Pace Law Review

Volume 34
Issue 2 *Spring 2014*

Article 8

April 2014

When Bigger Is Better: A Critique of the Herfindahl-Hirschman Index's Use to Evaluate Mergers in Network Industries

Toby Roberts

When Bigger Is Better: A Critique of the Herfindahl-Hirschman Index's Use to Evaluate Mergers in Network Industries

Toby Roberts*

I. Introduction

The Herfindahl-Hirschman Index ("HHI") operates under a very simple premise: industry behavior strongly correlates with industry structure; the larger a firm is within its industry, the more likely it is to engage in supracompetitive pricing or other anticompetitive conduct.[1] For more than 30 years, antitrust regulators have used the index to gauge whether prospective mergers would produce a firm of such magnitude that it would adversely impact societal welfare. When an HHI analysis of an impending merger suggests that a potentially harmful increase in concentration will result, the companies involved must demonstrate that the merger has other characteristics that mitigate its impact on prices in order to gain regulatory approval.[2]

* Staff attorney at the California Court of Appeal and former law clerk at the United States District Court for the Central District of California and the United States Court of Appeals for the Ninth Circuit. The views expressed in this Article are solely those of the author and do not reflect those of any court or judge. The author wishes to thank Professor Daniel Rubinfeld for his extremely helpful comments on an early draft of this Article and Dr. George Radics for encouraging its publication.

1. *See generally* Dennis C. Mueller, *Lessons from the United States's Antitrust History*, 14 INT'L J. INDUS. ORG. 415 (1996) (discussing the history of economic analysis in antitrust and citing the wide body of literature correlating structure with behavior). This paper will focus solely on the consumer welfare effects from increases in post-merger price.

2. For example, in many industries a post-merger firm will have lower unit costs due to economies of scale, which, in the absence of increased market power, tend to decrease prices. The Merger

One result of reliance on the HHI has been an erroneous conflation of market power (as proxied by market share) with consumer disutility. Although the close relationship between the two holds up in general, one significant exception arises in network industries. Network firms benefit consumers commensurately with their size. For example, given the choice between two equally priced credit cards, one accepted by merchants nationwide and the other by only half, most people would prefer to own the card with greater acceptance. Many people would willingly pay at least somewhat higher interest rates or greater fees for access to the larger network because the benefit of owning a universally accepted credit card outweighs the additional cost. But what if the larger network didn't exist? What if only smaller credit card companies existed? And what if two of those smaller firms desired to merge but could not offer the government a compelling reason why they would not subsequently exercise their increased market power to raise prices?

This Article argues that the current framework used by the Department of Justice ("DOJ") and Federal Trade Commission ("FTC") to evaluate mergers is inadequate in that it fails to account for network benefits. In particular, I argue for abandoning the use of the HHI in analyzing network industry mergers because the index generates little useful information about these mergers' effect on consumer welfare. Part II describes the HHI's historical and theoretical underpinnings and its integration into the current Merger Guidelines. Part III considers general objections to the HHI before turning to its problems in evaluating network industries. Part IV presents a formal model for evaluating the effects of mergers in network industries. Part V proposes an alternative framework for merger analysis to account for network effects. Part VI concludes.

Guidelines allow for such "efficiencies" to offset the anticompetitive concerns posed by large-scale consolidation. *See infra* Part II.B.

II. History of the Herfindahl-Hirschman Index

A. *The Search for an Optimal Measure of Concentration*

1. Early Measures of Concentration: The Golden Age of Graphs[3]

Statistical indexes of industrial concentration grew out of attempts by economists and statisticians in the early twentieth century to measure income distribution and inequality.[4] Theorists agreed that a society in which all members enjoyed equal incomes exhibited no income concentration; likewise, a state of affairs in which all income accrued to just one person would constitute the highest concentration possible.[5] The difficulty lay in developing a statistic that would meaningfully describe levels of inequality that fell between these antipodean states.[6]

Initial efforts proved fruitless because the models exhibited sensitivity to absolute income levels, rendering them useless for international or intertemporal comparisons.[7] In 1905, Max Otto Lorenz published a seminal paper on the subject of wealth

---

3. *See* JUDY L. KLEIN, STATISTICAL VISIONS IN TIME: A HISTORY OF TIME SERIES ANALYSIS, 1662–1938 17 (1997) (describing the late nineteenth and early twentieth centuries as "the golden stage for graphs," bridging periods of predominantly tabular and algebraic presentation of statistical data).

4. *See* ALBERT O. HIRSCHMAN, NATIONAL POWER AND THE STRUCTURE OF FOREIGN TRADE 157 (1945).

5. *See* Michael Schneider, Measuring Inequality: The Origins of the Lorenz Curve and the Gini Coefficient 6-8 (2000) (unpublished manuscript), *available at* http://www.latrobe.edu.au/data/assets/pdf_file/0011/130889/2004.01.pdf.

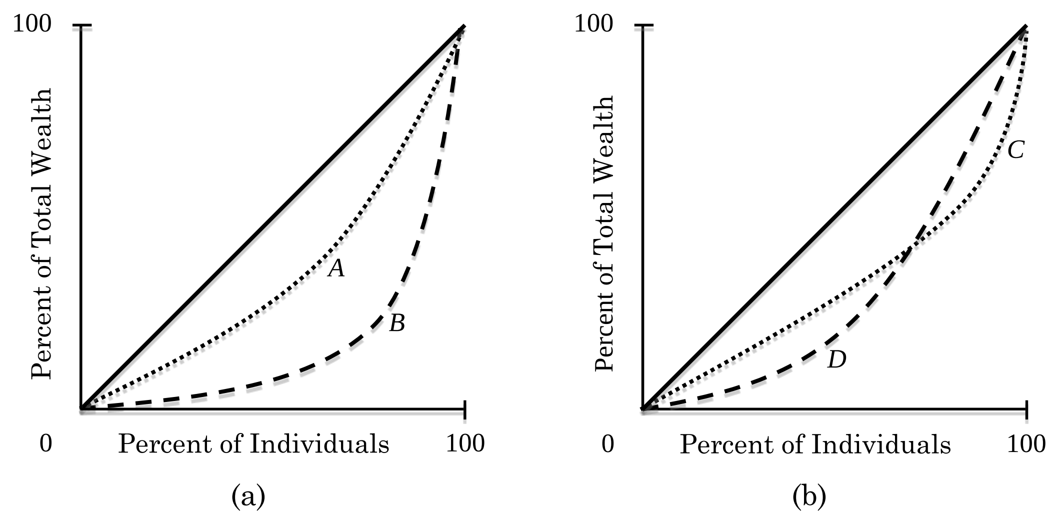
6. *See id.*

7. For instance, in 1895 Vilfredo Pareto observed a linear logarithmic relationship between income level, x, and the number of persons above a given income level, n: $\log n = \log A - \alpha \log x$ (where A and α are constants). Pareto proposed using α as a measure of concentration. His model fell prey to criticism, which asserted not only that α was sensitive to absolute income levels irrespective of concentration, but that α was insensitive to changes in income by the highest earning individual or group. *See id.* at 6.

concentration.[8] Lorenz proposed graphing the percent of individuals on the horizontal axis, arranged from poorest to richest, and their cumulative wealth on the vertical axis.[9] By using percentages rather than absolute wealth levels, Lorenz avoided the pitfalls that had ensnared his contemporaries.

The Lorenz curve, as the procedure became known, is illustrated in Figure 1(a) for two hypothetical countries. Country *A* has a more egalitarian wealth distribution than country *B*, given that curve *A* lies entirely above curve *B*, except at the endpoints. The straight line connecting the two endpoints—the equal distribution line—represents a situation of zero concentration.[10]

Figure 1.



8. M.O. Lorenz, *Methods of Measuring the Concentration of Wealth*, 9 PUBL'NS AM. STAT. ASS'N 209 (1905).

9. *See id.* at 216-19. Lorenz was inconsistent about which axis should depict "Percent of Number" and which should depict "Percent of Total Wealth." Gini followed his textual description rather than his graphs, as do I. *Compare id.* at 217-18, *with* Corrado Gini, *Sulla Misura della Concentrazione e della Variabilità dei Caratteri*, 73 ATTI DEL REALE ISTITUTO VENETO DI SCIENZE, LETTERE ED ARTI 1203, 1229 (1914).

10. More formally, on the line of equal distribution, x percent of the population holds $y = x$ percent of the wealth.

The situation illustrated by Figure 1(b) presents ambiguity: both countries *C* and *D* exhibit some degree of inequality in wealth distribution, but the Lorenz curves do not clearly show which country has the greater concentration of wealth. At the point where the curves cross, the richer and poorer cohorts in country *C* hold, respectively, the same percentages of total assets as in country *D*. In country *C*, there is less variance in wealth among the poorer cohort of the population and more disparity among the richer cohort relative to country *D*.

Italian statisticians, in particular Corrado Gini, devised a general method for comparing two or more Lorenz curves.[11] Gini had been working independently on a measure of inequality.[12] The traditional statistical measure of inequality—variance—focuses on the dispersion from a population's arithmetic mean. Gini came to believe that, in certain social science contexts, the appropriate measure of inequality should consider not "how much [] diverse outlying quantities differ from their arithmetic mean[,]" but rather "how much [] diverse actual magnitudes differ [from] each other."[13] He proposed an index of inequality that measured the "mean difference" in a population, obtained by taking every possible pair of observations, recording the absolute difference in value for each pair, and computing the average difference across all of the pairs.[14]

Gini demonstrated that the "mean difference" is equal to the area between the Lorenz curve and the line of equal distribution.[15] This relationship allowed a comparison between any two Lorenz curves—whichever curve had the greater area between it and the line of equal distribution represented the greater degree of inequality. Gini's coefficient of concentration,

11. *See* HIRSCHMAN, *supra* note 4, at 157-58 n.2.

12. *See* Schneider, *supra* note 5, at 11, 15 n.16.

13. *Id.* at 12 (citation omitted).

14. *See* HIRSCHMAN, *supra* note 4, at 157-58 n.2. In a population of *n* observations where each observation has a value of x_i ($i \in 1, 2, \ldots, n$), the mean difference is equal to:

$$\frac{2}{n(n-1)}\sum_{i=0}^{n-1}\sum_{j=1}^{n-1} x_{n-i} - x_{n-j}$$

15. *See* Gini, *supra* note 9, at 1229-33.

still in use today, took the area bounded by the Lorenz curve and line of equal distribution, and divided it by the entire triangular area underneath the line of equal distribution.[16]

As a measure of concentration, the Lorenz-Gini measure works well when the resource under measurement is allocated over a sufficiently large number of observations, such as in the distribution of wealth or income across a society. As applied to industrial concentration, however, the Gini Index fails to provide crucial information about the number of firms in the industry. Figure 2 illustrates the problem.

Figure 2.

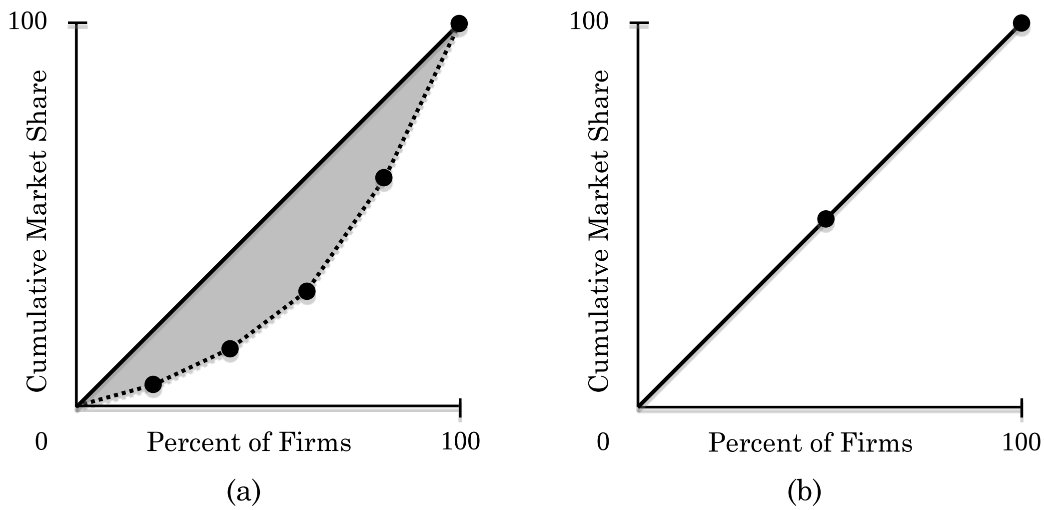


Figure 2(a) depicts the Lorenz curve for a five-firm industry with market shares of 5%, 10%, 15%, 30%, and 40%. The Gini coefficient for this industry is .36, which indicates a moderate amount of concentration. Suppose that the firm with a 30% market share acquires the firms with 5% and 15%

16. *See* HIRSCHMAN, *supra* note 4, at 157-58 n.2. If the Lorenz curve is obtained by sampling from the population, then this estimate of the Gini coefficient is biased, necessitating a multiplier of $n/(n-1)$. In many industrial organization contexts, no such correction is needed because the Lorenz curve is calculated from a complete population.

shares, while the firm with a 40% share merges with the remaining firm, such that a two-firm industry emerges in which each company enjoys a 50% market share. Figure 2(b) depicts the post-merger Lorenz curve. The Gini coefficient for the duopoly is 0, paradoxically indicating less concentration than before the consolidation. This is because the Gini Index measures only disparity in market share without regard to the absolute size of any firm's share. The Lorenz curve in figure 2(b) would look the same no matter how many firms were in the industry provided that each had an equal market share.

2. Other Attempts at Measuring Market Power

a. *Concentration Ratios*

One persistent yet unilluminating tool for measuring industry concentration is the concentration ratio.[17] It sums the market shares of the largest *x* firms in an industry, where *x* is typically 2, 4, or 8. This measure has been used since the time of the New Deal programs, when large volumes of industry statistics became increasingly available.[18] The concentration ratio provides little information about an industry's actual structure—it does not even amount to a point on a Lorenz curve.[19] Furthermore, a concentration ratio reveals nothing about the inequality among either the top *x* firms or the bottom

17. For general criticisms of concentration ratios, *see, e.g.*, Michael O. Finkelstein & Richard M. Friedberg, *The Application of an Entropy Theory of Concentration to the Clayton Act*, 76 YALE L.J. 677, 679-83 (1967).

18. *See, e.g.*, TEMP. NAT'L ECON. COMM., INVESTIGATION OF CONCENTRATION OF ECONOMIC POWER, 76TH CONG., 274-75 (Comm. Print 1941), *available at* https://archive.org/details/investigationofc27unit. The authors offer no justification for either their use of a concentration ratio or their choice of four rather than some other number of firms.

19. Without knowledge of how many firms comprise an industry, a concentration ratio provides only enough information to place a point on the vertical axis of a Lorenz curve; the point could fall anywhere to the right of the line of equal concentration, depending on the total number of firms. *See* Orris Clemens Herfindahl, Concentration in the Steel Industry 8-9 (1950) (unpublished Ph.D. dissertation, Columbia University) (on file with the Columbia University Library).

$n - x$ firms. For instance, an eight-firm industry dominated by one firm with a 65% market share followed by seven firms each with 5% market shares presumably would behave very differently than an industry in which the top four firms each had shares of 20% and the next four largest firms had shares of 5%. Yet the four-firm concentration ratio (CR-4) for both industries is the same—80%.

Concentration ratios have long held sway with courts in determining the competitive impact of mergers,[20] particularly after the DOJ formally adopted the four-firm concentration ratio in its initial 1968 Merger Guidelines.[21] The subsequent 1982 Guidelines ended the formal use of concentration ratios in favor of the HHI, and courts followed suit.[22] Although no longer part of the formal DOJ and FTC merger review process, concentration ratios nonetheless continue to appear in opinions considering Sherman and Clayton Act claims.[23]

b. *Lerner Index*

Economists predicated their search for a measure of industrial concentration on the assumption that an industry's structure influenced the conduct of its component firms—

20. *See, e.g.*, United States v. Gen. Dynamics Corp., 415 U.S. 486 (1974) (discussing the Clayton Act); United States v. Von's Grocery Co., 384 U.S. 270 (1966) (discussing the Clayton Act); United States v. Phila. Nat'l Bank, 374 U.S. 321 (1963) (discussing the Clayton and Sherman Acts).

21. *1968 Merger Guidelines*, DEP'T OF JUSTICE, http://www.justice.gov/atr/hmerger/11247.pdf (last visited Jan. 30, 2014) [hereinafter *1968 Merger Guidelines*].

22. *See* FTC v. PPG Indus., Inc., 798 F.2d 1500, 1503 (D.C. Cir. 1986) ("The FTC and the Department of Justice, as well as most economists, consider the [HHI] measure superior to such cruder measures as the four- or eight-firm concentration ratios which merely sum up the market shares of the largest four or eight firms . . . . This method, unlike the four- and eight-firm concentration ratios, shows higher market power as the disparity in size between firms increases and as the number of firms outside the first four or eight decreases.").

23. *See, e.g.*, Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 922-23 (2007) (Breyer, J., dissenting); Todd v. Exxon Corp., 275 F.3d 191, 208-09 (2d Cir. 2001); United States v. Manitowoc Co., No. 02-1509, 2002 WL 32060288, at *1 (D.D.C. Dec. 11, 2002); R.J. Reynolds Tobacco Co. v. Philip Morris Inc., 199 F. Supp. 2d 362, 367 (M.D.N.C. 2002).

conduct which in turn affected the degree of market power that individual firms enjoyed.[24] This aptly named structure-conduct-performance ("SCP") paradigm continues to enjoy cachet with both economists and courts.[25] Under the SCP framework, the purpose of a concentration measure is to provide as accurate a proxy as possible for the actual degree of monopoly power exhibited by firms.

The economist Abba Lerner questioned the need for such theoretical abstractions, proposing a more direct measure of monopoly power.[26] The Lerner Index (L), as his formula has become known, is a straightforward measurement of a firm's profits:

$$L = \frac{p - c}{p}$$

where L is the index coefficient, p is the price at which a firm sells a particular good, and c is the firm's marginal cost of producing the good. The Lerner Index thus avoids the difficulties inherent in choosing the relevant group of firms and products that comprise a given industry or market. Instead, it directly assesses the ability of a particular firm to charge a supracompetitive price for a particular product.[27]

Despite its theoretical attractiveness, the Lerner Index has garnered substantial criticism primarily leveled at its viability as a practical tool. Measurements of marginal cost are rarely, if ever, straightforward.[28] Small changes in the methodological

24. *See* Richard Gilbert & Oliver Williamson, *Antitrust Policy*, *in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 82 (Peter Newman ed., 1998).

25. *See, e.g.*, *In re* Urethane Antitrust Litig., No. 04-1616-JWL, 2012 WL 6681783, at *3 (D. Kan. Dec. 21, 2012) ("[S]tructure-conduct-performance analysis by an economist is well-accepted in [antitrust economics]."); TC Sys. Inc. v. Town of Colonie, 213 F. Supp. 2d 171, 183 (N.D.N.Y. 2002) (recognizing that "the S-C-P paradigm 'enjoys a long history and wide acceptance in the economic literature and in the antitrust courts'").

26. *See generally* A.P. Lerner, *The Concept of Monopoly and the Measurement of Monopoly Power*, 1 REV. ECON. STUD. 157 (1934).

27. This follows from the classical economic assumption that in a perfectly competitive industry firms choose to produce an output level at which the market-bearing price equals their marginal cost. *See, e.g.*, ROBERT S. PINDYCK & DANIEL L. RUBINFELD, MICROECONOMICS 267 (6th ed. 2005).

28. The American Airlines predatory pricing case illustrates the

assumptions for calculating marginal cost can lead to wildly divergent estimates of monopolistic profits.[29] Moreover, short-term demand shocks will change prices and, with them, the value of *L*, despite there being no underlying shift in a firm's market power.[30]

Even where accurate marginal cost calculations are possible, the Lerner Index often fails to reflect the competitive realities of a market. In a highly competitive industry where each firm faces a high, one-time sunk cost, the firms will need to recoup that cost, leading to prices above the industry-wide marginal cost. The Lerner Index would then falsely indicate the presence of some monopoly power.[31] Conversely, a firm with relatively high marginal costs may engage in predatory behavior and other tactics in order to preclude a potential rival with lower marginal costs from entering the market. By focusing on the incumbent firm rather than the more efficient potential entrant, the Lerner Index would paint a more sanguine picture of the competitive conditions than actually warranted.[32] Because the Lerner Index presents formidable

difficulties involved in calculating marginal costs. *See* United States v. AMR Corp., 335 F.3d 1109 (10th Cir. 2003). In 1999, the Government brought suit against American under Section 2 of the Sherman Act, alleging that American had responded illegally to new entrants in various markets involving its Dallas hub. *Id.* at 1111-13. In each case, American had expanded capacity, matched its rivals' low fares, and made those fares readily available. *Id.* Notwithstanding the significant drop in average fares in the relevant markets, the district court held for the airline and the Tenth Circuit affirmed, noting that "marginal cost, an economic abstraction, is notoriously difficult to measure and 'cannot be determined from conventional accounting methods.'" *Id.* at 1116 (citations omitted). The courts rejected all four measures of incremental costs proposed by the government, finding it undisputed that American had priced above average variable costs, a proxy for marginal costs. *Id.* at 1120.

29. *See* LYNNE PEPALL ET AL., INDUSTRIAL ORGANIZATION: CONTEMPORARY THEORY & PRACTICE 55 (3d ed. 2005); *see also* Ian Domowitz et al., *Market Structure and Cyclical Fluctuations in U.S. Manufacturing*, 70 REV. ECON. & STAT. 55 (1988); Robert E. Hall, *The Relation Between Price and Marginal Cost in U.S. Industry*, 96 J. POL. ECON. 921 (1988).

30. *See* IIB PHILLIP E. AREEDA ET AL., ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶ 505 (3d ed. 2007).

31. *See* PEPALL ET AL., *supra* note 29, at 54.

32. HERBERT J. HOVENKAMP, FEDERAL ANTITRUST POLICY: THE

practical difficulties, it has remained a theoretical construct not often used as a regulatory tool.[33]

3. The Herfindahl-Hirschman Solution

Albert Hirschman recognized the limitations of using the Lorenz-Gini methodology in an industrial organization context. In the appendix to his 1945 book on international trade, Hirschman pointed out the need for a measure of concentration that took into account not only equality of market shares, but also the number of total competitors:

> [T]he number of elements in a series the concentration of which is being measured is an important consideration. This is so whenever concentration means "control by the few," i.e., particularly in connection with market phenomena. . . One of the well-known conditions of perfect competition is that no individual seller should command an important share of the total market supply; this condition implies the presence of both relative equality of distribution *and* of large numbers.[34]

To this end, Hirschman argued that any index purporting to measure industrial concentration should increase as the dispersion in market shares increases and decrease as the number of firms increases.[35]

Hirschman proposed the following concentration index:

$$\sqrt{\sum_{j=1}^{n}\left(\frac{q_j}{Q}\cdot 100\right)^2}$$

where n is the number of firms, q_j is the output (or sales, profits, etc.) of the jth firm, and Q is the industry's total

LAW OF COMPETITION AND ITS PRACTICE 245-46 (3d ed. 1994).

33. *See* AREEDA ET AL., *supra* note 30, at ¶ 504.

34. HIRSCHMAN, *supra* note 4, at 158.

35. *Id.* at 160.

output.[36] Hirschman demonstrated that his index could be expressed equivalently as

$$100\sqrt{\frac{v^2+1}{n}}$$

where v is the coefficient of variation, equal to σ/μ, or the standard deviation of the series divided by its arithmetic mean.[37]

By breaking down the index into two components—one dependent on number of firms, n, and one dependent on the relative inequality of market shares, v—Hirschman fulfilled his self-imposed criteria. The index grows smaller as n increases and larger as v increases.

Five years later, Orris Herfindahl independently reached a very similar solution to the inadequacies of the Lorenz curve in measuring industrial concentration.[38] Examining concentration in the steel industry for his PhD dissertation, he proposed the familiar index:

$$\frac{\sum_{j=1}^{n} q_j^2}{\left(\sum_{j=1}^{n} q_j\right)^2}$$

which is functionally equivalent to Hirschman's index except for the square root sign and the scale.[39] As did Hirschman, Herfindahl noted that his index could be expressed as a relationship between the number of firms and the coefficient of variation.[40] Herfindahl surpassed Hirschman, however, in both

36. *Id.* at 159.

37. *Id.* The number 100 in the formulas merely allows the index to be expressed on a scale between 0 and 100; otherwise it would fall on the interval [0, 1]. Legal applications, following the practice in the Merger Guidelines, typically express the HHI on a scale of 0 to 10,000. Hirschman offers no reason for his use of the square root sign in the formulas.

38. *See generally* Herfindahl, *supra* note 19. Herfindahl, apparently unaware of Hirschman's earlier work at the time he began writing, acknowledges it in a footnote but adds that Hirschman "did not view the index as a weighted average nor give a graphic representation." *Id.* at 21 n.1.

39. *See id.* at 19. Hirschman offered no reason for his use of the square root sign. He may have simply applied it to his formula for the sake of balance.

40. *See id.* at 20.

his recognition of the legal applications and his conception of how to apply the index in such a context.

> [A]nti-trust law is assigning the size structure of firms a greater importance than formerly. Something less than one hundred per cent control of an industry is sufficient to make a showing of monopoly under the Sherman Act even in the absence of traditional acts in restraint of trade. The law will undoubtedly continue to use summary ideas resting on the size structure of firms. The economist will perform an important service if he can develop a more adequate account of the relationships between measures of concentration and market behavior.[41]

While acknowledging the imprecision with which any structural index indicated monopoly performance, Herfindahl suggested that his index might aid antitrust policy as a cost-efficient means of detection.[42] Presciently, he remarked that a strict reliance on his index would incur costs "through the harassment of some industries whose performance is actually quite competitive but whose structure, by the conventional standards, is not."[43]

Indeed, Herfindahl felt that any index of concentration had only a limited usefulness, as it would comprise only "one, or at most a few, of the many variables that determine the degree of monopoly in an industry."[44] In addition to the number of firms and the inequality of their market shares, he suggested several other factors that influence an industry's performance, such as the individual firms' "locational distribution," the "psychology" of corporate officers, and varying degrees of product substitutability.[45] Because structural indices have at best a limited correlation with industry performance, Herfindahl felt

41. *Id.* at 3 (footnote omitted).
42. *See id.* at 171.
43. *Id.*
44. *Id.* at 3.
45. *Id.* at 19, 170.

that his index should serve as just one tool among many in the application of antitrust policy.[46]

4. Towards an HHI Standard

Following the completion of Herfindahl's thesis, the HHI began to gain acceptance among industrial organization economists, in no small part due to the influence of Herfindahl's doctoral advisor, George Stigler.[47] The HHI had many attractive features for an index of concentration, both practical and theoretical.

One problem that plagued early forays into the study of industry structure was the lack of complete data and the lack of computing power to analyze the information available. The HHI advantageously required no complicated mathematical algorithms in its computation—merely addition and multiplication.

In the decades before the information age, many economists voiced concern that information about industry structure often only encompassed the largest firms in the industry, potentially omitting many if not most market participants.[48] This problem explains the popularity of the concentration ratio indexes despite their theoretical shortcomings. In theory, the HHI necessitated knowledge of every firm's market share, yet in practice one could often generate a very precise HHI approximation using only a few firms.[49] Moreover, to calculate the change in the index that a potential merger would produce, as is the current practice, it is necessary to know only the market shares of the two firms involved.[50]

46. *See id.* at 169-72.

47. *See* Stephen Calkins, *The New Merger Guidelines and the Herfindahl-Hirschman Index*, 71 CALIF. L. REV. 402, 409 (1983).

48. *See, e.g.*, Duncan Bailey & Stanley E. Boyle, *The Optimal Measure of Concentration*, 66 J. AM. STAT. ASS'N 702, 703 (1971) (observing the prevailing assumption "that 'better' indexes could be developed if more detailed firm data were available").

49. *See* HIRSCHMAN, *supra* note 4, at 160-62.

50. If two merging firms have market shares x_1 and x_2, then the change in HHI is equivalent to $2x_1x_2$. *See 1997 Merger Guidelines*, DEP'T OF JUSTICE & FTC 14 n. 8,

The index also allowed for an easy practical interpretation. Taking the inverse of the HHI yields the number of "effective competitors," or the number of equal-sized firms that would produce an equivalent HHI score.[51] The DOJ used this interpretation when it introduced the HHI in 1982.[52]

In addition, Herfindahl's index has an attractive theoretical link to the Lerner Index via a Cournot oligopoly model.[53] The economist Augustin Cournot, utilizing a framework that anticipated game theory by more than 100 years, demonstrated that two firms competing on output and letting the market forces determine the price would wind up selling at a price between that of a monopolist and that in a perfectly competitive market.[54] An extension of the Cournot duopoly model to include any number of firms reveals that as the number of firms decreases, the equilibrium price increases, the Lerner Index increases, and the HHI increases.[55] As the number of firms tends toward infinity, the equilibrium price

http://www.justice.gov/atr/hmerger/11251.pdf (last visited Jan. 30, 2014) [hereinafter *1997 Merger Guidelines*].

51. *See* M.A. Adelman, *Comment on the "H" Concentration Measure as a Numbers-Equivalent*, 51 REV. ECON. & STAT. 99, 101 (1969). For an example of the use of the phrase "effective competitors" in this context, see Steven A. Morrison & Clifford Winston, *The Dynamics of Airline Pricing and Competition*, 80 AM. ECON. REV. 389, 390 (1990). *But see* George Eads, *Intercity Passenger Transportation: Airline Capacity Limitation Controls: Public Vice or Public Virtue?* 64 AM. ECON. REV. 365, 367 (1974) (using "effective competitors" to denote those firms with a market share of greater than 10%).

52. *See 1982 Merger Guidelines*, DEP'T OF JUSTICE 1, 14, http://www.justice.gov/atr/hmerger/11248.pdf (last visited Jan. 30, 2012) [hereinafter *1982 Merger Guidelines*] (explaining that markets with an HHI above 1,800 "hav[e] the equivalent of no more than approximately six equally sized firms").

53. *See* TROUT RADER, A THEORY OF MICROECONOMICS 271-76 (1972); Robert E. Dansby & Robert D. Willig, *Industry Performance Gradient Indexes*, 69 AM. ECON. REV. 249 (1979).

54. *See* AUGUSTIN COURNOT, RESEARCHES INTO THE MATHEMATICAL PRINCIPLES OF THE THEORY OF WEALTH 79-89 (Nathaniel T. Bacon trans., 1897) (1838).

55. Cournot made this extension. For a modern exposition, see PEPALL ET AL., *supra* note 29, at 215-23. Formally, the relationship is

$$L = \frac{p - \bar{c}}{p} = \frac{\mathrm{HHI}}{\eta}$$

where $\bar{c}$ is the weighted average industry marginal cost and η is the price elasticity of demand.

approaches the perfectly competitive level, and the Lerner and Herfindahl-Hirschman Indexes approach zero.[56]

Of course, the Cournot model is not itself without detractors—one important early critic pointed out that many firms compete on price rather than output.[57] Nonetheless, it remains a workhorse in the field of industrial organization because it explains, using a very simple model of firm behavior, why a decreasing number of firms tends to result in an increasing ability to coordinate output and pricing, even in the absence of tacit collusion. Thus, the Herfindahl-Cournot-Lerner models provide a theoretical justification for the structure-conduct-performance paradigm.

B. *The Merger Guidelines*

In the same year that Herfindahl submitted his doctoral dissertation, Congress passed the Celler-Kefauver Act,[58] which broadened the government's power to challenge mergers and marked the beginning of modern merger law.[59] The DOJ and FTC originally derived their power to review and approve mergers from Section 7 of the Clayton Act,[60] and Federal Trade Commission Act,[61] respectively.[62] The Celler-Kefauver Act,

56. *See id.*

57. *See* Joseph Bertrand, *Review of* Theorie Mathematique de la Richesse Sociale *and of* Recherches sur les Principles Mathematiques de la Theorie des Richesses, 68 J. DES SAVANTS 499 (1883).

58. Pub. L. No. 899, 64 Stat. 1125 (codified at 15 U.S.C. §§ 18, 21 (2012)).

59. *See* William J. Kolasky & Andrew R. Dick, *The Merger Guidelines and the Integration of Efficiencies into Antitrust Review of Horizontal Mergers*, 71 ANTITRUST L.J. 207, 210-12 (2003). The first major case to apply the Celler-Kefauver legislation and provide an interpretation of its impact on merger regulation was *Brown Shoe Co. v United States*, 370 U.S. 294 (1962).

60. 15 U.S.C. § 18 (2012).

61. *Id.* § 45. The Act precludes the FTC from evaluating bank and airline mergers. *See id.* § 45(a)(2).

62. Most actions to enjoin mergers invoke the Clayton Act. The Sherman Act, 15 U.S.C. §§ 1-2, also grants the federal government power to stop mergers "in restraint of trade" or those that monopolize or attempt to monopolize a market. *See* Constance K. Robinson, *Mergers and Acquisitions*, *in* 1 CORPORATE LAW AND PRACTICE

enacted at a time of great concern over "a rising tide of economic concentration in the American economy," gave teeth to the existing antitrust legislation by closing a number of loopholes and granting the government the power to curb any potential lessening of competition at its outset.[63]

In 1968 the DOJ issued, for the first time, a set of guidelines explaining its approach to merger evaluation.[64] This document had two notable features: a focus on market structure and the use of the four-firm concentration ratio as a measurement tool.[65]

By concentrating its merger analysis on a few structural factors, the DOJ argued that it would "produce economic predictions that are fully adequate for the purposes of [the Clayton Act]."[66] Additionally, its approach promoted efficient decision-making and provided transparency to industry participants.[67] The DOJ thus adopted the prevailing belief by economists that industry performance cannot be measured in any consistent or accurate way, and that industry structure serves as a useful proxy.

The DOJ accorded "primary significance" to market shares in assessing market structure for horizontal merger evaluations.[68] For fourteen years, this meant using the CR-4.[69] Industries received a classification of either "Highly Concentrated"—those with a CR-4 of 75% or more—and "Less Highly Concentrated"—those with a CR-4 of less than 75%.[70] Concentration ratios do not convey sufficient information for a

COURSE HANDBOOK SERIES: 45TH ANNUAL ANTITRUST LAW INSTITUTE 297, 303 (2004).

63. *Brown Shoe Co.*, 370 U.S. at 315-20.

64. The Guidelines limit discussion to merger analysis under the Clayton Act only. *1968 Merger Guidelines*, *supra* note 21, at 1.

65. *See id.* at 1-2, 14-15.

66. *Id.* at 1-2.

67. *See id.*

68. *See id.* at 5.

69. *See id.* at 6.

70. *Id.* at 6.

mapping to a unique HHI score; an industry with a CR-4 of x% could reflect an HHI value between $\left(\frac{x}{4}\right)^2$ and x^2.[71]

The 1968 Guidelines state that the DOJ will "ordinarily challenge" mergers in highly concentrated markets where firms of certain enumerated sizes acquire other firms of certain enumerated sizes.[72] In roughly equivalent terms, mergers inducing a change in HHI of more than 30 points would trigger a challenge.[73] For less highly concentrated markets, mergers resulting in an HHI increase of 50 or more would trigger a challenge.[74]

In 1982 the DOJ promulgated a new set of merger guidelines, in which the HHI replaced the four-firm concentration ratio as the method of measuring industry structure.[75] The DOJ's decision to embrace the HHI was remarkable given that courts had mostly rejected or ignored it.[76] The change in methodology among DOJ lawyers reflected an implicit acceptance of the criticisms of the CR-4 and was also influenced substantially by the work of Herfindahl's mentor, George Stigler.[77] The HHI had appeared in the legal

71. For a given concentration ratio, the lowest possible HHI would occur in an industry with four equal-sized firms and a virtually infinite number of other firms, each with a market share of close to zero. The HHI's upper bound for the same concentration ratio would reflect an industry consisting of one firm with a 75% market share and a virtually infinite number of other firms, each with a market share of close to zero.

72. *1968 Merger Guidelines*, *supra* note 21, at 6.

73. The Guidelines offer three examples: a firm of 4% market share acquiring another firm of 4% or more; a firm of 10% acquiring a firm of 2% or more; and a firm of 15% or more acquiring a firm of 1% or more. These scenarios represent, respectively, minimum HHI increases of 32, 40, and 30.

74. For example, a firm with a 5% market share acquiring another firm of 5% or more (increasing HHI by 50 or more) would ordinarily receive a challenge, as would a firm with a 15% market share acquiring a firm of 3% or more (increasing HHI by 90 or more). *See 1968 Merger Guidelines*, *supra* note 21, at 6.

75. *See 1982 Merger Guidelines*, *supra* note 52, at 11-15.

76. Prior to the 1982 Guidelines, only six judicial opinions had mentioned the HHI. *See* Calkins, *supra* note 47, at 410.

77. *See* David Scheffman et al., *Twenty Years of Merger Guidelines Enforcement at the FTC: An Economic Perspective*, 71 ANTITRUST L.J. 277, 283 (2003) ("The appeal of the HHI was that it was related to Stigler's 'Theory of Oligopoly,' which was the

literature as early as 1969, when Richard Posner endorsed it, and Areeda and Turner later gave it their approval, "albeit without enthusiasm."[78] The FTC did not formally adopt the DOJ guidelines until 1992, when the two issued a set of joint guidelines, but it effectively endorsed the DOJ framework from 1982 onwards.[79]

The DOJ issued further versions of the Merger Guidelines in 1984, 1992, 1997, and 2010. Although the wording varied somewhat, the essential use of HHI in structural analysis remained substantially the same.[80] Classification of industrial concentration levels expanded from two to three regions: "highly concentrated" (HHI above 2,500), "moderately concentrated" (HHI between 1,500 and 2,500), and "unconcentrated" (HHI below 1,500).[81] More subtly, policy shifted from an examination of pre-merger market shares—used in determining the CR-4—to an analysis of post-merger shares.

The structural analysis thus considered the change in industry HHI rather than making a static examination of the merging firms' market shares (although the two are directly related).[82] By focusing on the change in the index values, the Merger Guidelines adhere to Herfindahl's own beliefs about the use of his index in that they use it comparatively rather than

foundation of the Guidelines' collusion analysis (now known as coordinated interaction).").

78. Calkins, *supra* note 47, at 409-10.

79. *See* Scheffman et al., *supra* note 77, at 284.

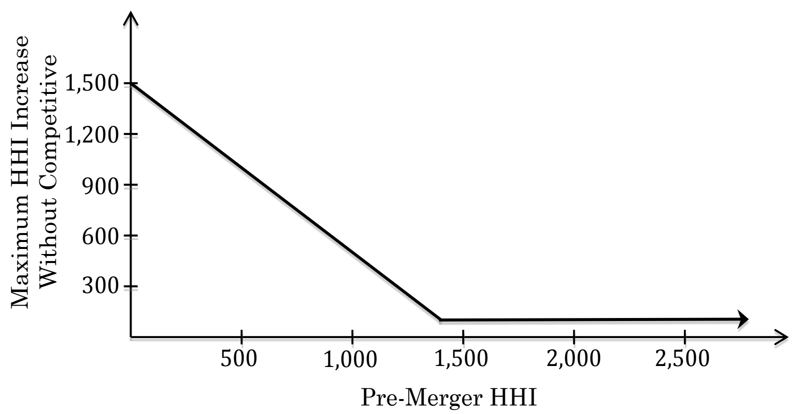
80. A proviso barring the leading firm in an industry from merging with a firm of 1% market share or more was dropped in the 1992 Guidelines. The subsequent guidelines did have several substantive changes in other respects, in particular relating to the use of efficiency as a defense of otherwise anticompetitive mergers. *See id.*; Kolasky & Dick, *supra* note 59.

81. *See 2010 Merger Guidelines*, DEP'T OF JUSTICE & FTC, http://www.justice.gov/atr/public/guidelines/hmg-2010.pdf (last visited Jan. 30, 2014) [hereinafter *2010 Merger Guidelines*]. The 2010 revisions to the Guidelines increased the region thresholds substantially. Previously, markets were highly concentrated if the HHI level was greater than 1,800 and moderately concentrated between 1,000 and 1,800. *See 1997 Merger Guidelines*, *supra* note 50, at 14-15.

82. *See 1997 Merger Guidelines, supra* note 50.

focusing on absolute HHI levels.[83] The Guidelines recognize that, all else equal, the more concentrated an industry is, the more likely a given change in HHI will raise competitive concerns. Nonetheless, the Guidelines do not set any maximum HHI beyond which all mergers are presumptively anticompetitive, as illustrated in Figure 3. Transactions resulting in a moderately concentrated industry post-merger "potentially raise significant competitive concerns and often warrant scrutiny" if the HHI increases by more than 100 points.[84] Those resulting in a highly concentrated post-merger industry are treated similarly if the HHI increases between 100–200 points and are "presumed to be likely to enhance market power" if the HHI increases by more than 200 points.[85]

Figure 3.



To the extent one can make a meaningful comparison between the CR-4 system of the 1968 Guidelines and the HHI scheme introduced in 1982, the threshold levels for regulatory scrutiny are not wildly divergent.[86] For its part, the DOJ

83. *See* Herfindahl, *supra* note 19, at 21-22 ("The usefulness of the measure lies in providing a definite description of gross changes and in furnishing a focus for further judgments about the data.").

84. *See 2010 Merger Guidelines*, *supra* note 81, at 19.

85. *See id.*

86. The 1982 Guidelines state that "the critical HHI thresholds at 1000 and 1800 correspond roughly to four-firm concentration ratios

believed that the market share thresholds for merging firms were comparable between the two sets of guidelines.[87] In contrast, the DOJ anticipated that its new thresholds for industry concentration would be received as an overly permissive departure from the 1968 Guidelines and thus chose the original safe-harbor threshold of 1,000 "as much as a political anchorage to windward as because anyone thought that nicely round number was just right."[88]

In practice, the government has not strictly adhered to the numerical thresholds. The 1982 Merger Guidelines themselves concede that "[o]ther things being equal, cases falling just above and just below a threshold present comparable competitive concerns."[89] An internal FTC study found that during the 1980s the minimum level of HHI leading to merger objections generally exceeded 1800 by a few hundred points.[90] While this evidence might indicate a decreased reliance on HHI,[91] it is equally consistent with higher unofficial HHI thresholds as part of the Reagan administration's generally laissez-faire economic policies.

III. Criticisms of the Herfindahl-Hirschman Index

Two primary factors justify the use of the HHI in merger evaluation. Sound economic theory links market structure, as measured by the HHI, to market power, and by extension, market performance. More practically, the Guidelines' use of the HHI provides an objective benchmark that mitigates the ebb and flow of political and academic favor towards antitrust

of 50 percent and 70 percent, respectively." *1982 Merger Guidelines*, *supra* note 52, at 11-15.

87. *See* William F. Baxter, *Antitrust Policy*, *in* AMERICAN ECONOMIC POLICY IN THE 1980S 610 (Martin Feldstein ed., 1994).

88. *Id.*

89. *1982 Merger Guidelines*, *supra* note 52, at 13. But the Guidelines justify deviations from their bright-line rules by noting "the numerical divisions suggest greater precision than is possible with the available economic tools and information," suggesting that in the future more precision will be possible. *Id.*

90. *See* Scheffman et al., *supra* note 77, at 300.

91. *See id.*

policy.[92] Nonetheless, several criticisms lie against the HHI, both in general and in particular, with regard to its use in network industry mergers.

A. *Criticisms Applicable to All Markets*

Herfindahl himself admitted that his index suffered the deficiency of considering only two indicia of industry behavior—the number of firms and the variance in market share distribution.[93] He recognized that other factors would also play a role, in particular geographic dispersion of firms.[94] The Merger Guidelines cleverly handle this issue by dividing the inquiry into two steps: first market definition, then concentration analysis. Because the relevant market inquiry considers both geography and demand cross-elasticities,[95] the HHI implicitly includes those factors.

Other criticisms of the HHI broadly fall into two areas. The first attack its use in practice, either through faulty or inconsistent application or because of inappropriate calibration. The second line of criticism focuses on the Index's theoretical underpinnings, arguing that it fails on a more fundamental level.

1. Criticisms of the HHI in Practice

The Guidelines' use of the HHI has garnered criticism for its arbitrary numerical thresholds.[96] Former Assistant Attorney General William Baxter, who was primarily responsible for the 1982 Guidelines, admitted as much, conceding the "arbitrary" lines "have no magical qualities"

92. But for criticism of both the theoretical and practical justifications, see Robert D. Joffe et al., *Proposed Revisions of the Justice Department's Merger Guidelines*, 81 COLUM. L. REV. 1543 (1981).

93. *See* Herfindahl, *supra* note 19, at 19.

94. *See id.*

95. The Merger Guidelines specifically address geographic markets and implicitly address demand cross-elasticities in their SSNIP test. *See 2010 Merger Guidelines, supra* note 81, at 8-14.

96. *See, e.g.*, Jay Greenfield, *Beyond Herfindahl: Non-Structural Elements of Merger Analysis*, 53 ANTITRUST L.J. 229 (1984).

beyond "the fact that we were born with ten fingers and have gotten used to a base ten system."[97] However, this critique does not really address the adequacy of the HHI per se, but merely the details of its implementation.

A related argument posits that the precision and sophistication of the HHI may cloak its limitations and create a false impression of scientific accuracy in the courts.[98] Exacerbating this problem, the argument continues, the government's discretion in market definition renders any apparent concreteness in concentration thresholds illusory.[99] Yet, it seems doubtful that judges intelligent enough to parse the results of the CR-4 with skepticism would be duped into blind acceptance of the government's case solely on HHI evidence. Moreover, any measure of concentration would fall prey to this criticism of misleading accuracy; surely it is better to have *some* objective standard in merger evaluation in the interests of fairness and consistency.

Another criticism of the HHI highlights its potential to lead to wide measurement errors. As Hirschman noted, measurement errors of smaller firms' market shares (perhaps from incomplete data) lead to relatively minor fluctuations in an industry's HHI value.[100] But if the errors involve larger firms, the HHI calculation for the industry can produce large errors.[101] This argument overstates the problem, since the Guidelines look at the change in HHI rather than absolute values—other than as a threshold matter. The error would presumably appear in both the pre- and post-merger calculations, resulting in a de minimis distortion to the increase in HHI. Moreover, any index of concentration will suffer from measurement bias, and the HHI does not suffer more in this regard than others.

Finally, the charge has been made that by lumping post-merger HHI into one of three regions, the DOJ does not treat

97. William F. Baxter, *A Justice Department Perspective*, 51 ANTITRUST L.J. 287, 292 (1982).

98. *See* Calkins, *supra* note 47, at 425.

99. *See* Steven C. Salop, *Symposium on Mergers and Antitrust*, 1 J. ECON. PERSP. 3, 10 (1987).

100. *See* HIRSCHMAN, *supra* note 4, at 157-62.

101. *See* Calkins, *supra* note 47, at 405.

all mergers equally.[102] Prior to the 2010 revisions, substantial arbitrariness was possible. Given two mergers, one with a higher HHI change could have passed muster under the former Guidelines while the merger with a lesser change in HHI triggered the presumption of anticompetitiveness.[103] Consider the two hypothetical industries given in Figure 4. In industry A, a merger between the firms with market shares of 2% and 13% presumptively would have been anticompetitive under the former Guidelines, whereas in industry B, the firm with a 2% share could have merged with the firm having a 24% share without any concern of antitrust action. Both mergers result in an industry HHI of close to 1,800, and the merger sanctioned by the Guidelines would raise HHI by nearly twice that of the merger that was presumed anticompetitive.

102. *See* George G. Szpiro, *A Note on the Equitable Treatment of Mergers*, 13 MANAGERIAL & DECISION ECON. 543 (1992). Szpiro misunderstands the Merger Guidelines to focus on pre-merger HHI, and offers no justification for his choice of just two of the potentially infinite data points from which to extrapolate a linear relationship between pre-merger HHI and the maximum permissible increase in HHI.

103. *See id.* This disparity was possible because, under the former Merger Guidelines, if a merger resulted in a moderately concentrated industry, the HHI had to increase by more than 100 points before anticompetitive concerns were triggered. If the resultant industry was highly concentrated, however, the HHI had to increase by only 51 points to create a presumption of enhanced market power. Under the current Guidelines, no merger raises concerns if the HHI change does not exceed 100 points. *Compare 2010 Merger Guidelines*, *supra* note 81, *with 1997 Merger Guidelines*, *supra* note 50.

Figure 4.

	Market Share Industry A	Industry B
Firm 1	27%	24%
Firm 2	17%	17%
Firm 3	14%	17%
Firm 4	14%	16%
Firm 5	13%	12%
Firm 6	13%	12%
Firm 7	2%	2%
Pre-merger HHI	1,752	1,702
Post-merger HHI	1,804	1,798
Change	52	96

The 2010 revisions to the Merger Guidelines fixed this problem by changing the HHI differentials necessary to trigger anticompetitive concerns.[104] In any event, this type of criticism ignores both the letter and the spirit of the Guidelines. As discussed above, the merger regulators' analyses do not turn on such technical minutiae.[105] Not surprisingly, most of the practical criticisms of the HHI arose around the time the DOJ adopted the 1982 Merger Guidelines. In the three decades since, these arguments have clearly proven unfounded. The theoretical criticisms offer more substance.

2. Criticisms of the HHI in Theory

The most common theoretical attacks on the HHI center on its weightings of the number of firms, *n*, versus their size distribution, *v*. For instance, some commentators have suggested that the HHI overstates the potential competitive impact of mergers involving large and small firms (i.e., *n* is overemphasized relative to *v*).[106] Unless a small firm "is a 'maverick' or has substantial excess capacity and competitive

104. *Compare 1997 Merger Guidelines*, *supra* note 50, *with 2010 Merger Guidelines*, *supra* note 81.
105. *See supra* p. 121.
106. *See* Scheffman et al., *supra* note 77, at 283.

costs[,]" its acquisition by a larger competitor should not raise the concerns an HHI analysis might indicate.[107]

In the opposite vein, others have claimed that the HHI understates the value of small competitors (i.e., v is overemphasized relative to n).[108] For example, if an industry has a dominant firm with market share s_1 and $(n - 1)$ smaller, equally sized firms, then the limit of the HHI is s_1^2 as $n \to \infty$. If the dominant firm merges with one of the smaller firms, the market's HHI necessarily will increase. Previously, the Guidelines held that *any* increase in HHI raised competitive concerns if the post-merger HHI resulted in a concentrated market. The problem was that if the dominant firm's pre-merger market share were sufficiently large (under the previous Guidelines, when s_1>41%), the merger would raise competitive concerns no matter how many additional firms occupied the market. A superior index, the argument goes, would not cap the effect of additional firms, no matter how small.[109] In other words, as the number of firms in a market approaches infinity, the concentration index should approach zero, no matter what the size distribution.[110]

The most recent Guidelines respond to this criticism by imposing a 100-point buffer before scrutiny is triggered, even in highly concentrated markets, because HHI increases of less than 100 points "are unlikely to have adverse competitive effects."[111] While the HHI does not approach zero as the number of competitors becomes sufficiently large, the number of competitors may now make a difference in whether the regulators investigate a merger in a concentrated market. If a dominant firm merges with one of its small but equally sized competitors, HHI will *never* increase by more than 100 points—no matter how large the dominant firm—if the firm has more than 50 pre-merger competitors.[112]

107. *Id.*

108. *See* Finkelstein & Friedberg, *supra* note 17.

109. *See id.*

110. George Stigler has described such a measure as "stimulating and appealing" although "it lacks any precise theoretical rationale." George J. Stigler, *Comment*, 76 YALE L.J. 718, 718 (1967).

111. *2010 Merger Guidelines*, *supra* note 81, at 19.

112. For the HHI to increase by 100 points or more,

Another theoretical criticism of the HHI points out its inability to assess dynamic aspects of competition.[113] The HHI provides a snapshot of the "before" and "after" images of market structure. But in the wake of a merger, sufficiently large price increases may draw in other competitors, deconcentrating the market. In the alternative, the presence of potential entrants may restrain significant price increases altogether. Dynamic models of concentration would recognize that "cost savings are generally longer lived than anticompetitive effects" and apply some sort of intertemporal discount factor to the effects of increased concentration.[114] However, such models make more sense when weighing the costs and benefits of a merger in an efficiencies context—after the HHI threshold has been reached.[115]

A final criticism of the HHI argues that as a structural measure applied uniformly across industries, it misses the industry-specific nuances relating structure to market power and thus wholly fails to provide regulators with useful information. For example, HHI provides no information about barriers to entry, economies of scale or scope, rapidly changing technology, or firm-specific characteristics, all of which may bear on the degree of competition in the industry.[116] I will make

---

$$\left(s_1 + \frac{100 - s_1}{n-1}\right)^2 - \left(s_1^2 + \left(\frac{100 - s_1}{n-1}\right)^2\right) \geq 100$$

or $n \leq \frac{s_1(100-s_1)}{50} + 1$. The maximum value of n required to satisfy this inequality is 51, which occurs when $s_1 = 50$. The 100-point buffer produces interesting outcomes in near-monopoly markets. If the dominant firm has a 99% market share and two other firms equally split the remaining 1% of the market, then a merger between the dominant firm and one of the other two firms will raise the HHI by only 99 points—ordinarily an insufficient amount to trigger scrutiny.

113. *See* Joseph Farrell & Carl Shapiro, *Horizontal Mergers: An Equilibrium Analysis*, 80 AM. ECON. REV. 107 (1990) (arguing that market shares among both merging and non-merging firms will change, necessitating an analysis of post-merger equilibrium to determine the welfare effects).

114. *See* Scheffman et al., *supra* note 77, at 314.

115. *See id.*

116. *See* Geraldine Alpert, *Is Structure All?*, 53 ANTITRUST L.J. 255 (1984).

a similar argument in the next section with respect to HHI's failure to account for network effects.

In fairness, attacking the HHI for failing to address every individual industry peculiarity somewhat misses the point of the index. The HHI offers a general framework for approximating competitive effects. It acts as a gatekeeper in ruling out cases in which consolidation presents no obvious harm to consumers. In this respect, it may prove underinclusive—committing the Type II error of allowing undesirable mergers to go through. But if the HHI does raise a red flag, the Merger Guidelines have alternative provisions to address some of the characteristics that might make an industry unique. For example, regulators consider barriers to entry when determining the relevant market (a prerequisite to applying the HHI), in that they affect the ability of firms to effectuate a "small but significant and non-transitory increase in price."[117] Economies of scale and scope should also receive consideration when the government considers efficiencies that would offset market power.

The HHI probably does provide the best general measure of industrial concentration, and the Merger Guidelines have the flexibility to apply it usefully in a variety of industry settings. Discarding the HHI because it does not perfectly forecast market power in every application would be senseless. At the same time, when it becomes apparent that particular types of industries do not fit well into the HHI mold for some critical reason, the Guidelines can and should be adjusted to account for that. I turn to such an argument now.

B. *Criticism of the HHI As Applied to Network Industries*

Networks contain nodes connected by links.[118] Structurally, a network can take many forms. One common

117. *1982 Merger Guidelines*, *supra* note 52, at 9.

118. *See generally* LAWRENCE J. WHITE, U.S. PUBLIC POLICY TOWARD NETWORK INDUSTRIES (1999) (offering a thorough overview of network industry economics and the attendant policy issues); Nicholas Economides, *The Economics of Networks*, 14 J. INT'L J. INDUS. ORG. 673, 674 (1996) (providing another excellent discussion of network industry economics).

arrangement, called a star or hub-and-spoke network, involves one or more common nodes (or "hubs") connected uniquely to many other nodes (or "spokes").[119] This form of network occurs, for instance, in the communications and airline industries. However, various topographical forms of networks exist in many diverse industries, and the term as used here in an economics context should be distinguished from its colloquial reference to specific products, such as hardware and software.[120]

The fundamental characteristic of a network industry is "positive consumption and production externalities."[121] These externalities occur when "the value of a unit of the good increases with the expected number of units to be sold."[122] In other words, users of a network good or service derive value both from the good or service itself and also from the direct or indirect effects of additional users. Computer operating systems offer a classic example: consumers benefit from the functionality that an operating system provides. Additionally, when they purchase a popular operating system, consumers benefit from not having to learn other platforms on computers outside the home (a direct benefit) and from having a wide variety of applications to choose from (an indirect benefit).

Positive consumption externalities not surprisingly have an impact on demand. As the expected number of users of a network commodity increases, so does the price a given number of potential users will pay—i.e., the demand curve itself shifts up, although it remains downward sloping.[123] As an illustration, consider a hypothetical single-firm network industry in which a finite number of potential consumers have a willingness to pay that is uniformly distributed between 0 and a when everyone uses the network. That is, the number of potential customers equals the number of actual customers at a

119. *See* Economides, *supra* note 118, at 675.

120. *See* Daniel L. Rubinfeld, *Antitrust Enforcement in Dynamic Network Industries*, 43 ANTITRUST BULL. 859, 861 (1998).

121. Economides, *supra* note 118, at 678.

122. *Id.* (emphasis omitted).

123. *See id.*

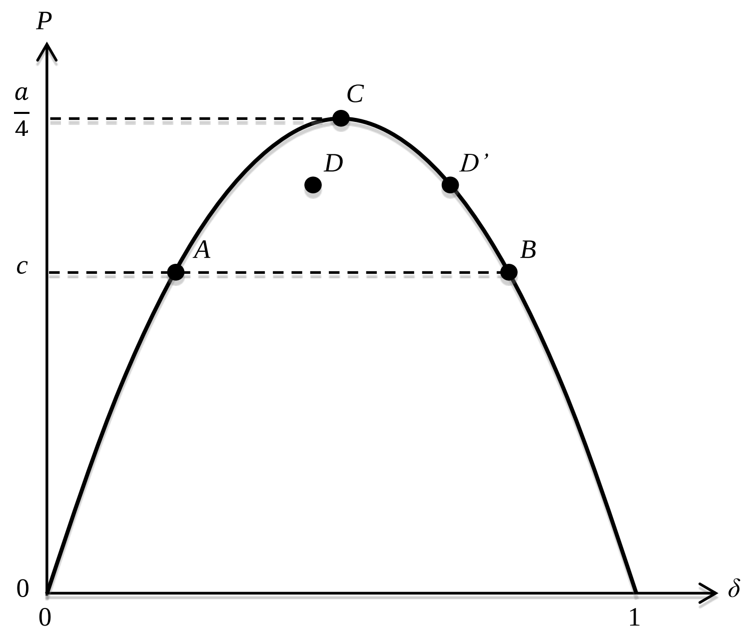
price of zero.[124] This implies the familiar linear demand curve. In accordance with the positive consumption externality, each potential customer discounts her willingness to pay by δ, the ratio of actual to potential customers, which is equivalent to the share of the market served by the firm.[125] In equilibrium, $P = a\delta(1 - \delta)$ where P is the price charged.[126] Figure 5 depicts this relationship.

124. This example is adapted from Jeffrey Rohlfs, *A Theory of Interdependent Demand for a Communications Service*, 5 BELL J. ECON. & MGMT. SCI. 16 (1974).

125. Of course there is no reason other than ease of exposition to assume a linear relationship between δ and willingness to pay. It may well be the case that there is a critical mass of δ somewhere between 0 and 1, before which there is an increasing marginal willingness to pay and beyond which a decreasing marginal willingness to pay.

126. As a proof, let x denote the large but finite number of potential customers, and Q the number of actual customers. Given the uniform willingness to pay, $Q = x(1 - P/a)$ or $P = a(1 - Q/x) = a(1 - \delta)$ when $\delta = 1$. Given that each potential customer discounts her willingness to pay by δ, the inverse demand curve becomes $P = a\delta(1 - \delta)$ for $\delta \in [0, 1]$. For an elaboration, see PEPALL ET AL., *supra* note 29, at 615-20.

Figure 5.



At any positive price greater than or equal to marginal cost *c* and less than *a*/4, two potential equilibria exist. For instance, at both points *A* and *B*, demand equals supply, the additional demand at *B* stemming from the additional value that the greater number of network users generates. At *D* the number of network users creates value exceeding that reflected in the price, and additional consumers (those with lower willingness to pay at any network size) will want to join the network. In equilibrium, *D* will shift to *D*'.

Thus, the upward sloping side of the demand curve represents a "critical mass."[127] At points to the left of this curve, the price is too high to sustain the number of users. Those with the lowest willingness to pay will drop out of the

127. *See* Rohlfs, *supra* note 124, at 29.

market.[128] But the reduced number of users diminishes the value of the network to the remaining users, and more drop out. The effect snowballs until no users remain and the network fails. To the right of the critical mass the opposite effect takes hold, as illustrated by the move from *D* to *D*'. The critical mass curve itself represents a set of unstable equilibria.[129] Outcomes along the curve remain tenable only so long as no exogenous forces perturb the steady state.

It is important to distinguish network industries from natural monopolies. Industries of both types have a tendency towards consolidation, albeit for different reasons. Natural monopolies possess increasing returns to scale, meaning that as they increase output their average costs decrease. Typically, this occurs when a firm incurs high fixed costs and low marginal costs. Larger firms will thus maintain cost advantages over smaller firms and in the long term the industry becomes a small oligopoly or monopoly in the absence of government intervention.

By contrast, network industries tend towards monopoly not because of increasing returns to scale—although they often exhibit that trait—but because, all else equal, consumers derive greater utility purchasing from firms with more customers. Even in a network industry with decreasing returns to scale (i.e., where average unit costs increase with output), consumers will purchase a more expensive product if a sufficiently large number of other consumers use it.

For example, an electric power company has high startup costs but relatively low marginal costs—once the expensive infrastructure is in place, it costs virtually nothing to add an additional user and its marginal cost is essentially the cost of electricity production. At the same time, there are no direct benefits and few indirect benefits to one user of adding additional users. The electric power industry would therefore be classified as a natural monopoly but not a network industry.

128. This raises the question of how a network firm reaches its critical mass in the first place without the help of a subsidy or below-cost pricing, an issue beyond the scope of this discussion.

129. *See* Rohlfs, *supra* note 124, at 29.

The telecommunications industry is classified as a natural monopoly for reasons similar to the electric power industry. Telecommunications also qualifies as a network industry because there are direct and indirect benefits to one user as the number of other users increase.

The apparel industry serves as an example of a network industry that is not a natural monopoly. Fixed costs are low relative to variable costs, most costs increase at least proportionally with output, and no appreciable economies of scale exist.[130] On the other hand, as more and more individuals don a particular brand of clothing, it becomes increasingly fashionable and other consumers' willingness to pay for it increases.[131]

This distinction between industries with network effects and natural monopoly industries is important for policy reasons. The government typically protects consumers from a natural monopoly by allowing the monopoly but strictly regulating prices, profits, or both.[132] No such rationale supports the imposition of similarly draconian measures on network industries that do not exhibit natural monopoly tendencies; consumers may in fact benefit from the concentration even in the presence of higher prices.

The use of the HHI in horizontal merger regulation presumes that consumer welfare suffers as firms increase their market share and take advantage of the consequent increase in market power to raise prices. In network industries, the validity of this presumption comes into question. Consumer

130. *See* Yoram Gutgeld & Damon Beyer, *Are You Going Out of Fashion?*, 3 MCKINSEY Q. 55, 58 (1995).

131. But only to a point there are also negative network effects that eventually outweigh the positive ones associated with more users. On the one hand, an increasing number of users reassures consumers that the garment has value. On the other hand, as an article of clothing becomes more ubiquitous each wearer feels less unique and the brand value becomes diluted. *See* Peter M. Kort et al., *Brand Image and Brand Dilution in the Fashion Industry*, 42 AUTOMATICA 1363 (2006).

132. *See generally* LUÍS M.B. CABRAL, INTRODUCTION TO INDUSTRIAL ORGANIZATION 75-77 (2000).

surplus—the traditional measure of consumer welfare—is the difference between the amount consumers actually pay for a product and the aggregate amount they would be willing to pay for the product.[133] Consumer surplus should fall when the post-merger price rises because of enhanced market power—the competitive effect.[134] Conversely, consumers will pay more for a product with positive consumption externalities following a merger that expands the network size. This increased willingness to pay tends to increase consumer surplus—the network effect.[135]

On balance then, it remains uncertain whether a network industry merger will increase or decrease consumer surplus. The result depends on which effect dominates: the competitive or the network effect. The next section explores an economic model of this relationship and examines situations in which the benefits from the network effect outweigh the losses to consumers from lessened competition. The model takes as its starting point the basic Cournot model of competition, with its direct link to the HHI.[136]

IV. A Model of Competition in Network Industries

A. *The Basic Model of Network Industry Competition*

In the basic model $n > 1$ firms face $x > 0$ consumers, where x is assumed to be a sufficiently large number. Willingness to pay is distributed uniformly among consumers on $[0, a]$. Thus, firms face a linear demand function at a given level of output.

133. *See, e.g.*, WALTER NICHOLSON, MICROECONOMIC THEORY: BASIC PRINCIPLES AND EXTENSIONS 145-47, 317-18 (9th ed. 2005). In general, different consumers would be willing to pay different amounts for the same product. It is this phenomenon that gives rise to the downward slope of the demand curve. At any given price, different consumers will realize different surpluses. Consumer surplus for a market is the sum of surpluses across all consumers whose willingness to pay equals or exceeds the price charged.

134. *See* Economides, *supra* note 118, at 691.

135. *See id.*

136. *See supra* pp. 115-116.

The importance of the latter limitation will be explained shortly.

Firms compete à la Cournot in that each firm $j \in \{1, \ldots, n\}$ simultaneously selects its output, q_j, taking into account the demand function and the industry cost function. Each firm incurs zero fixed costs and a constant marginal cost of $c_j = c$. Because costs must be positive and less than the amount at least one person is willing to pay, c is confined to the interval $(0, a)$.

Consumers' willingness to pay also depends on the network size of the firm from which they make their purchase. Accordingly, the demand function D that a particular firm faces is discounted by the fraction of the market that firm captures, $\delta_j = \frac{q_j}{x}$. In other words, $p_j = D(q_j, Q_{-j}) \cdot \delta_j$ where p_j is the price charged by firm j and $Q_{-j} = \sum_{k=1}^{n} q_k - q_j$. This departs from the Cournot model in that firms do not necessarily sell the good at the same price in equilibrium. Larger firms—those with greater output—will charge a higher price for their product. Nonetheless, the undiscounted prices must be equal regardless of output to avoid arbitrage, i.e., $p_1^u = \ldots = p_n^u = p^u$.

Accordingly, the undiscounted market demand function is

$$Q = x\left(1 - \frac{p^u}{a}\right)$$

and the inverse demand function is

$$p^u = a\left(1 - \frac{Q}{x}\right)$$

where $Q = \sum_{k=1}^{n} q_k$. Firm j will then charge $p_j = a\left(1 - \frac{Q}{x}\right)\delta_j$, yielding the profit function

$$\pi_j = \left[a\left(1 - \frac{Q}{x}\right)\delta_j - c\right]q_j$$

In equilibrium,[137] each firm will maximize profits when

137. Because the profit function is a third degree polynomial with respect to q_j, the first order conditions capture both relative maxima and minima. The additional constraints implemented below will eliminate the solutions involving relative minima.

$$\begin{cases} \dfrac{\partial}{\partial q_1}\pi_1 = \dfrac{a}{x^2}\left[3q_1^{*2} + 2q_1^*(Q_{-1} - x)\right] + c = 0 \\ \vdots \\ \dfrac{\partial}{\partial q_n}\pi_n = \dfrac{a}{x^2}\left[3q_n^{*2} + 2q_n^*(Q_{-n} - x)\right] + c = 0 \end{cases}$$

Solving the above system of n first order conditions yields solutions of the form $q_j^* \in \{q_A, q_B\}$ where i firms will optimize at q_A and $(n - i)$ firms will optimize at q_B $\forall i: 0 \leq i \leq n$. This follows from the fact that the n first order conditions are symmetrical quadratics. Thus, there will be $\binom{n}{i}$ solutions for each value of i and a total of 2^n unconstrained solutions.

To obtain values for q_A and q_B, the system of x first order conditions can be reduced to

$$\begin{cases} 3aq_A^2 + 2aq_A[(i-1)q_A + (n-i)q_B - x] + cx^2 = 0 \\ 3aq_B^2 + 2aq_B[iq_A + (n-i-1)q_B - x] + cx^2 = 0 \end{cases}$$

which has the solution

$$\begin{cases} q_A^* = x\left[\dfrac{a + \sqrt{a[a - c(2i+1)(2n-2i+1)]}}{a(2n+1)}\right] \\ q_B^* = x\left[\dfrac{a - \sqrt{a[a - c(2i+1)(2n-2i+1)]}}{a(2n-2i+1)}\right] \end{cases}$$

Additionally, there may be corner solutions in which $q_A = 0$ is an optimal response to some $q_B > 0$ and vice versa.[138]

In order for the solutions to make economic sense, they must conform to the constraints set forth above as well as the requirement that both output and profits are positive for all firms:

$$\begin{cases} q_B^* > 0 \\ \pi_B^* > 0 \\ i = 0 \end{cases} \text{or} \begin{cases} q_A^* > 0 \\ q_B^* > 0 \\ \pi_A^* > 0 \\ \pi_B^* > 0 \\ 0 < i < n \end{cases} \text{or} \begin{cases} q_A^* > 0 \\ \pi_A^* > 0 \\ i = n \end{cases}$$

except for corner solutions where

138. Both interior and corner solutions are Nash equilibria. That is, each firm produces at the most profitable level given the production decisions of every other firm, leaving no firm with an incentive to deviate from its equilibrium output.

$$\begin{cases} (q_A^* > 0 \cap \pi_A^* > 0) | q_B = 0 \\ q_B, \pi_B \notin \{s: s \in \mathbb{R} \wedge s > 0\} | q_A^* > 0 \end{cases}$$

or

$$\begin{cases} (q_B^* > 0 \cap \pi_B^* > 0) | q_A = 0 \\ q_A, \pi_A \notin \{s: s \in \mathbb{R} \wedge s > 0\} | q_B^* > 0 \end{cases}$$

and $0 < i < n$.

It turns out that the only valid interior solutions occur when $i = n$ (all firms produce at q_A^*) and $0 < \frac{c}{a} < \frac{1}{(n+1)^2}$. There also exist corner solutions where g firms choose to produce at $q_g = \frac{x\left(a+\sqrt{a(a-c(2g+1))}\right)}{a(2g+1)}$ and $(n - g)$ firms choose not to produce when $\frac{7-4\sqrt{3}}{\left(g+4-2\sqrt{3}\right)^2} < \frac{c}{a} < \frac{1}{(g+1)^2}$.

The different restrictions on values of $\frac{c}{a}$ for interior and corner solutions has important economic implications. The term $\frac{c}{a}$ can be thought of as the industry cost structure, as it represents the industry marginal cost relative to the maximum amount anyone would pay for the product. It falls on the range (0, 1), which can be divided into four distinct regions.

At sufficiently low industry cost structures, only interior solutions are possible or, put differently, all firms will choose to produce. From an economic interpretation, even if all $g < n$ of the producing firms in a potential corner solution could credibly threaten to set their output as if the other $(n - g)$ firms had left the market, the other firms would still find it profitable to enter as a result of sufficiently low costs. This in turn would induce the g firms to reduce their actual from their threatened output until the entire industry wound up at the interior equilibrium.

When industry costs exceed a certain level, a credible threat by g firms to set output at the g-firm oligopoly level will deter the remaining firms from entry. No positive amount of production by those $(n - g)$ firms will yield positive profits for them, resulting in corner solutions. At the same time, an interior equilibrium in which all firms set output equal to q_A^* remains feasible.

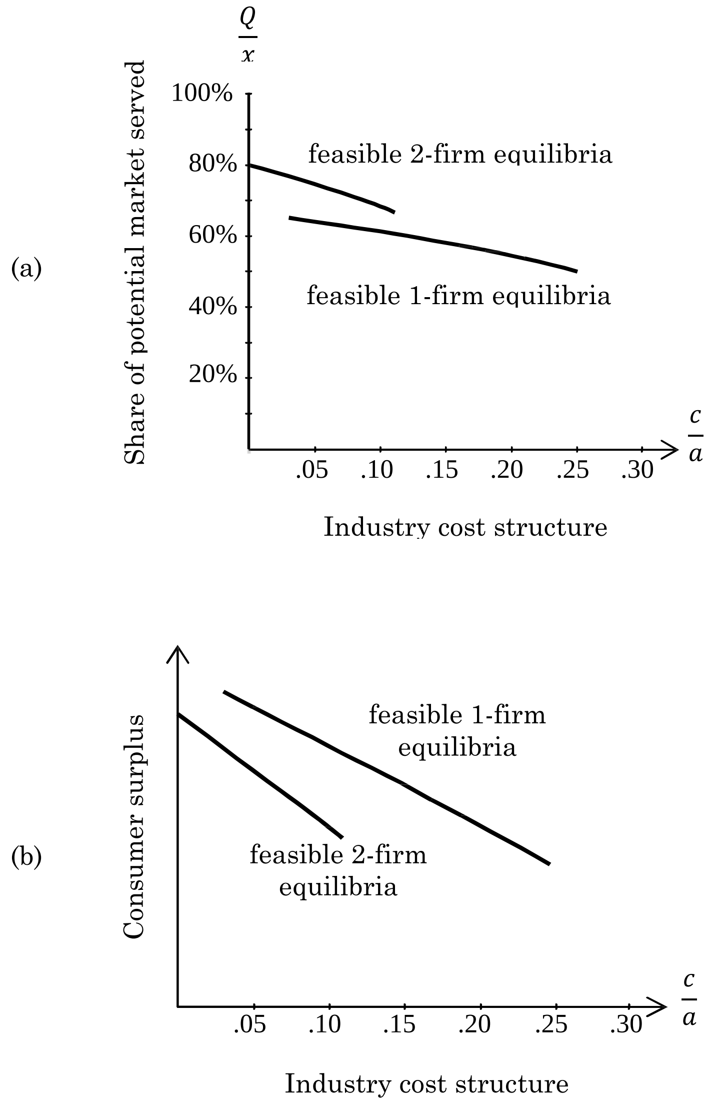
As the cost structure continues to rise, at some point the viability of interior solutions will end while at least some corner solutions will continue to be possible. The number of firms a given cost structure can support will decrease as the cost structure increases until only one firm can profitably produce. As the cost structure increases, firms can only recoup their greater costs by charging a higher price, which in turn can only be supported by a sufficiently large network. When the cost structure exceeds a certain point, $\frac{c}{a} > \frac{1}{4}$, not even a monopoly firm can make a profit and the only possible outcome will be for all firms to shut down.

Figure 6 illustrates these regions and the corresponding shares of the total market served as well as consumer surplus for the simple case where $n = 2$. To borrow a physics analogy, Figure 6 resembles a critical phase transition: over a certain range, either of two potential outcomes may occur, the actual one reached a result of path-dependence.[139] Left to the invisible hand, the market could settle at the socially inferior outcome, however defined.[140]

139. *See generally* PHILIP BALL, CRITICAL MASS: HOW ONE THING LEADS TO ANOTHER (2004) (discussing how phase transitions, path-dependence, and other concepts from the field of physics relate to social science phenomena, including economic and legal applications).

140. *Cf*. Rubinfeld, *supra* note 120, at 862-63 (describing a Pareto inferior outcome in which market forces cause a dominant network firm to drive a rival with superior technology out of the market because of incompatible standards). Often when two firms occupy the market, they will compete fiercely and the equilibrium can shift suddenly and dramatically, a phenomenon known as "tipping." *See id.* at 865-66.

Figure 6.



By comparison, the traditional *n*-firm Cournot model would not permit corner solutions. As long as all firms face the same constant marginal cost, it follows that if one firm finds it profitable to produce then all firms will find it profitable. In this model of network industry competition, nonproducing firms could be thought of as potential entrants that become actual entrants if the costs structure is permissively low.

B. *Accounting for Residual Value*

One deficiency in the model thus far presented is that it fails to account for residual value—that is, it treats a product produced by a network with zero users as worthless, having no intrinsic value apart from its network benefits. This extreme assumption rarely if ever holds true in the real world. As an illustration, consider a word processing program. It qualifies as a network product because its value to each user increases with the total number of users. However, even if no one uses the program, it may still offer value to a potential purchaser in that it performs a useful function.

Residual value will vary considerably from product to product. For example, an airline network will offer a very high residual value to customers. Although passengers no doubt appreciate some benefits of a larger network,[141] their immediate concern is to travel on a particular itinerary. Depending on their elasticity of demand, air travelers will sacrifice the conveniences of a larger network to some degree for a lower fare. Furthermore, negative network externalities may offset the positive ones, to which anyone who has ever been trapped in a middle seat on a crowded flight can attest.

Computer operating systems offer a moderate degree of residual value. Although the products offer functionality regardless of network size, the number of other users of a particular piece of software is an important consideration. Learning to operate multiple software programs where one would suffice results in a costly waste of time.

At the other end of the spectrum, credit cards typically offer very little residual value to users. A credit card accepted by only one vendor provides extremely limited utility to consumers. Although such cards do exist, vendors must usually

141. Most of these network benefits are indirect, such as frequency of flights, fewer connections, nearby alternative airports in case of service disruptions, better frequent flyer reward programs, etc.

dispense them for free or even with compensation for their use, i.e., at a negative price. Retailers typically offer a discount on the first purchase made with their store credit card as an incentive to acquire it. In contrast, large credit card networks like MasterCard and Visa frequently impose an annual fee and high annual percentage rate on their card members, who willingly incur these costs because of the widespread acceptance of the cards.

Because residual value is a factor of variable importance to demand for network products, it is desirable to capture this effect in the model and to examine its effects. Let θ denote a product's residual value, which can take values on the range [0, 1). In this more robust model,

$$\delta_j = \theta + \frac{q_j}{x}(1-\theta)$$

In the case where $\theta = 1$, no network effects are present and the model collapses into the garden variety Cournot setup. When $\theta = 0$, the product has no residual value, exactly as in the basic model presented above.

Similar to the basic model, interior equilibrium solutions are characterized by i firms producing at q_C and $(n-i)$ firms producing at q_D such that in equilibrium

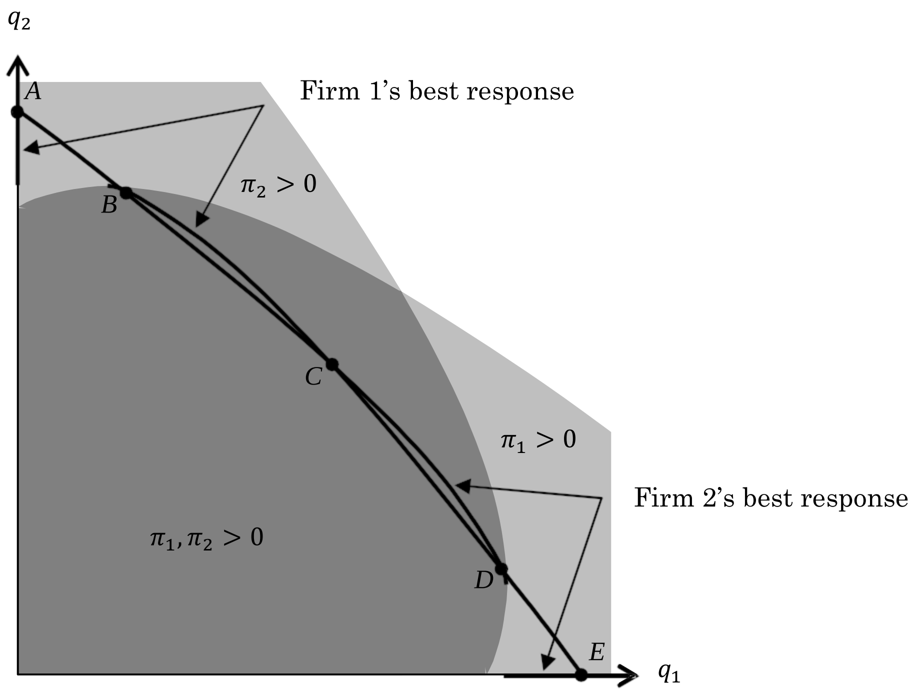
$$\begin{cases} q_C^* = x\left(\dfrac{a\big(2-\theta(3+2i-n)\big)+\sqrt{a^2(\theta(n-2)+2)^2+\big(a^2\theta^2-4ac(1-\theta)\big)(2i+1)(2n-2i+1)}}{2a(2i+1)(1-\theta)}\right) \\ q_D^* = x\left(\dfrac{a\big(2-\theta(3-2i+n)\big)-\sqrt{a^2(\theta(n-2)+2)^2+\big(a^2\theta^2-4ac(1-\theta)\big)(2i+1)(2n-2i+1)}}{2a(2n-2i+1)(1-\theta)}\right) \end{cases}$$

Unlike in the basic model, valid interior equilibria can exist, subject to certain constraints,[142] for values of i greater than zero and less than or equal to n. In other words, this more robust model has the potential for more interior solutions than the basic model. It also allows for corner solutions.

For example, a two-firm industry could reach one of five different equilibria under certain parameter values. Figure 7 illustrates this for the case when $\frac{c}{a} = .16$ and $\theta = .28$.

142. These constraints are set forth in the Appendix.

Figure 7



There are three interior solutions. At point C, each firm sells to approximately 28.4% of the potential market, x. At point D, firm 1 sells to approximately 44.1% of the potential market and firm 2 sells to 9.6%. Point B is the reverse of point D. Point C is not Pareto superior to either B or D—one firm will realize a greater profit and one a lesser profit than when their output is unequal.

Additionally, there are two corner solutions, points A and E, where one firm produces at the monopoly level, selling to approximately 51.5% of the potential market, and the other firm chooses not to produce. The non-producing firm cannot make a profit at any positive level of output given the producing firm's output choice.

For the purposes of the next section—evaluating the effect of mergers on consumer welfare—it is helpful to limit the number of potential equilibria. For that reason, it will be assumed that in the initial state all firms set output identically, equal to q_C and corresponding to point C above.

This has the advantage of requiring the fewest restrictions on parameter values. In particular, it is the only possible interior equilibrium when $\frac{2}{3} < \theta < 1$.

C. *The Effect of Mergers on Consumer Welfare*

The best way to evaluate the welfare of consumers under given market conditions is to measure the surplus S that they realize.[143] This model allows for two different prices concurrently in equilibrium, owing to the different discounts consumers apply to their willingness to pay for networks of unequal size. To evaluate consumer surplus, I calculate the aggregate consumer surplus at undiscounted prices and apply the discount for each network size in proportion to each network's share of the total output. In other words,

$$S = \int_{p^u|Q=Q^*}^{p^u|Q=0} D^{-1}(p^u)dp\left(\delta_C \frac{iq_C^*}{iq_C^* + (n-i)q_D^*} + \delta_D \frac{(n-i)q_D^*}{iq_C^* + (n-i)q_D^*}\right)$$

or

$$S = \frac{a}{2x^2}\left[\left(iq_C^* + (n-i)q_D^*\right)\left(iq_C^*(x\theta + (1-\theta)q_C^*) + (n-i)q_D^*(x\theta + (1-\theta)q_C^*)\right)\right] \qquad [1]$$

As discussed in the preceding section, this Article will investigate only equilibrium solutions where all firms produce q_C^* units of output. Equation [1] can then be simplified by setting i equal to n, yielding

$$S = \frac{a}{2x^2}(nq_C^*)^2(x\theta + (1-\theta)q_C^*) \qquad [2]$$

The expanded equation will still prove useful for evaluating consumer surplus in the short term.

The model thus far presented predicts the outcome of competition in network industries for a given number of firms facing a given cost structure. It provides an illustration of the competitive landscape in the steady state that precedes a merger, but offers little insight as to the market structure that results following a merger. This Article will consider two simple cases of the post-merger structure—one short term and one long term. These two cases present two extremes of the possible structural outcomes of a merger.

143. *See supra* pp. 133-34.

In the immediate aftermath of a merger, the overall production in the market remains substantially the same, the major difference being that one larger firm produces the amount of output previously contributed by two smaller firms. The Merger Guidelines take essentially a short run view in their application of the HHI.[144] In the long run, the merged firm will likely cut back on production to some extent, depending on demand elasticities. Other firms in the industry may increase their output.[145] In the most sanguine of worlds (from the consumer's standpoint), the remaining firms will eventually end up once more with roughly equal market shares.

Accordingly, the short run case assumes that one merged firm produces output equal to $2q_C^*$ while the remaining $(n - 2)$ firms continue to produce output equal to q_C^*. The long run case assumes that after the merger all $(n - 1)$ firms produce at q'^*_C where $i' = n' = n - 1$. In both cases, the pre-merger consumer surplus is exactly the amount from equation [2].

For the short run case, we can utilize equation [1], letting $q_D^* = 2q_C^*$. The post-merger consumer surplus therefore is

$$S' = \frac{a}{2x^2}[n(q_C^*)^2(nx\theta + (n+2)(1-\theta)q_C^*)] \qquad [3]$$

Subtracting [2] from [3] yields the quantity

144. In considering the change in HHI the Guidelines assume static market shares follow the merger and do not incorporate a dynamic equilibrium analysis. *See supra* p. 114 and note 50.

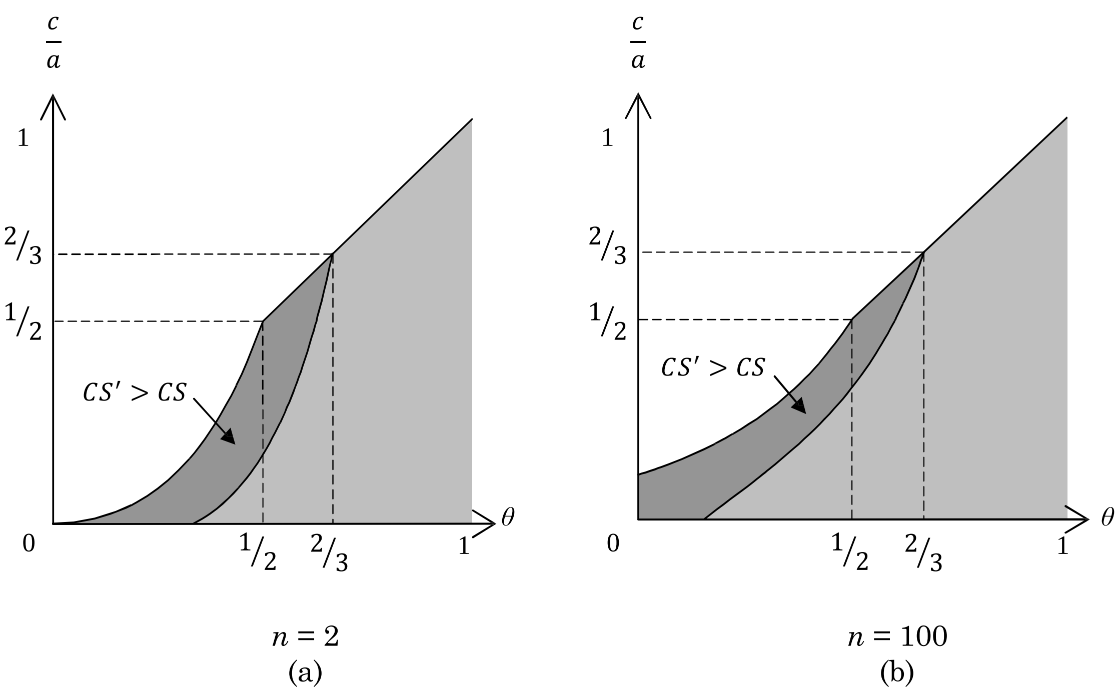
145. For empirical evidence, see John R. Baldwin & Paul K. Gorecki, *Mergers Placed in the Context of Firm Turnover*, *in* 5 BUREAU OF THE CENSUS, PROCEEDINGS OF THE ANNUAL RESEARCH CONFERENCE 53 (1990) (finding significant declines in market shares for plants acquired in horizontal mergers); Klaus Gugler et al., *The Effects of Mergers: An International Comparison*, 21 INT'L J. INDUS. ORG. 625 (2003) (considering international data and finding consistently reduced sales among merging firms); Dennis C. Mueller, *Mergers and Market Share*, 67 REV. ECON. & STAT. 259 (1985) (conducting large-scale study that found merging firms lost market share relative to non-merging firms); Martin Pesendorfer, *Horizontal Mergers in the Paper Industry*, 34 RAND J. ECON. 495 (2003) (finding that 74.1% of merging firms lost market share).

$$\frac{a}{2x^2}[n(q_C^*)^3(1-\theta)] \qquad [4]$$

which is always positive. Therefore, in the short run case post-merger consumer surplus will always exceed pre-merger surplus by the amount in equation [4] when two identically sized network firms combine.

The long run case offers similar but more qualified results. For most cost structures, consumer surplus will increase after a network industry merger, given a sufficiently low residual value. Figure 8 provides an illustration.

Figure 8.



The shaded areas represent feasible parameter combinations (i.e., those that produce profitable outcomes before and after the merger). The darker regions indicate parameter values for which the post-merger consumer surplus exceeds the pre-merger surplus.

This is not to say that mergers involving products with sufficiently low residual value will evince unequivocally good results; merely that such mergers will result in increased consumer surplus. There may of course be important considerations other than immediate welfare maximization, as will be discussed in the next part.

Regardless of the relative merits of welfare maximization versus other objectives, the long-term equilibrium case highlights the importance of proper merger evaluation by regulators. Although in many circumstances network industry mergers will increase consumer surplus, in many other circumstances such mergers will not be desirable from a public policy perspective.

V. A Suggested Approach for Incorporating Network Effects into Merger Evaluation

The usefulness of HHI analysis varies with the degree to which an industry exhibits network externalities. Prior to their HHI analysis but subsequent to market definition, antitrust regulators should attempt to determine the residual value of the relevant products. In terms of the model above, regulators should attempt to estimate θ. The greater the residual value, the less impact network externalities will have on post-merger consumer welfare.

Figure 8 also suggests a second consideration: cost. All else equal, a network industry with a greater marginal cost structure is more likely to experience an increase in consumer welfare from consolidation. When the number of firms in a network industry decreases, the industry output decreases, the price increases and the output per firm increases. The first two of these phenomena tend to decrease consumer surplus while the third tends to increase it, making the overall change in surplus uncertain. However, as the cost structure increases, the proportional increase in output per firm increases whereas the proportional decrease in industry output and increase in price both diminish in magnitude. As a result, the change in consumer surplus from industry consolidation—although not necessarily positive for all values of marginal cost—will increase with the marginal cost structure.

It is important then to understand both the amount of residual value attributable to a product and its marginal cost of production when evaluating network industry mergers. The higher the product's residual value and the lower its cost, the more the industry resembles a "traditional" industry and the more insight a subsequent HHI analysis will reveal. On the other hand, the HHI will provide scant guidance when evaluating mergers involving products with low residual values and high marginal costs of production.

The HHI may still play a useful role in the regulation of mergers in which network effects dominate. As acknowledged at the outset, increased concentration may influence industry behavior in anticompetitive ways other than through the price mechanism. For instance, a lack of competition could in theory suppress innovation, which would ultimately weigh on consumer utility.[146] Moreover, as networks grow large, firms can attempt to stifle competition by reducing the interoperability between their dominant networks and those of smaller rivals.[147]

Mergers will also have the result of suppressing total market output and raising price, similar to the difference

146. *See* JOHN E. KWOKA, JR. & LAWRENCE J. WHITE, THE ANTITRUST REVOLUTION: ECONOMICS, COMPETITION, AND POLICY 420-21 (4th ed. 2004) [hereinafter THE ANTITRUST REVOLUTION]. Empirical evidence in a non-network context suggests otherwise. *See* Richard Blundell et al., *Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms*, 66 REV. ECON. STUD. 529 (1999).

147. For example, in the U.S. government's monopolization suit against Microsoft, the government alleged, inter alia, that Microsoft had leveraged its considerable market power and reduced interoperability. For an illuminating discussion of the case and network issues involved, see Daniel L. Rubinfeld, *Maintenance of Monopoly:* U.S. v. Microsoft, *in* THE ANTITRUST REVOLUTION, *supra* note 146, at 476. Preventing network mergers that substantially increase concentration may combat such anticompetitive behavior. However, given the unstable and rapidly changing competitive dynamics in network industries, merger policy alone may not adequately guard against such concerns.

between the interior and corner solution outcomes illustrated in Figure 6 above. To the extent government policy favors low prices and market access over consumer surplus, even mergers with substantial network benefits may invite scrutiny.

For instance, if the network in question involved a telecommunications service, a policymaker might wish to maximize the number of consumers with access to some network, regardless of size. One means of achieving that outcome would be to prohibit large telecommunications mergers. A more efficient outcome would nonetheless allow consolidation; the government could increase network access through a subsidy or by creating a legal monopoly that could price discriminate, if possible.[148]

Although an unconsolidated industry in some sense compensates consumers with lower prices, society as a whole shoulders the opportunity cost of using smaller, less effective networks rather than larger ones with critical mass. The pre-merger equilibrium often will not capture the bulk of the positive network externalities. In any event, the government should acknowledge the HHI's inherent limitations in the network context regardless of extent to which it continues to rely on the index.

The DOJ and FTC could address these network issues in a variety of ways, ranging from a radical transformation of the current horizontal merger evaluation process to a minor tweaking of the existing guidelines. On one end of the spectrum, the agencies could choose to completely revamp their processes and adopt a dynamic model, such as the one presented above, which incorporates the concept of residual value.[149] If such a model indicated a lack of significant network effects or, alternatively, if it produced an inconclusive or insubstantial prediction of changes in consumer welfare, regulators might then use the HHI as a tiebreaker.

148. *See* PEPALL ET AL., *supra* note 29, at 620.

149. *See* Farrell & Shapiro, *supra* note 113, at 107-08 (calling for a dynamic "equilibrium analysis" rather than the static structural model embodied in the HHI, although not specifically addressing the issue of network industries or residual value).

Residual value, while not commonly calculated at present, should not present too difficult a challenge for government economists. Standard econometric techniques could easily estimate its value for a product made by several different competitors. Even in markets with a high degree of concentration and thus fewer data points, economists could still obtain a reasonable approximation of residual value by augmenting the data set with data from other time periods and comparable industries.

The more intractable problem, from an empirical perspective, is how to model the discount function. Even if regulators could estimate a product's value both when no one uses it and when everyone does, they still need to devise a function, which estimates a product's value for all intermediate network sizes. The model above assumed a linear relationship, but that may only infrequently hold true. In fact, the function may not even be strictly increasing, but may reach a maximum and then start to decrease, as in the fashion industry example.[150]

Given the time and other constraints faced by government regulators, such economic modeling may prove impractical. The discount function likely will differ substantially from industry to industry, necessitating a longer period of time to review mergers, reducing transparency and running the risk of reaching inconsistent results. Regulators could nonetheless incorporate a network effects analysis into their current practices.

Residual value is neither an abstruse concept nor difficult to approximate without resort to statistical data. If the antitrust authorities examine a merger in which the relevant product has a low residual value, they should proceed with their HHI analysis knowing it presents only one side of the post-merger landscape. Network effects, if not formally modeled prior to the HHI analysis, should at a minimum receive treatment similar to efficiencies—as an offset to a presumptively anticompetitive increase in the HHI score. The Merger Guidelines could easily adopt such a change without abandoning their existing framework.

150. *See supra* p. 133.

VI. Conclusion

Despite the allure of efficiencies and market dominance, one wonders why firms would choose to merge in the first place, given that mergers frequently prove unprofitable for the acquiring company.[151] Nonetheless, mergers among network firms can bring about substantial consumer benefits, even after taking into account the likely price increases from more concentrated market power. At the moment, the Merger Guidelines give short shrift to such network benefits, and their reliance on the HHI lies at the core.

The HHI serves a useful role in measuring changes in industrial concentration. In many industries it provides a reasonable initial indication, prior to an efficiencies calculation, of harm to consumers from potential consolidation. In network industry mergers, however, the presumption of harm from concentration obfuscates rather than clarifies the situation.

The strength of the DOJ and FTC's approach to the Merger Guidelines over the past thirty years has been their willingness to modify the Guidelines to better reflect economic realities. The beneficial effect of large networks to consumers is well documented and our antitrust regulators should correspondingly adjust their approach.

151. *See* Gregor Andrade et al., *New Evidence and Perspectives on Mergers*, 15 J. ECON. PERSP. 103 (2001).

Appendix: Parameter Constraints for Interior Solutions

General constraints: $\begin{cases} q_A^*, q_B^*, \pi_A^*, \pi_B^*, x > 0 \\ 0 < c < a \\ 0 \leq \theta < 1 \end{cases}$

For $0 \leq i \leq \frac{x-1}{4}$ there is no interior solution.

For $\frac{x-1}{4} < i \leq \frac{n}{2}$, there are interior solutions when

$$\begin{cases} \dfrac{2n-4i}{(2n-2i+1)(2i+1)-(n-2)(n-2i)} < \theta \leq \dfrac{1}{i+2} \\ \dfrac{(1+(n-i-1)\theta)^2}{(n+1)^2(1-\theta)} < \dfrac{c}{a} \leq \dfrac{(2(1-\theta)+n\theta)^2+\theta^2(2i+1)(2n-2i+1)}{4\big(2n+1+4i(n-i)\big)(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{1}{i+2} < \theta \leq \dfrac{2}{n-2i+3} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} \leq \dfrac{(2(1-\theta)+n\theta)^2+\theta^2(2i+1)(2n-2i+1)}{4\big(2n+1+4i(n-i)\big)(1-\theta)} \end{cases}$$

For $\frac{n}{2} < i \leq \frac{n-1+(n+1)\sqrt{3}}{4}$, there are interior solutions when

$$\begin{cases} 0 < \theta \leq \dfrac{4i-2n}{(2n-2i+1)(2i+1)+(n-2)(n-2i)} \\ \dfrac{(1+(n-i-1)\theta)^2}{(n+1)^2(1-\theta)} < \dfrac{c}{a} < \dfrac{(1+(i-1)\theta)^2}{(n+1)^2(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{4i-2n}{(2n-2i+1)(2i+1)+(n-2)(n-2i)} < \theta \leq \dfrac{1}{i+2} \\ \dfrac{(1+(n-i-1)\theta)^2}{(n+1)^2(1-\theta)} < \dfrac{c}{a} \leq \dfrac{(2(1-\theta)+n\theta)^2+\theta^2(2i+1)(2n-2i+1)}{4\big(2n+1+4i(n-i)\big)(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{1}{i+2} < \theta \leq \dfrac{2}{2i-n+3} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} \leq \dfrac{(2(1-\theta)+n\theta)^2+\theta^2(2i+1)(2n-2i+1)}{4\big(2n+1+4i(n-i)\big)(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{2}{2i-n+3} < \theta \leq \dfrac{2}{3} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} < \dfrac{\theta(1+(n-i-1)\theta)}{(2n-2i+1)(1-\theta)} \end{cases}$$

For $\frac{n-1+(n+1)\sqrt{3}}{4} < i \leq \frac{3n+1}{4}$, there are interior solutions when

$$\begin{cases} 0 < \theta \leq \dfrac{1}{i+2} \\ \dfrac{(1+(n-i-1)\theta)^2}{(n+1)^2(1-\theta)} < \dfrac{c}{a} < \dfrac{(1+(i-1)\theta)^2}{(n+1)^2(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{1}{i+2} < \theta \leq \dfrac{4i-2n}{(2n-2i+1)(2i+1)+(n-2)(n-2i)} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} < \dfrac{(1+(i-1)\theta)^2}{(n+1)^2(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{4i-2n}{(2n-2i+1)(2i+1)+(n-2)(n-2i)} < \theta \leq \dfrac{2}{2i-n+3} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} \leq \dfrac{(2(1-\theta)+n\theta)^2+\theta^2(2i+1)(2n-2i+1)}{4\big(2n+1+4i(n-i)\big)(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{2}{2i-n+3} < \theta < \dfrac{2}{3} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} < \dfrac{\theta(1+(n-i-1)\theta)}{(2n-2i+1)(1-\theta)} \end{cases}$$

For $\frac{3n+1}{4} < i < n$, there are interior solutions when

$$\begin{cases} 0 < \theta \leq \dfrac{1}{i+2} \\ \dfrac{(1+(n-i-1)\theta)^2}{(n+1)^2(1-\theta)} < \dfrac{c}{a} < \dfrac{(1+(i-1)\theta)^2}{(n+1)^2(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{1}{i+2} < \theta \leq \dfrac{1}{n-i+2} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} < \dfrac{(1+(i-1)\theta)^2}{(n+1)^2(1-\theta)} \end{cases}$$

or

$$\begin{cases} \dfrac{1}{n-i+2} < \theta < \dfrac{2}{3} \\ \dfrac{\theta(1+(i-1)\theta)}{(2i+1)(1-\theta)} < \dfrac{c}{a} < \dfrac{\theta(1+(n-i-1)\theta)}{(2n-2i+1)(1-\theta)} \end{cases}$$

For $i = n$, there are interior solutions when

$$\begin{cases} 0 \leq \theta \leq \dfrac{1}{2} \\ 0 \leq \dfrac{c}{a} < \dfrac{(1+(i-1)\theta)^2}{(n+1)^2(1-\theta)} \end{cases}$$

or

$$\begin{cases} \frac{1}{2} < \theta < 1 \\ 0 \leq \frac{c}{a} < \theta \end{cases}$$