Policy Implications of Endogenous Sunk Fixed Costs in Banking: Has U.S. Antitrust Policy Been on the Wrong Track?
David VanHoose

Abstract: The application of U.S. antitrust policy toward mergers in the banking industry is based on past research suggesting that there is a trade-off between adverse effects on consumer welfare owing to potentially augmented market power and possible welfare-improving effects arising from efficiency gains. Recent research applying the theory of endogenous sunk fixed costs to the banking industry suggests that the current focus of bank antitrust policy on this trade-off may be misplaced. According to initial results obtained by work along these lines, the presence of endogenous sunk fixed costs arising from non-price competition among banks on the basis of product quality yields a lower bound on concentration in banking markets, implying that a few large banks typically will predominate. This policy brief provides an overview of the current basis of U.S. antitrust policy with regard to bank mergers, outlines the theory of endogenous sunk fixed costs, explains why and how the theory might apply to the banking industry, and evaluates whether antitrust policy should be altered in light of evidence offered to date.

About the Author: David VanHoose is a Senior Fellow at Networks Financial Institute. VanHoose earned his Ph.D. at the University of North Carolina at Chapel Hill and is currently the Herman W. Lay Professor of Private Enterprise and Professor of Economics at the Hankamer School of Business at Baylor University. He has published articles in such professional journals as the Quarterly Journal of Economics, the International Economic Review, the Southern Economic Journal, the Journal of Money, Credit, and Banking, Economic Inquiry, and the Scandinavian Journal of Economics.

Keywords: Banking antitrust policy; endogenous sunk fixed costs; regulatory compliance costs.

JEL Codes: G28, L41, L13

The views expressed are those of the individual author and do not necessarily reflect official positions of Networks Financial Institute. Please address questions regarding content to David VanHoose at David_VanHoose@baylor.edu. Any errors or omissions are the responsibility of the author.

NFI working papers and other publications are available on NFI’s website (www.networksfinancialinstitute.org). Click “Research” and then “Publications/Papers.”
Policy Implications of Endogenous Sunk Fixed Costs in Banking: Has U.S. Antitrust Policy Been on the Wrong Track?
David VanHoose

1. Introduction

The application of U.S. antitrust policy toward mergers in the banking industry is based on past research suggesting that there is a trade-off between adverse effects on consumer welfare owing to potentially augmented market power and possible welfare-improving effects arising from efficiency gains. Recent research applying the theory of endogenous sunk fixed costs to the banking industry suggests that the current focus of bank antitrust policy on this trade-off may be misplaced. According to initial results obtained by work along these lines, the presence of endogenous sunk fixed costs arising from non-price competition among banks on the basis of product quality yields a lower bound on concentration in banking markets, implying that a few large banks typically will predominate. Furthermore, the theory suggests that if greater concentration created by a merger results in higher prices of financial services, the reason that prices are higher is not necessarily an expansion of banks’ market power. Instead, larger post-merger firms may provide higher-quality services that require them to incur higher fixed costs, which in turn necessitate higher prices in a long-run competitive equilibrium. Consumers, the theory suggests, are willing to pay for these quality enhancements because quality improvements raise consumer welfare. Thus, in many instances—perhaps more than permitted under current bank antitrust policy — bank mergers are likely to be welfare-enhancing.

This policy brief provides an overview of the current basis of U.S. antitrust policy with regard to bank mergers, outlines the theory of endogenous sunk fixed costs, explains why and how the theory might apply
to the banking industry, and evaluates whether antitrust policy should be altered in light of evidence offered to date. The main conclusions are as follows. First, the theory of how endogenous sunk fixed costs affect market structure in many industries is both interesting and persuasive. Second, there are compelling reasons to view the banking industry as one to which the theory might well apply. Third, recent research by Dick (2007) provides strong circumstantial evidence favoring the applicability of the theory to banking, along with the subsidiary implication that current antitrust policy fails to give adequate attention to the potential quality-boosting impacts of bank mergers. Fourth, fixed costs of regulatory compliance undoubtedly emerge endogenously as an outcome of a dynamic regulatory interaction between the interests of regulated banks and consumers, yet the role of this source of endogenous sunk fixed costs as an influence on bank market structure has not been considered to date. Fifth and finally, until these regulatory costs are taken fully into account in a full assessment of the applicability of the theory to the structure of the banking industry, it would be unwise for policymakers to alter their present merger guidelines, aside perhaps from incorporating greater attention to the implications of mergers for the quality of banking products.

2. Banking Antitrust Policy in the United States

Traditional policy-oriented economic analysis of the industrial structure of the banking industry has been guided by two intellectual paradigms: the structure-conduct-performance (SCP) hypothesis and the efficient-structure (ES) theory. The SCP hypothesis proposes that the level of concentration in a banking market influences banks’ conduct, which in turn has a bearing on loan and deposit quantities, qualities, interest rates,
and other market outcomes that determine consumer welfare. Specifically, the SCP hypothesis suggests that greater concentration gives banks more market power, which in turn leads to fewer loans and deposits and higher loan rates and lower deposit rates, all of which reduce consumer welfare.

The ES theory suggests that cost conditions faced by banks play a crucial role in determining the optimal scale of individual banking organizations and the appropriate scope of banking activities. Thus, in contrast to the SCP hypothesis, the ES theory indicates that greater cost efficiencies resulting from expansions of scale and/or scope can lead both to expansions in loans and deposits, with associated lower loan rates and higher deposit rates.

In practice, therefore, regulators contemplating proposed bank mergers and acquisitions have focused considerable attention on a perceived trade-off between resulting increases in market power versus cost-efficiency gains. In an effort to find a general point of balance along this trade-off, U.S. government and Federal Reserve authorities have developed a set of rules for evaluating the merits of proposed banking mergers.

Under terms of the Bank Merger Act of 1966, any contemplated banking merger must first be proposed to the applicable banking regulator (or regulators in cases in which the planned merger involves banks supervised by different regulatory agencies). If the applicable regulator renders a favorable judgment, then the merger cannot be completed for another 30 days, pending potential further review by the U.S. Department of Justice and private parties possessing legal standing to challenge a merger under U.S. antitrust laws. If no legal challenges arise by the end of this 30-day waiting period, then parties to the merger can engage in the proposed consolidation. At any point, however, banking regulators or the Department
of Justice can act to avert the proposed consolidation.

According to procedures developed by the Federal Reserve’s Board of Governors (Walter and Wescott, 2008; American Bar Association, 2007) each of the twelve Federal Reserve banks defines the relevant banking markets—clearly delineated, distinct markets for banking services—to be utilized in evaluating effects of any proposed mergers within the geographic area encompassed by its own Federal Reserve district boundaries. In keeping with the U.S. Supreme Court’s 1963 decision emphasizing the importance of local market considerations, staff economists and other officials at Federal Reserve banks, in defining the relevant banking market, emphasize geography in their delineations. Thus, as an initial estimate, they typically rely considerably on the U.S. Office of Management and Budget’s so-called “MSAs”—Micropolitan Statistical Areas containing at least one urban area and a population between 10,000 and 49,999 people and Metropolitan Statistical Areas containing at least one urban area and a population of 50,000 or more people—and on Rand McNally’s Ranally Metropolitan Areas, or “RMAs,” containing at least 70 people per square mile with at least 20 percent of its labor force commuting to a defined central urban location. In some cases, the Federal Reserve takes into account political county boundaries as well.

Once a geographic region has been identified as the relevant banking market, the relevant item sold in that market by banks must be identified. The 1963 Supreme Court ruling determined that item to be “the cluster of products and services” offered by banks but did not specify a definition of the appropriate “cluster” to be considered in banking antitrust analysis. In practice, banking regulators and the Department of Justice have chosen to utilize bank deposits as a surrogate measure of the appropriate cluster of
banking services. The Federal Reserve analyzes concentration of the relevant market utilizing data on deposits that banks report as of June 30 each year. It excludes deposits at banking institutions specializing solely in credit card lending, because these institutions raise their deposit funds in nationwide markets.

To assess the degree of concentration in the relevant banking market, the Federal Reserve follows the Department of Justice by relying on the Herfindahl-Hirschman index (HHI), the sum of squared percentage market shares of all banks in the market. It computes both pre-merger and post-merger HHIs, and it also calculates the resulting change in the HHI, which in the case of a merger of two rivals turns out to equal twice the product of percentage market shares of the two rivals. (The pre-merger HHI is equal to $\sum_{i=1}^{n} (S_i)^2$, where there are $n$ banking institutions in the relevant market and $S_i$ is the market share of a given bank $i$, so it follows that the post-merger HHI in the event of a merger by the first two banks, $i = 1, 2$, must be equal to $\sum_{i=1}^{n} (S_i)^2 - (S_1)^2 - (S_2)^2 + (S_1 + S_2)^2 = \sum_{i=1}^{n} (S_i)^2 + 2S_1S_2$; thus, subtracting the pre-merger HHI, $\sum_{i=1}^{n} (S_i)^2$, yields $2S_1S_2$.) Both the post-merger HHI and the change in the HHI are important elements in banking regulators’ assessment of a proposed merger.

In evaluating planned consolidations of nonfinancial firms, the U.S. Department of Justice considers a market to be unconcentrated if its HHI is less than 1,000, to be moderately concentrated if its HHI is between 1,000 and 1,800, and to be highly concentrated if its HHI exceeds 1,800. Figure 1 suggests that relying on these Department of Justice classifications of market concentration would imply that many U.S. banking markets are already relatively concentrated. As the figure indicates, the average HHI values for urban banking markets (defined for simplicity as MSAs) and rural banking
markets (defined as non-MSA counties) have trending very slightly downward recently. Nevertheless, average HHI values in urban markets have remained very close to the Department of Justice’s minimum level of 1,800 for defining high market concentration, and average HHI values for rural markets are well above this value.

**Figure 1: Average Herfindahl-Hirshman Index Values in U.S. Banking Markets**

[Source: Pilloff (2009)]

Indeed, Figure 1 suggests that many U.S. banking mergers involving relatively large institutions would fail to meet the Department of Justice’s screen for assessing proposed nonfinancial mergers, according to which a planned merger raises antitrust concerns if post-merger HHI level exceeds 1,800 and the change in the HHI resulting from the merger exceeds 50.

Proposed U.S. banking mergers, however, face a different screening procedure than do planned mergers of nonfinancial firms.

The U.S. bank merger screening mechanism consists of two screens. The first, known as Screen A, calculates the HHI for the relevant banking market giving 100 percent weight to the deposits of commercial banks but placing only a 50 percent weight on deposits of savings institutions, which
U.S. banking authorities perceive to provide a narrower cluster of banking products than commercial banks. This naturally reduces the effective value of the post-merger HHI that a proposed bank merger would yield. Under screen A, a proposed banking merger would raise antitrust concerns on the part of banking regulators and the Department of Justice only if the resulting post-merger \textit{weighted} HHI exceeds 1,800. In addition, in recognition of the fact that in the course of their day-to-day operations banking institutions face some competition from savings institutions and other myriad financial institutions, antitrust concerns about a proposed merger arises if the change in this weighted HHI exceeds a value of 200.

If the Screen A thresholds for the post-merger \textit{weighted} HHI and change in weighted HHI are exceeded, then the Department of Justice applies a second screen, called Screen B. This screen computes a post-merger \textit{unweighted} HHI and change in the \textit{unweighted} HHI for RMAs instead of the relevant markets as defined by the Federal Reserve. If the thresholds are again exceeded under Screen B, then antitrust policymakers consider other factors specific to the planned merger in reaching their final judgment regarding its merits. Among these so called “mitigating factors” are whether there currently is direct competition between the merging institutions, whether the merging institutions appear to specialize in providing different types of services, whether customer surveys suggest that the merger partners’ clusters of products are close substitutes, and whether the locations of their offices and customers exhibit significant geographic overlap within the relevant market.

As a matter of procedure, within the Federal Reserve System, Board of Governors can authorize a Federal Reserve bank to approve a merger application, but only the Board of Governors has authority to seek to block a
proposed consolidation. As noted above, however, even if the Federal Reserve initially approves a planned merger, the Department of Justice and private parties with legal standing can seek to block it in court under U.S. antitrust laws.

The formal merger guidelines utilized by banking regulators and the U.S. Department of Justice provide a relatively clear roadmap for banks contemplating a consolidation. Indeed, a Federal Reserve Web site, http://cassidi.stlouisfed.org/, provides access to HHI data that prospective partners can utilize to assess the prospects for approval of a contemplated merger. Nevertheless, recent research suggests that it is conceivable that the merger guidelines are based on faulty economic reasoning.

3. The Theory of Endogenous Sunk Fixed Costs: An Overview

Economists typically presume that a key long-run objective of any firm is to attain a cost-minimizing scale and scope of operations and that, given attainment of this minimum-cost objective, the firm continually adjusts its balance sheet with an aim to maximize profits. According to this perspective, a firm’s only sunk costs are costs of entry and setup that are outside of its control. A predictable consequence is that, as the size of an industry’s market expands, perhaps through growth in income or population, incumbent firms in an industry should reach their efficient scales. As profits increase, additional firms should enter and should as well pursue output expansions toward efficient scales. Thus, as market size expands, there should be an increase in the number of firms and an accompanying decline in each firm’s market share. Consequently, an increase in market size eventually should lead to a reduction in concentration.
Sutton (1991) proposes that in some industries sunk costs may be endogenous. An important consequence, Sutton suggests, is that industry structure may remain static even in a growing market. Firms in industries with endogenous sunk fixed costs, Sutton suggests, select a stream of fixed outlays on items such as research and development, advertising, or other characteristics that enhance the demands for their products. As a result, an expansion in market size encourages firms to proportionately increase their fixed expenses relating to such items in an effort to boost consumers’ willingness to pay. As firms’ fixed costs increase because of outlays aimed at boosting product demands, however, their profits decline, which removes the incentive for additional rivals to enter the market. In Sutton’s framework with endogenous sunk fixed costs, therefore, steady increases in market size ultimately generate no further increases in the number of firms. Hence, concentration eventually reaches a lower bound as market size continues to expand.

Sutton provides a three-stage theory of industry entry and competition. In the first stage, a potential competitor decides whether or not to enter the industry in light of the entry cost that would be entailed. In the second stage, each firm chooses a level of product quality to maximize its profits, taking into account the facts that a higher quality level will expand the demand for its product, that selecting a higher quality level entails a higher fixed cost, and that rival firms will be making quality decisions in the same manner. In the third stage, each firm decides how much of its product to produce with an aim to maximize product and taking into account the profit-maximizing outputs of other competitors.

Sutton shows that an end result of this three-stage process typically is a relationship of the type displayed in Figure 2. Initially, as market size
expands—that is, income increases or population rises—from a relatively low range of values, there is an increase in the number of firms that can earn sufficient accounting profits to compensate owners for the opportunity cost of allocating their resources to the particular industry in question instead of some other industry. Beyond a critical market size, however, the number of firms in the industry declines and eventually levels off at or below an upper bound. Thus, there is a lower bound on concentration in an industry in which endogenous sunk outlays on product quality are an important dimension of market rivalry. In effect, such an industry is a “natural oligopoly” that tends to settle out at a long-run equilibrium either with a relatively small number of firms when firms face similar cost structures and offer only slightly differentiated products or a few large firms and perhaps several fringe firms when firms possess different technologies and offer highly differentiated products.

**Figure 2: A Simulation of the Predicted Relationship between Market Size and the Number of Firms in an Industry with Endogenous Sunk Fixed Costs**

![Graph showing the relationship between market size and the number of firms in an industry with endogenous sunk fixed costs](image)

[Source: Shiman (2007)]
4. Applying Sutton’s Theory to the Banking Industry

Sutton’s theoretical framework applies to industries in which non-price competition is an essential feature. It has long been recognized in the banking literature that banks often compete along non-price dimensions as well as on the basis of explicit interest rates and service charges. In particular, when deposit rate ceilings were in place in the United States prior to rate deregulation in the early 1980s, numerous authors suggested that developing a framework of analysis of non-price competition was central to understanding rivalry in banking markets.

Initially, work along these lines focused on the idea that when confronted with legal deposit rate ceilings, banks offered “implicit interest” on deposits. For instance, Barro and Santomero (1972) measured the implicit deposit interest rate as a remission rate banks granted their customers, and Benjamin Klein (1974) sought to estimate a deposit rate that banks counterfactually would have paid if not prohibited from doing so by legal constraints. Becker (1975) measured the implicit deposit rate as noninterest expenses less service charges as a percentage of deposits, and this became the standard measure utilized in most other analyses, such as Mitchell (1979), Startz (1983), Merris (1985), and Bradley and Jansen (1986). As shown by VanHoose (1988), however, implicit deposit rates measured as remitted service charges arise in imperfectly competitive banking markets with or without deposit rate ceilings and must dissipate in perfectly competitive markets in which service charges fully adjust to reflect marginal costs of providing relevant services. Furthermore, as noted by Michael Klein (1978), such “implicit” rates are measured by economists after the fact and cannot function as market signals, because depositors do not have detailed knowledge of bank costs to compute implicit rates of interest. Indeed, bank
customers have no reason to care about bank costs per se but instead base their decisions on the characteristics of the products banks offer.

Thus, a more productive approach to contemplating how banks engage in non-price competition is to focus on how banks seek to alter consumers’ perceptions of the characteristics, or qualities, of their products. Heggestad and Mingo (1976), for instance, proposed that that quality of service—such as weekly office hours of walk-in or drive-in services and availability of 24-hour automated services derived from survey data—influences behavior of bank customers. Heggestad and Mingo offered evidence derived from banking data suggesting a positive relationship between bank concentration and service quality levels, implying that in banking markets more apparent scope for competition, banks tended to offer higher-quality services. White (1976) found a similar result using the number of bank branches as the relevant quality variable, and Carlson and Mitchener (2006) suggest that bank branches likewise were important elements of non-price competition in the 1920s and 1930s.

There is considerable evidence that branching and automated-teller-machine networks continue to play important roles in the competitive interplay among banks. Calem and Nakamura (1998) have studied U.S. deposit rate data from 1985 and from 1989-1990 and concluded that although branch banking helps banks differentiate their product, it ultimately brings banks into more direct competition and thereby reduces localized market power. Schmid (1994) has found evidence in 1980s data from four European nations that banks branch to an extent dictated by consumer preferences. In a study of bank-level data from Norway over the period 1988-1995, Kim and Vale (2001) have concluded that the placement of branches is a crucial strategic variable for banks. Their work suggests that
the relative size of a bank’s branch network influences its market share but that such networks have meager feedback effects onto competitors and hence do not have an impact on overall market size. 

Cerasi et al. (2002) likewise conclude that branching was a key strategic variable in European banking markets during the 1990s. Hirtle (2007) examines the competitive role of branch network size and concludes that banks with mid-sized branch networks may be at a competitive disadvantage with respect to rivals possessing larger networks. Dick (2006) explores the impacts of U.S. banking deregulation during the 1990s on branch banking. She finds that banks responded by competing through a significant expansion in the number of branches, which together with other adjustments in the deregulated environment boosted operating costs considerably. Increased branching generated greater revenues, however. On net, therefore, she concludes that bank profits were unaffected and, additionally, there was virtually no adjustment in banking market concentration—a result consistent with Sutton’s proposed endogenous-sunk-cost mechanism.

Sutton initially applied his theory to industries in which advertising is a commonly utilized approach to trying to boost consumers’ willingness to pay. Could outlays on advertising also represent a form of endogenous sunk costs that make his theory applicable to the banking industry? Unfortunately, there are a limited number of studies on the economic impacts of advertising in the banking industry. Martín-Oliver and Sals-Fumás (2008), for instance, examine whether advertising outlays by Spanish banks between 1983 and 2003 succeeded in boosting deposit supply and loan demand. They conclude that the answer is yes in both cases, although the magnitudes of responses are relatively small, with a deposit supply
response with respect to advertising of 0.22 and a loan demand elasticity of only 0.11. The other studies—Lapp (1976), Edwards (1973, 1976), Rhoades (1980), De Pinho (2000), DeYoung and Örs (2004), Hasan et al. (2002), Kohers and Simpson (1981), Örs (2006), Scott (1978) and Wolken and Derrick (1986)—have focused mainly on the relationship between bank market concentration and advertising outlays. These studies reach contradictory conclusions. For instance, Rhoades and Kohers and Simpson find evidence of negative effects of concentration on advertising, but Hasan et al. and De Pinho conclude that there is a positive relationship. Edwards found no evidence of a relationship. Some studies, such as Scott’s, find some evidence favoring an inverted-U-shaped relationship between concentration and advertising, implying greater advertising by banks in oligopolistic markets, while others, such as Wolken and Derrick, do not.

Several of these bank advertising studies suffer from small-sample problems or reliance on survey data and consider banking data from much earlier periods. The Hasan et al. and De Pinho analyses apply to more recent data, and both conclude that there is a positive relationship between concentration and advertising expenditures at depository financial institutions, as do the more recent and broader studies by DeYoung and Örs (2004) and Örs (2006). DeYoung and Örs examine data from almost 1,900 U.S. thrift institutions in more than 600 deposit markets from 1994 through 2000 and find evidence of a positive association between market concentration and advertising outlays (except for mutual institutions). Örs evaluates data from nearly 4,500 commercial banks between 2001 and 2004 and additionally seeks to control for potential endogeneities between concentration and advertising expenses. He finds some evidence favoring an inverted-U-shaped relationship but a positive effect of concentration on
advertising outlays within the sample. In addition, Örs concludes that advertising outlays per dollar of deposit decline as a bank’s scale increases and that advertising has a positive effect on a bank’s profitability.

Overall, therefore, the evidence from more recent data covering a broader range of banking institutions indicates that advertising is an important facet of bank competition. Along with outlays for expansions of branches and ATM networks, therefore, bank advertising outlays arguably constitute endogenous sunk costs that arguably might make Sutton’s theory applicable to the banking industry. Indeed, Gual (1999) argues that bank market structure ultimately depends on whether the main determinants of competition in banking are variable and exogenous sunk costs versus endogenous sunk costs.

A recent study by Dick (2007) has examined the relevance of Sutton’s theory for the banking industry. Dick considers U.S. banking data encompassing more than 300 regional banking markets, with populations—her primary measure of market size—ranging from less than 100,000 people to more than 2 million. She finds little variation in bank market concentration across different markets sizes, as depicted in Figure 3. As shown in panels (a) of the figure, the one-bank concentration (C1) ratio, measured as the fraction of deposits held by the largest bank in the regional market, varies relatively little around Dick’s sample average of 0.30. Panel (b) shows that the same pattern holds true for the Herfindahl-Hirschman Index (HHI) measure of bank market concentration. The average HHI exhibits little variation around the sample mean value of about 1,850. Dick shows that scatterplots relating either the C1 or HHI to market size appear to imply a lower bound on market concentration. Estimated values of lower bounds on concentration depend on the presumed underlying distribution.
of markets, but Dick shows that under alternative distributional assumptions lower bounds are in fact implied by U.S. banking data.

**Figure 3:**  The Relationship between U.S. Regional Bank Market Concentration and Market Size

Panel (a)

Panel (b)

[Source: Table 2 in Dick (2007)]

Dick considers various measures of bank quality, including advertising intensity (outlays as a fraction of assets), branch density
(branches per square mile in a regional market), and alternative measures such as employees per branch, salary per employee, and number of states in which a bank operates. She finds that, consistent with the Sutton model, each of these potential measures of quality increases with market size. She also concludes that within a given market, larger banks typically provide higher levels of quality than smaller banks do, a result that turns out to be consistent with an implication of Sutton’s theory when extended to heterogeneous banks offering highly differentiated products.

Thus, Dick provides evidence that endogenous sunk costs could well be an important characteristic of competition in the banking industry. If this conclusion is justified, then banking is an example of a “natural oligopoly,” and antitrust analysis may be overly focused on industry concentration. In particular, the current critical post-merger HHI value of 1,800 and change in HHI of 200 specified by U.S. bank merger guidelines conceivably could be inconsistent with the banking industry’s “natural” structure. Furthermore, in principle the average level of quality of banks’ products could be enhanced by mergers, implying that more concentrated banking markets conceivably could yield higher welfare for consumers.

5. Evaluating the Policy Implications of Applying the Theory of Endogenous Sunk Fixed Costs to Banking

Based on the results of her study, Dick (2007) suggests that an important implication for U.S. antitrust policy with respect to banking is that bank quality is a key variable that should be incorporated into policy evaluations of the consumer welfare impacts of changes in structure within banking markets. According to Dick, “if consumers are receiving higher quality and are benefiting as a result, they are not necessarily hurt
by...higher prices they have to pay” for the higher quality. By implication, if a bank merger that results in a more concentrated market also boosts the fixed costs associated with the provision of better products made possible, both prices and consumer welfare could rise. In principle, therefore, violations of the critical HHI thresholds specified by U.S. bank merger guidelines actually could yield higher-quality banking products and improvements in consumer welfare.

Much depends, Dick suggests, on whether a proposed merger will maintain rivalry among a few dominant banks and potentially several fringe competitors. Her analysis indicates that as long as scope for active quality competition remains in place, consumer welfare will not necessarily be harmed by a merger that creates a more concentrated banking market.

Dick’s analysis is carefully done and very persuasive, which suggests that it be given serious consideration by U.S. banking antitrust authorities. Certainly, her work indicates that incorporating analytical tools for measuring consumer welfare effects of potential quality changes brought about by proposed mergers would be worthwhile. Dick’s empirical techniques for evaluating such welfare effects, or perhaps future refinements on those methods, could prove very useful, for instance, evaluating proposed mergers—particularly mergers that are on the outer edges of the merger guidelines’ critical values for the post-merger HHI and change in the HHI.

Nevertheless, there are good reasons to exercise caution in assessing applications of Sutton’s theory for the banking industry. Sutton’s own work (1991, 2000) has applied his theory to advertising- and technology-intensive industries in which endogenous sunk costs are important. While many such industries face a number of social regulations, such as product quality
regulations, most arguably face nothing like the broad range and depth of economic regulations confronted by the banking industry. Indeed, costs imposed on banks by regulations constitute a substantial component of the fixed outlays that banks incur. In a review of regulatory compliance costs imposed on banks, Elliehausen (1998) concludes that the total cost of complying with U.S. bank regulations as of the early 1990s amounted to about 12 to 13 percent of banks’ noninterest expenses. Although Elliehausen suggests that labor costs are the major component of startup and ongoing costs of complying with banking regulations, the fact that banks must comply suggests that an important consequence of regulation is to transform a significant portion of labor costs that otherwise would be variable into fixed costs. In fact, Elliehausen argues about half of the required activities of bank employees relating to regulatory compliance are undertaken only because they are required, suggesting that at least half of labor expenses related to regulatory compliance are purely regulatory fixed costs.

The fixed-cost burden of bank regulation undoubtedly has increased as a consequence of implementation of the Basel I capital standards beginning in the early 1990s and the gradual phase-in of the more highly capital-intensive Basel II regulations since the mid-2000s. Based on estimates of the total discounted present value of all implementation costs, VanHoose (2007) concludes that the total Basel II compliance cost faced by the banking industry may be equivalent to a charge against a single year’s earnings of as much as 3.5 percent, with a significant component of this cost likely being fixed costs associated with utilization of information technologies required to implement Basel II requirements.

It is arguable that fixed costs of regulation should be regarded as
exogenous sunk costs. After all, even though there are several regulators—the Federal Reserve, the Office of the Comptroller of the Currency, the Federal Deposit Insurance Corporation, the Office of Thrift Supervision, the National Credit Union Administration, and relevant state authorities—every federally insured U.S. banking institution confronts the same regulatory superstructure and faces essentially similar compliance-cost-creating layers of supervision and regulation. Surely a portion of these regulatory compliance costs are essentially exogenously fixed across institutions.

Nevertheless, a significant component of regulatory costs surely is endogenous. For instance, VanHoose (2007) documents that estimated average costs of complying with Basel II differ considerably across larger and smaller banks. In part, these cost differences reflect fluidity in the implementation of the Basel II across banks in different size categories, with larger banks confronting more rapid and costly phase-in requirements. Yet larger banks are able to spread the compliance costs across greater asset scales, which actually yields lower average compliance costs for large banks than for smaller banks. The recent subprime mortgage crisis appears to have generated a slow-down in efforts by regulators to implement Basel II beyond the largest banking institutions officially judged to be “too big to fail.” This outcome suggests an endogenous response of regulatory fixed costs aimed at reducing the impact on smaller institutions deemed by regulatory policy “sufficiently small to fail.”

Indeed, endogeneity of regulatory fixed costs is a prediction that naturally flows from key economic theories of regulation, such as theories advanced by Peltzman (1976) and Becker (1983). These theories suggest that the interaction of lobbying activities by regulated firms and by consumer interests produces a “regulatory equilibrium.” Over time, changes in
external factors—such as a subprime mortgage crisis—alter the terms of the trade-off between the interests of regulated firms and consumers, resulting in a shift in the nature of the regulatory equilibrium—and hence the magnitudes of fixed costs faced by banks in complying with the shifting terms of regulation that they confront. Banks have considerable input into the process by which regulations are adopted and adapted in response to external events. Hence, a significant portion of the fixed regulatory compliance costs they face arguably are endogenous.

Direct application of Sutton’s theory to the banking industry presumes that privately incurred fixed bank expenses—such as those associated with advertising or, as in Dick’s (2007) work, branch and ATM networks and other outlays aimed at enhancing product quality—are the primary components of banks’ endogenous sunk fixed costs. Available evidence suggests, however, that the majority of regulatory compliance costs are also sunk fixed costs that comprise a considerable portion of banks’ noninterest expenses. Work to date on applying Sutton’s framework to banking ignores this important source of endogenous sunk fixed costs in the banking industry. Of course, integrating such costs into a Sutton-style theory would require accounting for endogeneity of the process by which a regulatory equilibrium is attained and maintained over time, which likely would be a challenging undertaking.

Regulation has altered, currently impinges on, and will continue to affect significantly the history, present status, and likely future configuration of the banking industry. A key component of banks’ sunk fixed costs is compliance costs generated by regulation, which by their nature are both endogenous and subject to influence by banks themselves. Consequently, to the extent that endogenous sunk fixed costs help to explain
the present structure of banking markets, one fundamental reason may well be that an endogenous regulatory framework, rather than a natural, unregulated process may be at least partly—or perhaps even largely—responsible. Until the likely role of endogenous sunk fixed regulatory costs in determining the equilibrium structure of the banking industry has resolved, policymakers should not accept at face value the notion that banking is a “natural oligopoly.”

6. Conclusion

Sutton’s theory of the role of endogenous sunk fixed costs in determining the long-run industry structure is compelling for industries in which advertising and research and development expenses are important factors. Sutton’s theory presumes, however, that industries to which it is aimed to apply are unregulated. Thus, it is most clearly applicable to industries confronting minimal forms supervision, such as basic social regulations faced by all firms in an economy.

In light of considerable evidence that the banking industry is characterized by considerable non-price forms of competition on the basis of product quality, the suggestion that banks experience endogenous sunk fixed costs is persuasive. Furthermore, evidence amassed by Dick (2007) provides a compelling circumstantial case favoring applicability of Sutton’s theory to the banking industry. As predicted by Sutton’s theory, bank market structure does indeed seem to be invariant across markets of differing sizes. Furthermore, quality enhancements that benefit consumers appear to be prevalent in concentrated markets dominated by large banks, also consistent with the theory. An inescapable implication of these results appears to be that present U.S. antitrust policy regarding banking mergers
may have been formulated on a false premise. Focusing on a traditionally perceived trade-off between potential augmentations of market power of mergers versus possible merger efficiency enhancements could fail to reflect product-quality improvements and related consumer welfare enhancements that mergers may generate.

Nevertheless, an important element missing from application of the theory of endogenous sunk fixed costs to the banking industry is the important role of regulatory compliance costs. The fixed costs that banks confront in complying with regulations are significant. In addition, the regulatory framework confronted by banks is determined through a dynamic interplay between banks, consumers, and government policymaking. Hence, a substantial portion of fixed regulatory compliance costs are also endogenously determined. To date, the theoretical and empirical analysis of the impact of endogenous sunk fixed costs on the structure and performance of the banking industry has failed to account for regulatory fixed costs. Until it does, the policy implications of endogenous sunk fixed costs for the banking industry will remain unsettled.
REFERENCES


Barro, Robert, and Anthony Santomero, 1972, Household money holdings and the demand deposit rate, Journal of Money, Credit, and Banking 4, 397-413.


Dick, Astrid, 2007, Market size, service quality, and competition in banking,
Journal of Money, Credit, and Banking 39, 49-81.


Heggestad, Arnold, and John Mingo, 1976, Prices, nonprices, and concentration in commercial banking, Journal of Money, Credit, and Banking 8, 107-


Klein, Benjamin, 1974, Competitive interest payments on bank deposits and the long-run demand for money, American Economic Review 64, 931-949.


Merris, Randall, 1985, Explicit interest and implicit demand deposit interest, *Journal of Money, Credit, and Banking* 17, 528-533.


White, Lawrence, 1976, Price regulation and quality rivalry in a profit-maximizing model: The case of bank branching, Journal of Money, Credit, and Banking 8, 97-106.