would be negative; it is a subsidy equal to the residual surplus. This is because the private benefit of the firm's investment is simply the transfer (T). The social benefit of its investment is the sum of the transfer and residual surplus ($T+W$). In order to align private incentives to invest with social incentives, the punishment authority should award the monopolizing firm with a bounty equal to the residual surplus. These conflicting welfare concerns are not easy to reconcile.

---

[25] This example, based partially on *United States v. Dentsply International, Inc.*, 399 F.3d 181 (3d Cir. 2005), is taken from Evans & Hylton, *supra* note 4.

We assume that the monopolizing firm cannot engage in price discrimination. If the firm implements perfect price discrimination in the monopolization stage, charging each consumer the maximum that he is willing to pay, there will be no economic basis for imposing a penalty, or for providing a subsidy. The perfectly discriminating monopolist will not destroy any surplus; and, given this, there will be no need to impose the static penalty in order to regulate the monopolization incentive. And since the perfectly discriminating monopolist will not externalize any surplus that it generates from innovation to consumers, there will be no need to provide a subsidy in order to optimally regulate the investment incentive. An alternative scenario to consider is where the firm can discriminate only enough to capture the residual surplus. In this case the optimal dynamic penalty will be the same as the static penalty. It will still be socially desirable in this case to regulate the monopolization incentive, but not the investment decision.

In our study of the dynamic enforcement model,[26] we found that the optimal penalty is a weighted average of the static penalty and a subsidy based on the residual surplus. Specifically, the optimal antitrust penalty in the dynamic setting is of the form

$$\textit{Dynamic Penalty} = (1-\theta)\left(\frac{T+D}{P}\right) + \theta\left(\frac{-W}{P}\right) + C,$$

where the weighting parameter θ, itself an increasing function of the penalty, varies with the relative responsiveness of the firm's monopolization and investment incentives to changes in the penalty. If a change in the penalty would have no effect on ex ante investment, while discouraging the monopolizing act, θ will be close to zero, and the dynamic penalty will be roughly the same as the static penalty. This might be observed if the firm's discount rate is so high that a change in the penalty has little effect on ex ante investment incentives. If the change in the penalty has a big impact on ex ante investment, the optimal penalty probably will be negative—specifically, a subsidy based on the residual consumer surplus.

The sign and size of the optimal antitrust penalty depend on several factors. If the expected enforcement cost is greater than the residual surplus ($PC > W$), then the optimal dynamic penalty is always positive.[27] This is unlikely to be observed in general. It would be an enormously inefficient administrative process that resulted in enforcement costs that always exceeded the residual surplus generated by a new product. Alter-

---

[26] Hylton & Lin, *supra* note 4.
[27] *Id.* at Appendix, Case 1.

natively, the case in which the expected enforcement cost exceeds the residual surplus can be viewed as encompassing the instances in which the residual surplus from innovation is trivial, another scenario that should be regarded as a special case.

If the expected enforcement cost is less than the residual surplus ($PC < W$), the optimal dynamic penalty could be a penalty or a subsidy depending on the parameters. If the residual surplus is sufficiently large relative to the consumer harm, the optimal penalty is always a subsidy equal to the residual surplus divided by the enforcement probability less the enforcement cost ($W/P - C$).[28] When the residual surplus is below that level, the optimal penalty could be either a positive weighted average of the static penalty and the residual-surplus subsidy (as in the formula above), or just the residual-surplus subsidy.

We present the results of a model simulation in Figure 2. The simulation assumes the parameter values are such that the optimal dynamic penalty is positive—i.e., it is not a subsidy.[29] The probability of enforcement is assumed to be less than 1, specifically 0.7; the cost of enforcement is set at 2, which is the minimum value for the static penalty.[30] Even in this case, the dynamic penalty is substantially less than the static penalty and the implicit multiplier on consumer harm is less than 1. Unlike the static penalty, the dynamic penalty converges toward a fixed amount as the total consumer harm grows. The ceiling on the dynamic penalty represents the limit at which investment is excessively discouraged by any further penalty increase.

II. IMPLICATIONS

The purpose of the model summarized in the previous part is to incorporate dynamic efficiency considerations directly into the economic analysis of antitrust enforcement. This is in contrast to the traditional approach, which relies on a static analysis and then adds dynamic efficiency considerations on at the end as footnotes, afterthoughts, and qualifications. Incorporating the dynamic efficiency issue at the outset forces the analyst to design and evaluate basic features of antitrust enforcement policy with dynamic cost considerations in mind.

---

[28] *Id.* at Appendix, Case 2.

[29] This biases our simulation in the direction of larger (dynamic) penalties. If we examined a wider array of parameter values, the divergence between the optimal static and dynamic penalties would be greater than observed in Figure 2.

[30] Recall that the static penalty is the sum of two terms: (1) the consumer harm divided by the enforcement probability and (2) the enforcement cost. Thus, for a given enforcement probability, as the consumer harm approaches zero, the optimal static penalty approaches the enforcement cost.
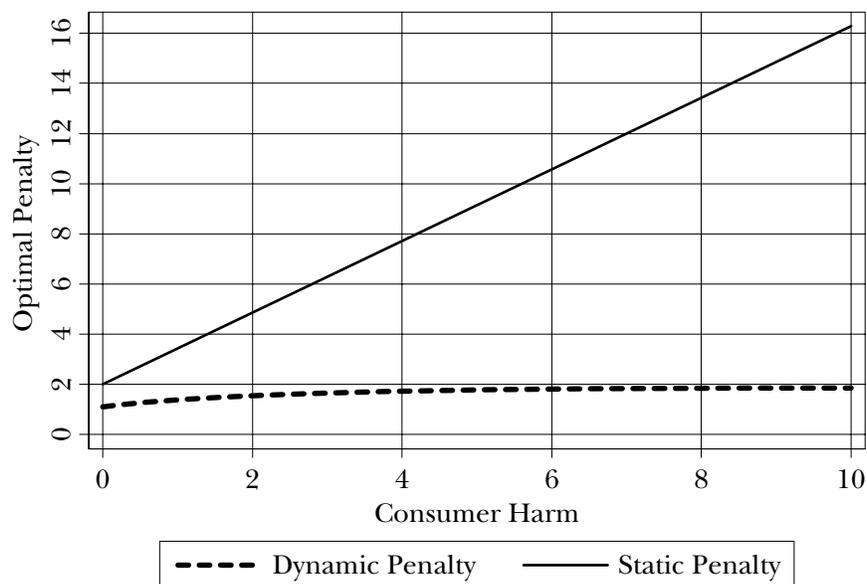
FIGURE 2: SIMULATION OF OPTIMAL PENALTY IN DYNAMIC
AND STATIC ENFORCEMENT MODELS

The dynamic enforcement model has implications for antitrust policy. We will first discuss the implications for the interpretation and reform of existing law, and then consider implications for changing economic conditions.

A. The Regulatory Function of Profits

The investment that we refer to in the model above could take the form of any decision to sink resources into an activity that creates or expands a market. The typical example is the development of a new product. But there is no reason to limit the application of this model to the creation of new products. Any investment that enhances the value of a product or service within an existing market could also fit within the model.[31] If the new source of value that is created by costly investment is

---

[31] Admittedly, this is a partial equilibrium analysis. Suppose we examined a model with two markets. An innovation in one might reduce demand in the other market, destroying surplus in the other market. Still, if consumers leave one market (the static one) to enter the innovating market, they presumably would do so only if there was a net increase in surplus for each one of them. Total surplus should therefore rise following innovation. There may be a special case in which net surplus falls, but that is likely to be a peculiar case.

easily replicable by rivals, the firm that invests initially will have an incentive to limit follow-on competition in order to secure a normal return on the investment.

Within this model, profit, rather than being a direct transfer from consumer welfare as in the static economic model, serves a regulatory function. Profit is the reward for market-creating and market-expanding investment, and the expectation of profit is the signal that generates such investment.

In the standard economic analysis of the market, profit serves as a signal for entry and exit. This is a socially productive role, but its value is diminished in antitrust analysis by the assumption that entry is blocked by the action of the monopolizing firm. With profit unable to serve as a signal that actually induces entry, the conclusion follows in the standard antitrust treatment that any profit existing beyond the short run serves no socially desirable purpose beyond the provision of a normal return on existing resources devoted to production and selling.

When profit is recognized to serve the socially productive role of encouraging investment that creates or expands markets, the argument for eliminating all opportunities for firms to charge a price that exceeds average avoidable cost is no longer as persuasive as in the static framework. Of course, this does not lead to the conclusion that all efforts to exclude rivals should be treated as per se lawful under the antitrust laws. However, the dynamic model suggests that practices that might enable a firm to exclude rivals, such as tying and exclusive dealing, should not be regarded as presumptively socially undesirable even when there is no offsetting static efficiency, such as a reduction in selling costs.

Indeed, in the special scenario where the entire anticipated profit from exclusion is necessary to induce the firm to invest in innovation ex ante, the static costs from exclusionary conduct are largely illusionary. The forgone surplus from supracompetitive pricing would never have been enjoyed by consumers if the firm did not invest in the initial time period.[32] Consumers suffer no costs from exclusion. They gain to the extent of the residual surplus.

---

[32] This has implications for the importance of the welfare standard used to assess anticompetitive conduct; specifically, the total harm versus consumer harm debate. We noted earlier that our optimal enforcement prescriptions are based on a total welfare objective. But even if we adopted (short-run) consumer welfare as the appropriate maximand, there would still be some constraint on the total fine that would be optimal to impose on an antitrust violator. The implications of our enforcement model for antitrust analysis may be weakened to some degree but they do not become irrelevant once the objective function is switched from total welfare to consumer welfare.

B. UNDERSTANDING SECTION 2 DOCTRINE

One longstanding paradox of Section 2 doctrine was at the core of Judge Learned Hand's decision in *Alcoa*. Hand noted that cartels are per se illegal and then argued that a monopoly is equivalent in effect to a cartel.[33] From this analogy it followed that the exercise of monopoly power should be treated as an offense under the Sherman Act.[34] The paradox created by the law is that the mere exercise or exploitation of monopoly power is perfectly legal under American law. What explains this paradox?

Evans and Hylton argue that the safe harbor created in American law for monopoly exploitation reflects the recognition by courts that Section 2 should not be enforced so broadly as to discourage innovation—innovation understood capaciously to encompass the creation of new markets and expansion of existing ones.[35] Under this view of the law, the paradox described by Judge Hand is not a paradox at all. The courts are unable to fashion precise case-specific penalties of the sort envisioned in the optimal enforcement model. If the courts were able to do so, they would award subsidies in some antitrust cases involving innovation antecedent to monopolization. But given the inability (or, more precisely, lack of technical capacity) of courts to design optimal case-specific penalties, they have shaped antitrust doctrine in a way that effectively shields firms from antitrust liability in a substantial set of the instances in which innovation is antecedent to monopolization. The law prohibits a narrow set of monopolizing acts, while leaving firms a great deal of room to adopt competitive strategies—whether advertising, product differentiation, protection of trade secrets, acquisition of patents, packaging, pricing, allocation of supply—that have the effect of excluding rivals.[36] Indeed, although the scope of monopolization law has appeared to wax and wane over various time periods,[37] the most expansive theories of exclusionary conduct, associated with the school of industrial organization scholarship initiated by Joe Bain,[38] have never gained a serious foothold in American monopolization law.

---

[33] United States v. Aluminum Co. of Am., 148 F.2d 416, 427–28 (2d Cir. 1945).

[34] *Id.* at 428.

[35] Evans & Hylton, *supra* note 4.

[36] *Id.* at 211–12.

[37] *See, e.g.*, HYLTON, ANTITRUST LAW, *supra* note 14, at 186–229.

[38] JOE S. BAIN, BARRIERS TO NEW COMPETITION: THEIR CHARACTER AND CONSEQUENCES IN MANUFACTURING INDUSTRIES (1956).

The static enforcement model is incapable of providing a positive theory of this basic feature of monopolization law.[39] Under the static enforcement model, firms that exploit monopoly power gained through luck should be penalized to the same degree as those that exploit monopoly power gained through exclusionary conduct.[40]

To say that the dynamic enforcement model provides a positive theory for the exploitation safe harbor in existing Section 2 case law is not to say that the law is optimal in its current state. The safe harbor is surrounded by soft edges rather than a hard shell, and the current trend in the enforcement agencies is to press in at the edges of the shell. Moreover, there is no way to tell on a priori grounds whether it would be optimal to broaden the scope of the safe harbor. The approach we take to policy is to suggest that cases involving innovation, considered broadly, would be ideal candidates to be swept into the safe harbors of Section 2 case law.

C. DAMAGES

The dynamic enforcement model suggests that treble damages for Section 2 violations may be socially undesirable. The static enforcement model, as we noted, probably implies this already. Given that monopolizing conduct is seldom conducted in secrecy,[41] the probability of detection and enforcement is likely to be high in most serious cases of exclusionary conduct. Even within the static enforcement model, the argument for applying a multiplier to penalties based on consumer harm would not be uncontestable in monopolization cases. If, as seems plausible, the joint detection and enforcement probability is high in monopo-

---

[39] Evans & Hylton, *supra* note 4, at 220–24. The exclusion versus exploitation distinction appears to be an international norm. Although EU antitrust law formally prohibits monopoly exploitation, in practice the European competition authorities seldom bring prosecutions against cases of exploitation. *Id.* at 221–24.

[40] Obviously, one could argue that the law is indifferent to exploitation and is really concerned only with exclusion. But this leads to the question, why bother? Why should the law punish exclusion when it is not followed by exploitation? In actual practice, every monopolization case involves a claim of exploitation. Courts justify punishment on the ground that exploitation hurts consumers and destroys surplus, even when it merely transfers wealth between firms.

[41] Of course, there are counterexamples. A firm might obtain a patent on fraudulent grounds. In this case, one might say that the monopolizing act (the acquisition of the patent) was carried out in secrecy. But even in this example, the final result of the patent—the monopoly—and its immediate cause (the patent) are both open and obvious to everyone. If the patent monopoly is substantial, rival firms would have incentives to challenge it. Those firms would have incentives to invest resources into discovering the fraud, or at least challenging the monopoly. In general, almost every example of monopolization involves conduct that is open and obvious in its effect, and in this sense is different from the collusion case. We thank Maurice Stucke for suggesting this counterexample.

lization cases, the optimal static penalty is roughly equal to the sum of the transfer and deadweight loss. If the treble damages remedy is roughly three times the transfer, then the treble damages remedy probably will be greater than the optimal static penalty.

Of course, this argument assumes that detection is the key factor causing variability in the probability of private enforcement efforts. If detection is not the key factor, and instead it is weak incentives for private enforcement actions, then the argument for a damages multiplier in monopolization cases becomes stronger. But this raises the question why private enforcement incentives would be weak, especially when rivals (excluded firms) can easily complain to competition authorities or sue under a statutory framework that provides attorney's fees to plaintiffs.

The dynamic model adds another reason to question the general desirability of trebling in Section 2 cases. In the innovation antecedent to monopolization scenario, there are two levels for the optimal dynamic penalty: a pure subsidy equal to the residual surplus multiplied by the reciprocal of the enforcement probability, and a weighted average of the static penalty and the residual surplus subsidy (both multiplied by the reciprocal of the enforcement probability). No matter which form is correct for a particular case, the optimal dynamic penalty is less than the optimal static penalty. If the probability of enforcement is high in monopolization cases, then the optimal dynamic penalty, because it is less than the static penalty, will also be less than the treble damages measure.

It should be clear that this is a default approach we are suggesting. The model incorporates the probability of enforcement as a parameter that can take different values. If the probability of enforcement for some particular set of monopolization law violations is unambiguously low, then the enforcement model could generate a corresponding dynamic fine that is greater than the treble damages measure. But this is not necessarily so. Unlike the static enforcement setting, the dynamic penalty may actually decline as the probability of enforcement falls.[42] And

---

[42] *See* Hylton & Lin, *supra* note 4. First, the optimal responses of the dynamic penalty to changes in key parameters are more complicated than in the static model. A small reduction in the probability of enforcement could increase the optimal dynamic penalty, the opposite of the static enforcement model. The reason is that a small reduction in the enforcement probability alters the sensitivity of investment to the penalty. This imparts a type of "substitution effect" (in contrast to the "scale effect" of diluting enforcement), which may result (paradoxically) in a reduction in the optimal penalty level. For large changes in the probability of enforcement, the results are also more complicated. For example, as the enforcement probability goes to zero, the model's recommendation shifts from the scenario in which a positive penalty is optimal to that in which a negative penalty (the subsidy) may be optimal. The reason is that as the enforcement probability falls, the

there are additional complications that could be brought into the model that would limit the optimal size of any multiplier based on the probability of enforcement.[43] The most important complication of this sort is the possibility of false convictions and the existence of claims (baseless or frivolous) that take advantage of such errors.

D. Expanded Safe Harbors and No-Fly Zones

Of course, the courts do not attempt to calculate optimal case-specific penalties and subsidies, as envisioned in the model. Instead of subsidies, the general approach of courts is to create porous exemptions or safe harbors in the case law. The most important safe harbor is that for mere exploitation of monopoly power. The exploitation safe harbor can be viewed as a rough approximation of the dynamic enforcement model's implications. Firms are not penalized for merely charging the monopoly price, but then they are not given subsidies for instances in which the residual surplus from their market investments is large relative to the consumer harm resulting from exploitation.

The optimal enforcement model provides a guideline for safe harbors and interpretation rules that move the law toward exempting, on dynamic efficiency grounds, some instances of alleged monopolization. For example, the model provides a justification for the decision in *United States v. Jerrold Electronics Corp.*,[44] where the court applied the rule of reason to Jerrold's tying policy during the period in which the policy was thought to be necessary for the successful launch of its product (television antennae systems) in an uncertain market environment. The result is equivalent to a safe harbor; the rule of reason applied to tying permits a court to take efficiencies into account, which often results in a finding of legality.

A broader application of the model is possible to market power issues. In order to find a firm guilty of monopolization, a court must first determine that the firm has market power. Market power is understood as the ability to raise price without the constraint of competition from existing firms or from new entrants. Although easy to state, the market power

_____

ratio of enforcement costs to the residual surplus falls too. If the residual surplus is large relative to the consumer harm, the subsidy may become the optimal enforcement outcome after a large change in the enforcement probability. These technical questions are addressed in Hylton & Lin, *supra* note 4.

[43] One factor is the type of test that is used to determine legality. If it is a rule of reason type test, which is similar in form to a negligence test, the social desirability of multiplying damages is ambiguous, see Richard Craswell, *Deterrence and Damages: The Multiplier Principle and Its Alternatives*, 97 Mich. L. Rev. 2185 (1999).

[44] 187 F. Supp. 545 (E.D. Pa. 1960), *aff'd per curiam*, 365 U.S. 567 (1961).

test inevitably involves a complicated assessment of facts and the exercise of discretion, especially with respect to burdens of proof. In short, market power is a legal construct that permits courts to exempt some cases from penalization for a large set of reasons that may not be entirely explicable in terms of short-term price-setting ability.

This model suggests that in the innovation antecedent to monopolization setting, market power should be used more expansively as a screen to exempt some firms from liability under the Sherman Act. In a case where the residual surplus from market investment is large relative to the consumer harm, the dynamic enforcement model implies that a subsidy, rather than a fine, may be the socially desirable response from the enforcement authority. Given the inability of courts to award subsidies, the appropriate practical response from a court in such a case is immunization through the market power test. This implies that in the innovation antecedent to monopolization setting, where the residual surplus is likely to be large relative to consumer harm, courts should demand more of the evidence offered to prove the existence of market power, and err on the side of the broader market definition with more substitutes.

This argument can be approached in terms of the ratio of false conviction costs to false acquittal costs.[45] The dynamic efficiency concern, when taken into account, raises the ratio of the false conviction to the false acquittal cost. This implies that the burden of proof that plaintiffs have to meet on the monopoly power issue should be increased.

Consider, again, the artificial tooth hypothetical. In a monopolization case, the first issue that would confront a court is whether to treat the innovating firm as having monopoly power. Suppose the new artificial tooth is superior to existing types. The court has a choice whether to define the market as one based narrowly on the specific features of the new artificial tooth, or on the general features of many varieties of artificial tooth. If it chooses the former definition, the defendant has a market share of 90 percent; if the latter, the defendant's market share is only 10 percent. Choosing the latter (broad) market definition effectively exempts the innovating firm from Section 2 liability. The dynamic enforcement model suggests that where the residual surplus from inno-

---

[45] We are referring to the error cost approach to antitrust. *See* Frank H. Easterbrook, *The Limits of Antitrust*, 63 TEX. L. REV. 1 (1984); Keith N. Hylton & Michael Salinger, *Tying Law and Policy: A Decision-Theoretic Approach*, 69 ANTITRUST L.J. 469 (2005); James C. Cooper, Luke M. Froeb, Dan O'Brien & Michael G. Vita, *Vertical Antitrust Policy as a Problem of Inference*, 23 INT'L J. INDUS. ORG. 639, 639–64 (2005).

vation is large relative to the consumer harm, a court should be biased toward the broad market definition.

The lessons of this model can be used to examine the summary judgment decision for the defendant, Del Monte, in a recent case involving pineapples.[46] After nearly a decade and a half of research and development, Del Monte created a new variety of extra-sweet pineapple called the MD-2.[47] A monopolization class-action lawsuit was filed against Del Monte for efforts that it took to discourage rival firms from producing and marketing the same variety. In an analysis of the market power issue, the court noted that the plaintiffs' expert had excluded from the relevant market other varieties of sweet pineapple, effectively treating the MD-2 as a single market.[48] The court reviewed the arguments of the plaintiffs' expert and granted summary judgment in favor of Del Monte. The court's opinion on the summary judgment motion is a careful examination of the plaintiffs' expert's arguments on the market power question. However, the analysis at bottom is a judgment on the burden and standard of proof. The submarket doctrine of *Brown Shoe*,[49] analyzed by the court, involves several discretionary decisions that a court has to make on the sufficiency of evidence. The *Del Monte* court's opinion reduces to the simple statement that the plaintiffs did not bring enough evidence on the question of market power.

Under the analysis of this article, a demanding proof standard on market power is precisely what courts should apply in the innovation antecedent to monopolization setting. Such an approach strikes the appropriate balance from a decision theory perspective because it treats false convictions as having a relatively high cost. Setting a high standard of proof effectively creates a zone of safety from antitrust liability for the firm that creates the very market it is accused of monopolizing.

The court noted that the evidence suggested that Del Monte had a legitimate business justification for its conduct, and that Del Monte's efforts to discourage rivals did not appear to have had a significant effect in delaying their efforts to enter the MD-2 market. If these conclusions are correct, then the pineapple dispute is best viewed as a case of innovation without any real exclusion. A penalty applied to such a case probably would do more to discourage innovation than monopolization, because it would reward erroneous enforcement and frivolous litigation.

---

[46] *In re* Fresh Del Monte Pineapples Antitrust Litig., No. 04-1628 (S.D.N.Y. Sept. 30, 2009).

[47] *Id.* at 6.

[48] *Id.* at 10–22.

[49] Brown Shoe Co. v. United States, 370 U.S. 294 (1962).

Assuming the residual surplus from innovation was substantial, Del Monte would have received a subsidy, rather than a penalty, under the dynamic enforcement model. A summary judgment is not the same as a reward, but it is the best that an antitrust court can do in this setting.

E. MERGERS

A merger is another way in which a firm can enhance its power over price. In a horizontal merger that results in a combined entity with market power, the combined entity will increase its price above the competitive level, resulting in short-run harm to consumers. In a vertical merger that enables the acquiring firm to exclude competitors, the combined entity will raise its price, with the same impact.

As a preliminary matter, one might argue that the innovation model as described earlier in this paper is not applicable to the merger setting. Mergers are often determined by several "big picture" factors and are unlikely to be connected to the desire to secure a return on some specific product or service innovation. However, where the innovation is sufficiently substantial to create a new market, the innovating firm would have an incentive to consider how the long-term profits from innovation will be affected by the market's structure. Merger, under these conditions, would appear to be a plausible option for protecting or appropriating the rents from innovation. The availability of the merger option could then affect the incentive to invest in innovation.

Discussions of dynamic efficiencies in the merger context generally focus on the prospect that innovation will generate new rivals to the merging parties in the immediate future, dampening their ability to hike prices after the merger.[50] The merger of satellite radio firms XM and Sirius was permitted by the U.S. Department of Justice on the basis of this argument.[51] When innovation can generate new competitors, dynamic competition provides a justification for choosing a broader, rather than narrow market as the relevant market for analysis. However, in the applications suggested by the dynamic enforcement model presented here, the pricing power gained by a merger is viewed as having ambiguous implications in the innovation setting. Consumers lose surplus to the extent the post-merger price exceeds the competitive level. But they gain surplus to the extent that anticipated post-merger price increases support investment in innovation.

---

[50] Sidak & Teece, *supra* note 4, at 623.

[51] *Id.*

The dynamic enforcement model implies that in the innovation ante-cedent to merger setting, a relaxation of liability and regulatory prohibitions might be socially desirable. To implement this suggestion, merger enforcement authorities would have to distinguish antecedent innovation mergers from other mergers involving no antecedent innovation. Where the residual surplus to consumers is likely to be large relative to the potential consumer harm, the merger enforcement authority should err toward the broader market definition, as in the monopolization analysis discussed above.[52]

Within a model, it is easy for courts to distinguish antecedent innovation cases from others. In the real world, it would be difficult to separate the two types. Every firm proposing a merger would attempt to characterize it as an antecedent innovation case. Still, the objective evidence should be sufficient to identify mergers that have substantial implications for innovation.

Consider *FTC v. Staples*[53] from the perspective of this model. The court had a choice between the broad market of office supply outlets and the narrow market of office supply superstores. Within the broad market, the proposed merger between Staples and Office Depot would represent roughly 5 percent of office supply sales.[54] Within the narrow market, the combined entity would represent at least 45 percent of sales.[55] Staples had pioneered the high-volume office supply superstore. Although the time between the initial innovation and the proposed merger may appear to take this case outside of the innovation-antecedent-to-merger framework, the principles of this analysis still apply. We are not arguing that a firm that has earned its expected return from innovation within a reasonably long time period should still be permitted to act in a manner that lessens competition in order to sustain a high return. The problem, however, is uncertainty; when enforcers do not know whether the firm has or has not earned its expected return, they (as well as reviewing courts) should at least take into consideration the investment implications of blocking a merger.

---

[52] This is distinguishable from the argument for regulatory forbearance offered by Sidak and Teece. *Id.* Their analysis of dynamic competition considers markets in which potential rivals have the capacity to constrain the merging firms in the immediate future. Our argument applies to the case in which no such potential competition exists.

[53] 970 F. Supp. 1066 (D.D.C. 1997).

[54] To be precise, 5.5 percent in 1996. *Id.* at 1073.

[55] The court referred to the twenty-seven metropolitan areas where the merger would lead to a reduction from three office supply superstores to two. Within this set of geographic markets, the post-merger share of the combined entity would range from 45 percent to 94 percent. *Id.* at 1081.

Suppose the purpose of the merger was to enhance profitability through lessening competition in the submarket of office supply superstores, as the FTC charged. The FTC's policy, which prevented Staples from using the acquisition to recapture profit levels that had been eroded by entry into the submarket, discourages the very innovation that created the office supply superstore market. Forward-looking firms, realizing that merger as a cheap form of exit would be effectively prohibited by the policy, would demand a higher rate of return in order to develop a superstore submarket. If entry into the submarket is not difficult, the stream of future dynamic welfare gains forgone could easily be greater than the stream of future static welfare losses.[56]

Admittedly, this example suggests virtually all mergers may use this argument since virtually all of them involve some kind of innovation in product or process. But the fact that the argument could be used by nearly all mergers is not, by itself, a reason to reject it. The preferable route would incorporate the innovation defense in a manner that enables it to be used in the cases in which the costs of "false convictions" (rejecting a socially desirable merger) are likely to be high and the costs of "false acquittals" (permitting a socially undesirable merger) are low. To avoid overly broad and exaggerated claims of innovation that agencies are not well positioned to evaluate, the defense could be limited to specific products and time periods.[57]

Plenty of industries have been characterized at times as needing consolidation because of weak profits. The argument for consolidation is unpersuasive, on social welfare grounds, simply as a justification for higher profits and nothing more. But the argument is typically based on some feature of the industry that reflects a failure to efficiently meet consumer preferences. For example, the relatively poor condition of the banking sectors in the United States and in Germany has been attributed to market fragmentation, in addition to government support.[58]

---

[56] The present value of dynamic efficiency gains would be determined by the residual surplus component *W* (see Figure 1 *supra*). The present value of static costs would be determined, in the short run, by the consumer harm *T+D*. As entry occurred and the market became more competitive, the consumer harm would fall to zero and *W* would expand by the amount of the original consumer harm.

[57] Similarly, the dynamic story presented in this article could create difficulties in the enforcement context when a firm attempts to constrain competition in a different product than the one for which innovation investments are made. These difficulties suggest that an innovation-based defense (e.g., as observed in *Jerrold Electronics*) might appropriately be limited in scope to a single product for a limited time period.

[58] *The Sick Banking System of Europe*, ECONOMIST, May 9, 2009, at 77. The number of credit institutions per 10,000 population is substantially higher in the United States and in Germany than in other western countries. *Id.* This is distinguishable from the too big to fail problem, which is more accurately viewed as a problem of political connectedness

With little to lose in terms of a secure profit base, these banks invested in the riskiest assets available (e.g., subprime mortgages). The U.S. airline industry regularly earns negative profits. In a consolidated airline industry, firms would devote more resources to meeting consumer preferences, including safety. Even Japan's famously innovative mobile phone industry has been described as suffering from excessive fragmentation that limits the flow of resources to the most innovative and productive firms.[59]

Unless all of the reports of excessive industrial fragmentation are to be regarded as nonsense, antitrust policy should attempt to reconcile itself with this set of facts. When profit serves as a signal for investment, mergers for the sole purpose of gaining pricing power cannot be regarded as presumptively socially undesirable. If too much pricing power is attained, entry will occur, pushing prices back down.[60] The ordinary process of entry and exit will regulate profits to a level sufficient to compensate investments in market expanding products and services.

On a more general level, there is an open question about the fit between merger policy and general antitrust doctrine under Section 2. Given monopolization law's distinction between exploitation and exclusion, the policy against mergers for market power seems difficult to explain. Mergers toward monopoly are efforts to exploit market power. They sometimes include a risk of exclusion in the future, but it is by no means clear that a merger should be prohibited simply because of a risk that the merged entity may later exclude rivals. Let the law operate on the merged entity at the moment it attempts to exclude a rival.

The possibility that a merger toward monopoly could result in exclusionary acts in the future could, in theory, reduce the innovation incentives of firms that compete against the merging firms. This argument is based on two levels of prediction or conjecture: first that the merging firm will exclude rivals after the merger, and second, that those rivals would have invested in innovation if not for the threat of exclusion in the future. If the innovation disincentive created by a merger is suffi-

---

within the largest U.S. banking institutions. *See* Lawrence J. White, Financial Regulation and the Current Crisis: A Guide for the Antitrust Community (June 9, 2009), *available at* http://ssrn.com/abstract=1434454. Although the news coverage of banking failures has focused on large banks, the U.S. banking industry is fragmented in comparison to most other national markets. Most failures and regulatory shut-downs have involved banks that are not in the "too big to fail" category.

[59] *No Exit: What Japan Needs Is More Bankruptcies, Not Fewer*, ECONOMIST, June 20, 2009, at 16.

[60] We have assumed that the pricing power created by such mergers does not result from any restrictions on new entry. As long as there are no entry barriers, new firms can erode the pricing power created by a merger.

ciently large, the optimal enforcement policy, taking dynamic efficiencies into account, may tighten, rather than relax, regulatory constraints. However, this argument is often weakened by its speculative nature and the dearth of evidence supporting it. Moreover, it is difficult to distinguish rivals who object to a merger because they fear the effects of increased future competition from rivals whose objections are based on conjectural theories of exclusion. The arguments offered by the two groups will often be based on the same set of facts. Conjectural theories of exclusion and displaced innovation are relatively easy to state, and difficult to prove or to disprove. In the absence of concrete evidence, a position of skepticism toward conjectural theories of displaced innovation seems warranted.[61]

F. Changing Economic Conditions

The dynamic enforcement model directs us to examine (1) the ratio of the residual surplus to the consumer harm and (2) the relative elasticities of investment and monopolization with respect to the penalty. The optimal dynamic penalty varies with these factors. The relationship between economic conditions and changes in the penalty, as implied in this model, is an empirical matter.

As wealth declines, there is no unambiguously clear direction for the ratio of the residual surplus to the consumer harm from supracompetitive pricing. On one hand, consumers are likely to consider a wider range of substitutes as wealth declines, which would cause demand curves to become more elastic. With less pricing power the ratio of residual surplus to consumer harm from supra-competitive pricing probably increases. However, demand for some items may increase as consumer wealth declines: cigarette sales expanded dramatically during the Great Depression, increasing the pricing power of cigarette firms.[62]

In spite of the lack of an unambiguous answer that applies to every market, the likely general result is that the residual surplus increases relative to consumer harm during severe recessions. The Great Depres-

---

[61] *See generally* William Baumol & Janusz A. Ordover, *Use of Antitrust to Subvert Competition*, 28 J.L. Econ. 247 (1985).

[62] *See, e.g.*, Michael A. Bernstein, The Great Depression: Delayed Recovery and Economic Change in America 1929–1939 71 (1989) ("The cigarette division of the tobacco industry fared well in the depression. After a slight drop in 1932, cigarette output increased steadily."). Bernstein cites contemporaneous studies of the tobacco industry. In contrast to this account, Scherer, in the third edition of his textbook on industrial organization, argues that the profits earned by the cigarette companies were due to collusion in a period of falling demand. F.M. Scherer & David Ross, Industrial Market Structure and Economic Performance 250–51 (3d ed. 1990).

sion was associated with a decline in prices, and recessions have generally been associated with deflation. Pricing power falls in bad economic times. Given this, the general direction of the ratio of the residual surplus to the consumer harm is likely to be up during recessions, even though each component falls absolutely.

If the ratio of residual surplus to consumer harm is countercyclical, then the instances in which a subsidy will be optimal, rather than a penalty, will be greater during recessionary periods. In sectors in which firms invest in creating or expanding markets, the optimal enforcement model implies an expansion of safe harbors.[63]

The other factor to consider in the optimal enforcement approach is the elasticities of innovation and monopolization with respect to the penalty. As economic conditions worsen, will an increase in the penalty have a greater impact on the incentive to set a supracompetitive price or on the incentive to invest in market innovation? The incentive to set a supracompetitive price is already weak in recessions. The payoff from innovation is also weaker than in boom times. However, the alternative opportunities to an innovator also decline during recessions. The incentive to invest in innovation is likely to be relatively constant because both the payoff and the opportunity cost of innovation decline during recessions. If this is correct, the innovation incentive will be more responsive than is the monopolization incentive to a change in the penalty during recessions, and this implies a shift in the dynamic enforcement model toward the subsidy.

The direction of policy during recessions is ultimately an empirical question. The main contribution of a model such as this is to offer preliminary answers and to direct the analysis toward a narrow set of issues.

G. OTHER CONSIDERATIONS

Our discussion of mergers can be applied to collusion. Does the dynamic enforcement model imply that collusion should be permitted under the law?

The first thing to say about this is that the dynamic enforcement model does not imply a blanket exemption to collusive practices. This is a model of antitrust enforcement. Following the tradition of the litera-

---

[63] Since some degree of cyclical variation in demand must be assumed at the innovation stage, the downturns we are discussing should be assumed to be very serious recessions that are outside of the parameters that people typically anticipate in business. For recessions that are within the parameters typically anticipated, there is no clear innovation-related reason to modify enforcement.

ture on the economics of law enforcement, we have examined a model in which the enforcement authority applies a monetary assessment to all apprehended violators of the law. The novel feature in the dynamic enforcement model, in contrast to the static model, is that the assessment may vary between a positive penalty and a subsidy.

It is not entirely obvious that the dynamic model applies to the case of collusion. In the model we presented, one firm innovates and then seeks to exclude rivals in order to secure a positive return on the innovation. In the collusion setting this description would have to be modified. Suppose one firm innovates and then seeks to get others to collude in order to secure a positive return to the innovator. This is an implausible story of collusion. The rational strategy on the part of the non-innovating firms is to promise in the first period to collude during the second period, in order to encourage innovation, and then cheat when the second period arrives. Few instances of stable collusion would occur under this description.

An alternative description of dynamic competition in the collusion context would be a scenario in which all of the firms in an industry innovate and then attempt to collude in order to secure a positive return on the innovation. Cheating might still occur, but the knowledge that cheating erodes the incentive to invest might be a deterrent, especially in a repeat play setting. This is a plausible story of collusion and it appears to fit with the dynamic enforcement model. Moreover, the description is consistent with that of the airline industry, where negative profits have arguably discouraged firms from making investments in new technologies. This model may imply the social desirability of a safe harbor for collusion in the airline industry, and other industries suffering under similar conditions. The safe harbor would be more productive if applied during demand downturns.

However, creating safe harbors for collusion, although a possible implication of the dynamic enforcement model, creates real-world difficulties that the model as described does not address. One difficulty is the probability of detection. In the monopolization applications discussed previously, the probability of detection is high. In the collusion setting, detection probabilities are low, precisely because firms that engage in socially unproductive collusion (in contrast to the antecedent innovation setting) will hide it from the public. Once a safe harbor is recognized in the law, many firms may rush into collusion on the assumption that if they are caught they can persuade courts or regulators that they deserve to be within the safe harbor. This perception would reduce the deterrent effect of enforcement. A workable alternative might be a bi-

furcated system in which the general prohibition remains, but some firms are permitted to openly petition for the right to collude.

Another real-world difficulty with collusion is that of political influence. Once a safe harbor is opened in the law for collusion, industries will pressure legislators for a larger opening. The blanket prohibition, though inconsistent with the dynamic enforcement model, may be preferable in light of the rent-seeking costs that would be generated by a safe harbor.

Although we have limited ourselves to incorporating the dynamic effect on the monopoly side, there is quite possibly a dynamic effect on the side of the firms that are excluded by the monopolizing firm. A monopolizing firm may reduce the incentive of rivals to innovate by excluding them from a substantial market. The FTC's recent case against Intel is based in part on the theory that Intel has discouraged innovation by its competitors. Of course, one important distinction between the dynamic effects on the monopoly side and on the excluded-firm side is the difference between seeing and believing. A monopoly that has gained its status through innovation has demonstrated the link between its innovation and the reward from monopoly status. A reduction in the reward is likely to discourage future investment in the monopoly and in other firms that promise a similar sort of expansion through innovation. On the other hand, when rival firms complain that they would have innovated if not for the exclusionary acts of a dominant firm, there is always the question of validity lurking behind the claims. After all, plaintiffs have strong financial incentives to assert such claims. For this reason, claims that innovation would have taken place in some counterfactual scenario should be discounted in the absence of strong evidence.

It is not difficult to modify the dynamic antitrust enforcement model to incorporate the innovation effects on victims from exclusionary acts. If such effects are included, they would appear in the model as external costs that should be internalized to the monopolizing firm. But this would not change the core implications of the model.

III. CONCLUSION

Crises and recessions are good times to rethink antitrust policy. But a rethink is likely to bring out useful implications only if it deals with fundamental questions of policy. One fundamental question largely ignored in the modern literature is that of dynamic competition—or the static versus dynamic efficiency tradeoff in antitrust. When the dynamic competition perspective is brought into the theory of optimal antitrust enforcement, some substantial changes in enforcement policy recom-

mendations emerge. These changes apply to all seasons and they have implications for bad economic times too. The dynamic perspective implies that investment and innovation incentives, viewed broadly, should be taken into account in the assessment of liability under the Sherman Act.

APPENDIX:
SIMULATION OF DYNAMIC ENFORCEMENT
MODEL PENALTY

Here we set out some basic features of the model (whose results are summarized in the text) and the simulation shown in Figure 2. Let P = probability of apprehension. Let C = the cost to the state of apprehending the offender, $C > 0$; F = the fine imposed on apprehended offender; M = the gross gain to the offender from committing an offense, $M > 0$. M is governed by the probability distribution function H with corresponding density function h. The offender will commit the offense (monopolization) if $M > PF$; therefore, the probability that the offender does not commit an offense is $H(PF)$. The monopolization offense occurs with probability $1-H(PF)$ causing a loss of $T+D$; enforcement occurs with probability P; the offender is apprehended at cost C, and then punished with a fine equal to F. The potential offender invests in the first period, at cost k_o. The potential offender will invest if the expected return from monopolization, net of the penalty, exceeds his investment cost. Let the investment cost k_o be governed by the probability distribution Ψ with corresponding density ψ. The potential offender invests when $k_o < \bar{k}_o = (1-H(PF))[E(M \mid M > PF) - PF]$. The probability of investment is therefore $\Psi(\bar{k}_o)$. The optimal antitrust penalty is[64]

$$F = (1-\theta)\left(\frac{T+D}{P}\right) + \theta\left(\frac{-W}{P}\right) + C$$

where the weighting parameter is[65]

$$\theta = \frac{\psi(\bar{k}_o)(1-H(PF))}{\Psi(\bar{k}_o)h(PF) + \psi(\bar{k}_o)(1-H(PF))^2}.$$

If we assume H and Ψ are exponential functions with parameters λ^H and λ^Ψ respectively, then

---

[64] The optimal penalty maximizes the sum of the gain from investment and consumer surplus. *See* Hylton & Lin, *supra* note 4, at equation (2).

[65] *See id.* at equation (5) (for derivation).

$$\theta = \frac{\psi(\overline{k}_o)}{\lambda_H \Psi(\overline{k}_o) + \psi(\overline{k}_o)(1 - H(PF))}$$

$$= \frac{\lambda_\Psi \exp(-\lambda_\psi \overline{k}_o)}{\lambda_H(1 - \exp(-\lambda_\psi \overline{k}_o)) + \lambda_\Psi \exp(-\lambda_\psi \overline{k}_o)\exp(-\lambda_H PF)}$$

where $\overline{k}_o = \exp(-\lambda_H PF)/\lambda_H$.

To simulate the model, as shown in Figure 2, we used the following parameter values: $W=1$, $C=2$, $\lambda^H = 0.6$, and $\lambda^\Psi = 0.8$, $P = 0.7$.