2 Competitive analysis in banking: Appraisal of the methodologies

16 Bank Structure Conference announcement

18 Agglomeration in the U.S. auto supplier industry

35 The new view of growth and business cycles
## Contents

First Quarter 1999, Volume XXIII, Issue 1

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><strong>Competitive analysis in banking: Appraisal of the methodologies</strong>&lt;br&gt;&lt;br&gt;Nicola Cetorelli</td>
</tr>
<tr>
<td></td>
<td>How do we measure market power in the banking industry? This article provides an</td>
</tr>
<tr>
<td></td>
<td>overview of the methodology currently used in competitive analysis and highlights an</td>
</tr>
<tr>
<td></td>
<td>alternative technique that could be used to complement this methodology. Given the</td>
</tr>
<tr>
<td></td>
<td>ongoing process of consolidation in U.S. banking, assessing the competitiveness of</td>
</tr>
<tr>
<td></td>
<td>financial services markets is an important issue for policymakers.</td>
</tr>
<tr>
<td>16</td>
<td><strong>Bank Structure Conference announcement</strong></td>
</tr>
<tr>
<td>18</td>
<td><strong>Agglomeration in the U.S. auto supplier industry</strong>&lt;br&gt;&lt;br&gt;Thomas H. Klier</td>
</tr>
<tr>
<td></td>
<td>Analysis of a large set of plant-level data shows the auto supplier industry to be highly</td>
</tr>
<tr>
<td></td>
<td>spatially concentrated. Data on location of the plants’ customers, however, suggest that</td>
</tr>
<tr>
<td></td>
<td>immediate proximity to assembly plants is not necessary, despite a production system</td>
</tr>
<tr>
<td></td>
<td>that emphasizes low inventories and tight linkages.</td>
</tr>
<tr>
<td>35</td>
<td><strong>The new view of growth and business cycles</strong>&lt;br&gt;&lt;br&gt;Jonas D. M. Fisher</td>
</tr>
<tr>
<td></td>
<td>Evidence on the cost of business equipment investment supports a new way of understand-</td>
</tr>
<tr>
<td></td>
<td>ing growth and business cycles. The equipment price has been falling for most of the</td>
</tr>
<tr>
<td></td>
<td>last 40 years and it tends to fall more the faster the economy is growing. This suggests</td>
</tr>
<tr>
<td></td>
<td>that technological change embodied in new capital equipment has a substantial effect on</td>
</tr>
<tr>
<td></td>
<td>growth and business cycles.</td>
</tr>
</tbody>
</table>
Competitive analysis in banking: Appraisal of the methodologies

Nicola Cetorelli

Introduction and summary

Over the last 20 years, the U.S. banking industry has experienced significant structural changes as the result of an intense process of consolidation. From 1975 to 1997, the number of commercial banks decreased by about 35 percent, from 14,318 to 9,215. Since the early 1980s, there have been an average of more than 400 mergers per year (see Avery et al., 1997, and Simons and Stavins, 1998). The relaxation of intrastate branching restrictions, effective to differing degrees in all states by 1992, and the passage in 1994 of the Riegle-Neal Interstate Banking and Branching Efficiency Act, which allows bank holding companies to acquire banks in any state and, since June 1, 1997, to open interstate branches, is certainly accelerating the process of consolidation.

These significant changes raise important policy concerns. On the one hand, one could argue that banks are merging to fully exploit potential economies of scale and/or scope. The possible improvements in efficiency may translate into welfare gains for the economy, to the extent that customers pay lower prices for banks’ services or are able to obtain higher quality services or services that were not previously offered. On the other hand, from the point of view of public policy it is equally important to focus on the effect of this restructuring process on the competitive conditions of the banking industry. Do banks gain market power from merging? If so, they will be able to charge higher than competitive prices for their products, thus inflicting welfare costs that could more than offset any presumed benefit associated with mergers.

In this article, I analyze competition in the banking industry, highlighting a very fundamental issue: How do we measure market power? Do regulators rely on accurate and effective procedures to evaluate the competitive effects of a merger?

The U.S. Department of Justice, the Federal Reserve System, the Federal Deposit Insurance Corporation (FDIC), and the Office of the Comptroller of the Currency (OCC) enforce the antitrust laws in banking. The procedures to evaluate the competitive impact of a proposed merger may differ in some details among the agencies, but they all share the same approach, based on structural analysis of the banking market affected by the merger. The basic guideline, established by the Justice Department, requires the evaluation of the concentration of deposit market shares held by banks operating in the affected market. The importance of market concentration finds its theoretical justification in the so-called structure-conduct-performance paradigm (Bain, 1951), which postulates that fewer and larger firms (higher concentration) are more likely to engage in anticompetitive conduct. For example, a small number of large firms may be able to cooperate and act as a monopoly (cartel). Alternatively, one or more firms together may be large enough to set higher than competitive prices (acting as a dominant firm), while the other (smaller) firms would act as a competitive fringe, following the dominant firm’s behavior.

The most common measure of concentration, and the one used by regulators, is the Herfindahl-Hirschman Index (HHI), which is defined as the sum of the squared market shares of all banks in the market (box 1 explains how the index is calculated). According to the current screening guidelines, if the post-merger market HHI is lower than 1,800 points, and the increase in the index from the pre-merger situation is.

Nicola Cetorelli is an economist at the Federal Reserve Bank of Chicago. The author thanks Eli Breuer, Betsy Dale, Bob DeYoung, Doug Evanoff, Hesna Genay, David Marshall, and Paula Worthington for their comments.
less than 200 points, the merger is presumed to have no anticompetitive effects and is approved by the regulators. Should those threshold values be exceeded, the regulators will check for the existence of potential mitigating factors that would make it unlikely that the merger could result in anticompetitive behavior. The regulators also seek to identify those extreme cases in which the potential welfare loss from the exercise of market power would be smaller than the loss produced by maintaining the status quo (for example, the merger might prevent the failure of one of the parties involved, thus preserving the stability of the market). If the mitigating factors are not enough to justify the merger, the regulators may require the divestiture of some branches and offices, in order to bring the concentration indicator closer to or below the threshold level. If divestiture would not accomplish this goal, the merger application is denied. If the merger does not violate the 1,800/200 rule, the application is approved without further investigation.

Over the years, very few mergers have been denied. However, this fact should not lead one to conclude that the rules are not sufficiently stringent. The official statistics do not show attempts to file merger applications that were abandoned because of a voluntary decision of the banks involved or informal dissuasion by the regulators.

Does the ongoing merger and consolidation process represent a real competitive threat? A survey of local markets shows that concentration is a widespread characteristic of the banking industry. For example, in 1994, about 40 percent of metropolitan statistical areas (MSAs) had HHIs greater than 1,800 (Rhoades, 1995b). If indeed high concentration implies noncompetitive conduct, then policy concerns about the welfare effects of future mergers may be justified.

First, I review the appropriateness of the use of the HHI as a main screening factor in merger analysis. I examine the theoretical foundations of the market concentration–market power relationship and how focusing on market structure to infer firms’ conduct may lead to ambiguous or even misleading conclusions about the potential effects of a merger.

Next, I survey the state of the art of the empirical literature. If there are consistent and convincing empirical results confirming the existence of the market concentration–market power relationship, then it may be appropriate to use it in policy analysis, even in the absence of a solid theoretical explanation. While

---

**BOX 1**

**Calculation of the Herfindahl-Hirschman Index (HHI)**

The HHI formula is

$$HHI = \sum_{i=1}^{n} MS_i^2,$$

where $MS_i$ is the market share of bank $i$ and $n$ is the number of banks in the market.

Suppose a market has five banks. The share of total deposits of each bank is as follows:

<table>
<thead>
<tr>
<th>Bank</th>
<th>Deposit market share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

The $HHI = 30^2 + 25^2 + 21^2 + 16^2 + 8^2 = 2,286$. Suppose that banks 3 and 5 merge. After the merger, the $HHI = 30^2 + 29^2 + 25^2 + 16^2 = 2,622$, with a post-merger increase $\Delta HH = 336$. In antitrust evaluation this merger may be rejected, because it violates the 1,800/200 rule.

By construction, the HHI has an upper value of 10,000, in the case of a monopolist firm with 100 percent share of the market, and tends to zero in the case of a large number of firms with very small market shares.

The HHI synthesizes information on both the distribution of market shares and the number of banks in the market. With some manipulation it could be rewritten as

$$HHI = \frac{\sum_{i=1}^{n} MS_i^2}{n},$$

where $V$ is the coefficient of variation of deposit market shares, and $n$ is the number of firms in the market. This feature of the HHI makes it more popular than other concentration indicators, such as the $n$-firm ratio, calculated as the sum of the market shares of the $n$ largest firms in the market, where $n$ is usually 3 or 4.
there have been important contributions confirming a positive and significant relationship between market concentration and the exercise of market power, other recent work has cast doubt on the overall empirical strength of such a relationship.

I then describe an alternative methodology of competitive analysis that does not infer banks’ conduct through the analysis of market structure. This methodology recognizes that firms’ behavior differs depending on whether they operate in a perfectly competitive market, a monopolistic market, or any other prevalent market structure. I survey the applications of this methodology, which is based on the estimation of a direct indicator of firms’ behavior, for the banking industry.

Finally, I present some results of a specific empirical application of this methodology to the Italian banking industry. The analysis of Italy is relevant because the Italian banking industry has experienced a similar pattern of structural and regulatory changes as U.S. banking. In particular, as the result of an ongoing process of consolidation, the Italian HHI has been steadily increasing. The results of my empirical analysis indicate a steady convergence toward competitive conditions, providing evidence that changes in market concentration may not always provide correct information about the exercise of market power.

Theory behind the Herfindahl–Hirschman Index

As discussed above, the use of concentration ratios to evaluate competitive conditions relies on the theoretical predictions of the structure–conduct–performance paradigm. According to this paradigm, structure affects the conduct of firms, which ultimately determines their performance. Concentration of market shares will facilitate the adoption of collusive conduct and, ultimately, the setting of prices departing from the perfectly competitive benchmark. In a perfectly competitive market, firms are considered too small to have an individual impact on the price of the good they produce. From the point of view of social welfare, perfect competition represents an ideal benchmark, since consumers (in this case bank customers) pay the lowest possible price for the product they demand. Any situation in which firms command some degree of market power and are therefore able to set higher than competitive prices implies a social cost in terms of welfare loss for consumers.

The structure–conduct–performance paradigm predicts that there is an increasing relationship between the level of market concentration and market power. Some authors are more precise in stating that the relationship, while it is increasing, may not be linear. One would expect that at low levels of concentration, conduct is close to competitive, and an increase in concentration would generate a substantial increase in market power. At high levels of concentration, conduct is already very far from the competitive benchmark, and an additional increase would not increase market power very much. Given this argument, the market concentration–market power relationship should be S-shaped, as shown in figure 1 (Carlton and Perloff, 1989).

Is it possible to derive an optimal behavior rule from a model of industrial organization theory that predicts an increasing relationship between market concentration and market power? Can we rely on such a model to find a theoretical justification for, say, the 1,800/200 rule? The answer is yes, but only if one makes strong, restrictive assumptions about firms’ behavior, such as assuming that firms behave as Cournot oligopolists. Under Cournot conduct, a firm makes the simplistic assumption that all other firms have no reaction to a change in its behavior (see the technical appendix for the analytical derivation of this result). However, in more general (and plausible) theoretical models that allow for active interactions among firms, the market concentration–market power relationship is less obvious.

Thus, it seems that we cannot rely too much on theory to justify the postulated market concentration–market power relationship. Before surveying the approach taken in the profession, which has been to turn to a direct empirical corroboration of the postulated relationship, I present some simple numerical examples showing that, in the absence of a complete theory that can explain the market concentration–market

![Figure 1](attachment://image.png)

**Figure 1**

Theoretical relationship between market concentration and market power

- **Market power**
- **Monopoly**
- **Competition**

Note: HHI represents the Herfindahl–Hirschman Index.

Economic Perspectives
power relationship, it is possible to generate ambiguous or even incorrect predictions about the effects of a structural change on competition.

**Numerical examples**

These examples demonstrate the following two assertions: First, even when the 1,800/200 rule is not violated, a merger may generate anticompetitive conduct. Second, a merger may be procompetitive even when the 1,800/200 rule is violated.

In the first two examples, the basic guidelines are not violated. However, the mergers may generate the right conditions for monopoly power, not necessarily exercised only by the banks involved in the merger.

Table 1 summarizes the examples.

In a pre-merger market with 20 banks, each with a 5 percent market share (see Table 1, example 1), the HHI \((5^2 + 5^2 + \ldots + 5^2 = 500)\) characterizes a market with a relatively large number of banks with equal and small market shares and is presumably associated with a low likelihood of anticompetitive behavior. Suppose five of the banks are involved in a series of mergers. When all the mergers are completed, the market has one bank with a 25 percent market share and 15 banks with 5 percent each. The post-merger HHI of 1,000 would still be considered (borderline) un­con­cen­trated.\(^6\) However, the newly created bank, with a 25 percent market share, may be able to act as a **dominant firm**, setting noncompetitive prices, with the remaining 15 banks behaving as a *competitive fringe*, adjusting to the noncompetitive choices of the dominant firm.

In the second example, the pre-merger market has 15 banks, two with 15 percent market shares, one with 10 percent, and 12 with 5 percent (see Table 1, example 2). The two larger banks, \(B_1\) and \(B_2\), taken separately, may still be too small to behave as dominant firms. In addition, tacit or explicit collusion between them to act together as a dominant firm may still be unlikely, given the fact that the combined market share may not generate the market power and extra profits necessary to offset the costs associated with collusion.\(^7\) The HHI of 800 may therefore be correct in characterizing a competitive market.

Suppose banks \(B_3\) and \(B_4\) merge. The post-merger structure now has three banks with a 15 percent market share each and 11 banks with 5 percent each. The post-merger HHI is now 950. As in the first example, according to the guidelines the market would still be considered unconcentrated. However, the three major banks may now be able to coordinate (explicitly or tacitly) their action, thus producing adverse competitive conditions. (Note also that the two larger banks in the pre-merger market are benefiting from a merger that did not directly involve them).

The third example describes a market in which some degree of collusive behavior might have been observed prior to the merger (see Table 1, example 3). The merger could create conditions under which the stable collusive agreement would break down, thus restoring market competition. However, since the basic guidelines are violated, the merger could be rejected and the exercise of market power preserved.

The pre-merger market has seven banks, three with 20 percent market shares, two with 15 percent shares, and two with 5 percent shares. The HHI of 1,700, classifying the market as moderately concentrated, may not fully account for a situation in which the three largest banks, \(B_1\), \(B_2\), and \(B_3\), may be able to collude. In the event of a merger between banks \(B_4\) and \(B_5\), the post-merger market would have six banks, one with a 30 percent market share, three with 20 percent each, and two with 5 percent each. The post-merger HHI of 2,150 identifies this as a highly concentrated market. In addition, since the change in the HHI would be more than 200 points, there are grounds for the regulator to reject the merger application. However, the stability of a

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examples of pre- and post-merger markets</strong></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>Example 1</strong> Pre-merger market (20 banks)</td>
</tr>
<tr>
<td>Bank</td>
</tr>
<tr>
<td>Market share (%)</td>
</tr>
<tr>
<td>Post-merger market (16 banks)</td>
</tr>
<tr>
<td>Bank</td>
</tr>
<tr>
<td>Market share (%)</td>
</tr>
<tr>
<td><strong>Example 2</strong> Pre-merger market (15 banks)</td>
</tr>
<tr>
<td>Bank</td>
</tr>
<tr>
<td>Market share (%)</td>
</tr>
<tr>
<td>Post-merger market (14 banks)</td>
</tr>
<tr>
<td>Bank</td>
</tr>
<tr>
<td>Market share (%)</td>
</tr>
<tr>
<td><strong>Example 3</strong> Pre-merger market (7 banks)</td>
</tr>
<tr>
<td>Bank</td>
</tr>
<tr>
<td>Market share (%)</td>
</tr>
<tr>
<td>Post-merger market (6 banks)</td>
</tr>
<tr>
<td>Bank</td>
</tr>
<tr>
<td>Market share (%)</td>
</tr>
</tbody>
</table>
collusive agreement is known to decrease with the number of participants. In the new market structure, with four large players, the collusion might break down. In that case, the merger would actually be procompetitive.

In considering whether to reject the merger application, the regulator may impose some degree of divestiture on the banks involved in the mergers. Ironically, banks B₁, B₂, and B₃, which were not involved in the merger, could benefit in this case, as the post-divestiture B₄ may not be strong enough to undermine the stability of their pre-merger collusive agreement.

The market dynamics described in these numerical examples are all hypothetical. My point is that whether a merger will generate (undetected) anticompetitive conditions or actually improve competition cannot be determined unambiguously just by looking at market structure. Banks’ behavior can only be measured accurately through direct empirical analysis.

**Empirical evidence**

The empirical evidence for the existence of the market concentration–market power relationship is mixed. Some influential papers have suggested a positive relationship between concentration and the degree of market power. For example, Berger and Hannan (1989) analyze a cross-section of banking markets in 1983–85. After controlling for various factors affecting price-setting behavior, the authors find that deposit rates are significantly lower in the most concentrated markets.

Other work compares the time-series behavior of the deposit interest rate (and/or the loan rate) with the benchmark money market rate, which is not controlled by the banks. If banks have market power, they will, for example, quickly lower the deposit rate when the money market rate decreases, but the deposit rate will be sluggish when the money market rate increases. Conversely, in perfect competition one should expect quick reactivity in both cases. Hannan and Berger (1991) and Neumark and Sharpe (1992) find evidence of deposit rate rigidity and, thus, evidence of market power in the U.S. banking industry. Importantly, they find a higher level of rigidity in markets with higher HHIs.

However, recent research casts doubt on the market concentration–market power relationship. Reviewing Berger and Hannan’s (1989) results, Jackson (1992b) suggests that the market concentration–market power relationship may not be monotonic. He finds that such a relationship already holds at low levels of concentration, but in markets with middle levels of concentration the relationship vanishes, and it actually changes sign in highly concentrated markets (although this is a less robust result). In other words, at higher levels of concentration, an increase in concentration may imply less anticompetitive behavior, as suggested in example 3 of table 1.

In another work focusing on the rigidity of deposit rates, Jackson (1997) presents additional evidence that the market concentration–market power relationship may not be monotonic. He finds that while it is true that at high levels of concentration price rigidity increases, this is also the case at low levels of concentration. This suggests a U-shaped relationship between market power and market concentration which is not consistent with the structure–conduct–performance hypothesis.

Similarly, Rhoades (1995a) observes that structural characteristics may vary widely for markets exhibiting similar HHI levels. In particular, the market share distribution may differ substantially. As shown in example 1 above, firms’ conduct may be very different depending on market share distribution. Rhoades shows that market share inequality and the number of firms in the market have an effect on banks’ profitability that is independent of the HHI, despite the fact that (as shown in box 1) the HHI incorporates information on both market share variability and the number of firms. Finally, in an analysis similar to Berger and Hannan’s (1989), Hannan (1997) extends Rhoades’s (1995a) contribution by analyzing the impact of these two factors on deposit rate levels. His results for a cross-section of banking markets using November 1993 data show, first, that the HHI was not significant in explaining deposit rates and, second, that it was not able to take into account the separate importance of market share inequality and the number of firms.

Thus, a lack of strong theoretical foundations and mixed empirical evidence motivate the search for alternative methodologies to investigate firms’ competitive behavior.

**Oligopoly theory and the measurement of market power**

Methodologies in the “new empirical industrial organization” literature analyze firms’ conduct directly, instead of relying on observation of the market structure. Following this approach, the relationship between theory and firms’ conduct becomes unambiguous. For instance, as mentioned earlier, if banks are behaving as Cournot oligopolists, the market concentration–market power relationship would be theoretically grounded and the use of the HHI to infer firms’ conduct would be appropriate. This alternative methodology allows us to test whether indeed banks behave as Cournot oligopolists. However, the methodology is flexible enough to allow us to test for behavior that could be consistent with alternative models of
oligopoly theory. In such a case the market concentration—market power relationship would not be as clearly identified as in the Cournot case, but one would still be able to quantify the departure from perfect competition and, hence, to assess the degree of market power exercised in the industry.

The technical appendix provides details of the methodology. The following example illustrates the intuition. Suppose there is an exogenous increase in the demand for bank loans. In response, banks will take into account the cost they would incur in increasing the quantity of loans, the reactivity of demand itself to possible increases in the loan rate, and the expected reaction of the other banks in the market to their chosen course of action. In particular, the degree of interaction with the other banks in the market could differ substantially, depending on whether banks are in perfect competition with each other or enjoy some degree of market power. More precisely, the parameter of banks’ interaction should be equal to 0 if the market is perfectly competitive, equal to 1 if it is monopolistic, and should take intermediate values between 0 and 1 if banks are neither perfectly competitive or monopolistic but still exercise a positive degree of market power. Using appropriate econometric modeling techniques, one can estimate this parameter of interaction and, therefore, a quantifiable measure of market power.

The advantage of this approach is that it is rigorously based on theory and does not require indirect (and perhaps ambiguous) inferences about market power through measures of market concentration. The major limitation of the approach is that it requires detailed information, mainly on cost and demand conditions at the firm level.

Applications to the banking industry

Spiller and Favaro (1984) estimate the parameter of banks’ interaction for the Uruguayan banking industry in a period characterized by a significant relaxation of entry regulations. They apply a refinement of the methodology proposed by Gollop and Roberts (1979) to see whether different groups of banks have different reactions to other groups’ change in behavior. They reject Cournot conduct and find evidence of dominant firm–competitive fringe behavior, with a significant degree of oligopoly power, although this is substantially reduced after deregulation. Gelfand and Spiller (1987) extend the analysis of Uruguayan banks, treating the banks as multiproduct firms, the products being loans in the domestic currency and in U.S. dollars. They find evidence of noncompetitive behavior and, in particular, behavior consistent with mutual forbearance, whereby firms avoid changing behavior in one market fearing retaliation in another market, and with spoiling, whereby firms adopt predatory strategies. Applying the methodology to the Norwegian banking industry, Berg and Kim (1994) find that Cournot behavior is strongly rejected by the data and that instead banks behave as if they expect retaliation from their competitors in response to a change in their own behavior. Berg and Kim (1996) also investigate Norwegian banks as multiproduct firms, distinguishing between the retail and corporate banking markets. They find banks’ degree of oligopoly power to be relatively high in the retail market and lower in the corporate market. Interestingly, the Herfindahl indicators for the two markets analyzed suggest opposite findings. Shaffer (1989), using aggregate data for the U.S. banking industry, finds no evidence of oligopoly power. Similarly, in a study of Canadian banking, Shaffer (1993a) finds that despite structural and regulatory changes, Canadian banks operate in a market exhibiting perfect competition. Shaffer and Di Salvo (1994) focus on a local market in Pennsylvania with only two banks. They find that banks’ conduct is imperfectly competitive, but closer to perfect competition than one would expect, given the very high degree of concentration in that market.

Measuring market power: Results from an application to the Italian banking industry

Next, I present some results from an application of the methodology outlined above to the Italian banking industry. The remainder of the section is based on Angelini and Cetorelli (1998).

As mentioned in the introduction, there are at least two reasons the evolution of the Italian banking industry is of interest. First, the Italian banking industry is experiencing a similar pattern of regulatory and structural changes as that observed in the U.S. In the late 1980s, the requirement that Italian banks obtain a specific authorization from the central bank to open an additional branch was eliminated. Consequently, from 1983 to 1993 the number of branches increased by 67 percent. At the same time, mainly based on the anticipated opening of Italy’s national borders to international competition, widespread merger activity reduced the number of banks by more than 10 percent, to a total of approximately 900. It is not clear a priori whether such changes have actually enhanced competition. Second, the results for Italy highlight the possibility that changes in market concentration may provide misleading information on the exercise of market power.

To determine an average indicator of banks’ interaction, Angelini and Cetorelli (1998) analyze the market for commercial loans in 1983–93, pooling data
on all individual banking institutions, in substance treating the market for commercial loans as having a national dimension. It is usually argued that, especially for wholesale loans, the market boundaries are indeed very wide. Given that Italy is about as large as a mid-size U.S. state, using such a broad market definition seems appropriate. Also, performing the analysis at the national level increases the potential for finding evidence of perfect competition. This is true at least in terms of the structure–conduct–performance approach, since, as we will see below, market concentration is very low at the national level. With a possible bias in the study toward a finding of competition, therefore, evidence of noncompetitive behavior would be a strong result.

Angelini and Cetorelli (1998) make the following observations about the level of concentration of the Italian banking industry. First, the HHI, calculated on both deposits and loans, remained practically unchanged in the first part of the sample period, but increased noticeably after 1990, clearly due to the wave of bank mergers mentioned above. Second, in absolute terms the HHI remains very low, going from about 200 to 260 points over the entire period. Figure 2 plots the HHI time series for both deposits and loans. Following the predictions of the structure–conduct–performance paradigm, these two observations would imply that, given the extremely low level of concentration, the Italian banking industry should exhibit a very high degree of competition over the entire sample period, but with gradual movement toward conditions more appropriate to the exercise of market power.

In fact, the results of the econometric estimation contradict both predictions of the structure–conduct–performance paradigm. Figure 3 shows the estimates of the parameter of banks’ interaction for each year between 1983 and 1993, a period including years before and after the regulatory changes. As explained earlier, the parameter should take values between 0 and 1, with 0 representing the perfectly competitive benchmark and 1 the monopolistic benchmark. However, the results show the parameter is significantly different from 0 (and from 1) for almost the entire sample period, thus rejecting the hypothesis that the Italian banking industry is perfectly competitive (as well as the hypothesis that it is a perfect monopoly). This finding contradicts the inference one would draw from the HHI. Indeed, given the very low level of concentration, one might expect the market for commercial loans at the national level to be very competitive.

A further observation is that the parameter is well above 0 in the initial part of the sample, prior to deregulation, and shows an approximately steady decline throughout the rest of the sample period. This can be seen as evidence that the regulatory and structural changes have indeed enhanced the overall competitiveness of the banking industry. Finally, the parameter approaches 0, suggesting the presence of perfectly competitive conditions, toward the end of the period. This represents a second element of contradiction with the information in the HHI, which is increasing in the final years of the sample period.

In addition to the estimation of the parameter of interaction, Angelini and Cetorelli (1998) estimate a parameter measuring the elasticity of demand for commercial loans. As mentioned earlier, in deciding on behavior, banks have to take into account not only the
expected reaction of other banks but also the reaction of customers. Whether the market for loans exhibits a high or low elasticity to changes in the loan rate is crucial to banks' ability to exercise market power and affect profits. The intuition is simple. Suppose the parameter of interaction is very high, close to 1, approximating ideal conditions for the exercise of market power. Banks would attempt to keep a high loan rate, or to increase it, to maximize their profits. However, if market demand elasticity is also high, borrowers are likely to reduce substantially their demand for loans in the case of a price increase. In such a case, banks will be constrained in their ability to profit from their market power. The opposite would be true in the case of a rigid demand schedule.

This consideration is important, therefore, if we are interested in exploring the actual welfare cost of market power, in terms of how high the loan rate is relative to what it would be under perfect competition. To obtain a quantifiable measure of this, Angelini and Cetorelli (1998) compute the ratio of the parameter of banks' interaction and the parameter measuring demand elasticity. When this ratio is close to 0, it means that the market exhibits competitive conditions, regardless of banks' potential ability to exercise market power. Figure 4 reports estimates for this ratio for every year in the sample period. Between 1984 and 1986, interest rates on loans charged by banks were about 2 percentage points above the level that would have been charged under competitive conditions (interest rates on loans averaged around 21 percent). This gap declined to about 1 percentage point in 1987–89, then dropped to practically 0 at the beginning of the new decade. This provides evidence that the Italian banking industry has changed substantially as a result of the process of deregulation and consolidation that began in the late 1980s.

**Conclusion**

This article has presented an overview of the methodologies used in competitive analysis of the banking industry. Given the ongoing process of consolidation in the U.S. banking industry, properly identifying the conditions for the exercise of banks' market power is highly relevant for policy analysis.

I have briefly outlined the antitrust analysis procedure currently followed by the regulators. Drawing on the existing literature, I have highlighted some challenges to the theoretical foundations of the current approach, which is based on the identification of an increasing, monotonic relationship between market concentration and market power. Only under rather strong, restrictive assumptions about the behavior of banking firms is this relationship identifiable. As shown in the numerical examples, relying on concentration measures alone to infer industry conduct may lead to possibly incorrect conclusions. The empirical evidence on the existence of the market concentration–market power relationship is rather mixed, in light of several recent works that cast doubt on the robustness of such a relationship.

An alternative methodology for the identification of parameters of firms' conduct and the degree of market power, which does not rely on indirect inferences of market structure analysis, requires an econometric estimation of market demand and supply conditions. The testable implications associated with this approach allow us to unambiguously identify firms' conduct. The results from an empirical application of this methodology to the Italian banking industry provide evidence that contradicts the inferences of the structure–conduct–performance approach.

Adopting this alternative methodology to identify the parameter of banks' interaction brings a higher rigor to the antitrust analysis, implicit in the econometric exercise required to extract information from industry data. This is, however, also its principal shortcoming, in terms of the need for more detailed data and the greater difficulty associated with the implementation and interpretation of the econometric work. Conversely, the main advantage of the current approach to competitive analysis is that HHI indicators are relatively easy to compute and allow the regulators to formulate objective statements (for example, setting the 1,800/200 guideline) and deliver opinions that are less subject to arbitrary judgements. Nonetheless, it is important to recognize the potential...
shortcomings of the current approach and to test for accountability when developments in economic research provide the appropriate tools.

For example, the alternative methodology presented in this article could be applied to markets in which mergers have been approved to analyze banks’ conduct before and after the change in market structure. In addition to an “after the fact” analysis, the methodology could be used routinely to overview market conditions and to provide ex ante information that could be used by regulators when a merger application is filed, perhaps to resolve potential ambiguities associated with mere observation of market structure. In this way, the methodology could be adopted to complement the current procedure for antitrust analysis.

TECHNICAL APPENDIX

Details of the methodology

Estimating market power

The basic elements of the methodology can be illustrated as follows. In an industry producing a single good, let \( p \) be the market price of product \( y \) and let \( y_j \) be the quantity produced by firm \( j, j = 1, \ldots, m \), and let \( \sum_j y_j = y \). Let the demand function, written in inverse form, be \( p = p(y, z) \), where \( z \) is a vector of exogenous variables affecting demand. In addition, let \( C(y_j, \omega_j) \) be the cost function for firm \( j \), where \( \omega_j \) is the vector of the prices of the factors of production employed by firm \( j \).

Firms in this industry behave as profit maximizers. The profit maximization problem for firm \( j \) is written as

1) \[
\text{Max}_{y_j} p(y, z)y_j - C(y_j, \omega_j).
\]

If firms were in perfect competition with each other, they would set their optimal quantities at the point where the marginal cost of production would equal the market price, that is,

2) \[
p = C'(y_j, \omega_j),
\]

where \( C'(y_j, \omega_j) \) is the marginal cost of firm \( j \).

At the opposite extreme, suppose there is only one firm in the industry, operating as a monopolist. In such a case, we know that the firm would set quantities to a level where marginal revenue equals marginal cost, or

3) \[
p = C'(y, \omega) - \frac{dp}{dy} y,
\]

where \( p + \frac{dp}{dy} y \) is the monopolist marginal revenue \( \left( \frac{dp}{dy} y < 0 \right) \). In intermediate oligopolistic structures, with \( m \) firms operating in the market, conduct would be summarized by the general expression

4) \[
p = C'(y_j, \omega_j) - \frac{dp}{dy} y_j \theta_j.
\]

where the parameter \( \theta_j \) is an index of oligopoly conduct, quantifying the departure from the competitive benchmark. Equation 4 is a very general expression embedding various models of oligopoly behavior, which can be estimated econometrically. To appreciate its generality, it is perhaps convenient to interpret \( \theta_j \) as a parameter measuring the “conjectured” or “perceived” response of the entire industry to a change in quantity operated by firm \( j \). Solve the maximization problem in equation 1 in more extensive form as

5) \[
p + \frac{dp}{dy} y_j - C'(y_j, \omega_j) = 0.
\]

Multiply and divide the second term of equation 5 by \( y \). Then, rearranging terms, the equation can be rewritten as

6) \[
p = C'(y_j, \omega_j) - \frac{\theta_j}{\bar{c}},
\]

where

7) \[
\bar{c} \equiv \frac{dy}{dy} \frac{1}{y}, \bar{c} < 0
\]

is the semi-elasticity of demand and

8) \[
\theta_j = \frac{\bar{c} y_j}{\bar{c}_j y}
\]

is the so-called conjectural elasticity, that is, the percentage variation in aggregate output due to firm \( j \)’s
change in $y_j$. It should be clear that one does not need to impose any a priori restriction on $\theta_j$, that is, any behavioral model is a priori plausible, and the more appropriate one can be tested and identified econometrically. For example, if firms were Cournot oligopolists, then $\frac{\partial y}{\partial y_j} = 1$. Recall that under Cournot behavior, firm $j$ expects that all other firms will not adjust their quantities to a change in $y_j$. Therefore, since $y = \sum_j y_j$ incorporates firm $j$ quantity, the total variation in output to a change in $y_j$ must equal unity. Thus, under Cournot, $\theta_j$ would reduce to the market share of firm $j$.

If firms were instead in perfect competition, then $\frac{\partial y}{\partial y_j} = 0$, hence $\theta_j = 0$. In the case of monopoly, $\frac{\partial y}{\partial y_j} = 1$ and $y_i = y$, hence $\theta_j = 1$. Therefore, the convenient feature of this approach is that it specifies well-defined boundaries in terms of industry equilibrium conditions (perfect competition at one end and monopoly at the other), within which it is possible to identify the actual underlying characteristics of firms' conduct.

Given the generality of the methodology, one can also test whether $\theta_j \neq \theta_i$, where $h = 1, ..., l$ and $i = 1, ..., n$ and $l + n = m$. This would allow us to test, for example, whether firms behave according to dominant firm or leader-followers models.

**Analytical derivation of the market concentration–market power relationship**

We can also see now under what behavioral restrictions it is possible to identify a relationship between market concentration and market power. Define the degree of market power of firm $j$ as

\[
\alpha_j = \frac{p - C'(y_j, \omega_j)}{p} = \frac{\theta_j}{\varepsilon},
\]

where $\varepsilon \equiv \frac{\partial y}{\partial p} / p$ is the elasticity of demand.

Now define the degree of market power of the industry as a firm average, weighted by firms' relative size,

\[
\alpha = \sum_j \left[ \frac{p - C'(y_j, \omega_j)}{p} \right] \frac{y_j}{y} = \sum_j \frac{\theta_j y_j}{\varepsilon y}.
\]

Given the definition of $\theta_j$, we can rewrite this last expression as

\[
\alpha = \sum_j \frac{1}{\varepsilon} \frac{\partial y}{\partial y_j} \left( \frac{y_j}{y} \right)^2.
\]

Assume now that all firms form the same, identical conjecture about how the rest of the industry would react to a change in their own quantities. In addition, assume that these identical conjectures will also stay the same over time and over changes in market structure (for example, distribution of market shares and number of firms). Under these conditions, $\frac{\partial y}{\partial y_j} = \gamma, \forall j$, where $\gamma$ is a given constant.

Consequently

\[
\alpha = \frac{1}{\varepsilon} \gamma \text{HHI}.
\]

The Cournot model, where $\gamma = 1$, is an example of a model that would identify a proportional relationship between market concentration and market power. However, we have already remarked that the Cournot conjecture is rather restrictive. It seems even more restrictive to assume identical conjectures equal to some arbitrary constant $\gamma$. Moreover, note the importance of the assumption that the identical conjectures will have to stay unchanged over time and in case of a change in market structure. This implies assuming that $\gamma$ and HHI are independent from each other. Yet, as we argued earlier, a change in market structure, such as the one determined by a merger, whereby the distribution of market shares and the number of firms operating in the industry vary, will have an effect on how firms perceive the conduct of one another. This effect on conduct will not necessarily be the same for all firms (see, for example, the numerical examples section of the text). Therefore, the behavioral restrictions required to derive the market concentration–market power relationship from theory would indeed seem too strong to be accepted.

In the more general (and more plausible) case where $\frac{\partial y}{\partial y_j} \neq \frac{\partial y}{\partial y_i}, j \neq i$, the expression for $\alpha$ does not allow one to derive the HHI. Therefore, under these more general conditions, we cannot rely on the HHI to make predictions regarding firms' conduct. Nonetheless, as stated above, we can test econometrically whether the Cournot or the constant $\gamma$ restrictions can be rejected against alternative theoretical specifications. As Bresnahan (1989, p. 1031) stated, “Only econometric problems, not fundamental problems of interpretation, cloud this inference about what has been determined empirically.”
**Details of the empirical implementation**

As we saw above, estimating the degree of market power means being able to identify the conduct parameter $\theta$ in equation 4, here rewritten for convenience of exposition as

$$p = C'(y_j, \omega_j) - \frac{dp}{dy} y_j \theta_j,$$

where $p$ now indicates the interest rate on commercial loans, $y$ indicates the quantity of commercial loans, and $\omega$, the vector of factor prices, includes labor cost, capital expenses, and the interest rate on deposits.

For the identification of the parameter of conduct $\theta$, we need information on the marginal cost function $C'(y_j, \omega_j)$ and on the inverse of the semi-elasticity on loans demand, $\frac{1}{\bar{e}} = \frac{dp}{dy} y$. One can obtain this additional information at different degrees of refinement, depending in practice on data availability. Angelini and Cetorelli (1998) estimate the parameters of the marginal cost function using the widely used trans-log specification, deriving the following expression:

$$C'(y_j, \omega_j) = \frac{C}{y_j} \left[ a_1 + a_2 \ln(y_j) + \sum_{i=3}^{n} b_i \ln(\omega_{i,j}) \right].$$

In addition, the parameter $\bar{e}$ is recovered by estimating simultaneously a loans demand function, specified as

$$\ln(y) = d_1 + d_2 p + d_3 \ln(z) + d_4 [\ln(z) p],$$

where $z$ is an exogenous shifter of demand, such as investments or GDP.

Finally, although it would be feasible in terms of data availability to test various models of oligopoly, thus identifying distinct parameters of conduct, $\theta_1 \neq \theta_2$, Angelini and Cetorelli (1998) focus on the determination of an average indicator of conduct, $\theta$ (see Bresnahan, 1982, for details). Such an indicator gives a first approximation of the overall conditions for the exercise of market power in the industry. Since such a study has never been conducted before for the Italian banking industry, I believe there is high informational value in the average indicator $\theta$.

---

1. The remainder of the section is based on Appelbaum (1982) and Bresnahan (1989).
2. The derivation is based on Cowling and Waterson (1976).

---

**NOTES**

1. Examples of research work on the impact on efficiency of bank mergers include Berger and Humphrey (1992), DeYoung (1997), Hughes et al. (1996), Rhodea (1993b), and Shaffer (1993b). Other authors have sought to evaluate the impact on profitability (for example, Berger and Humphrey, 1992; Cornett and Tehrannia, 1992; Piloff, 1996; and Akhavan et al., 1997) and on production decisions, in particular on lending to small business (for example, Berger et al., 1997).

2. An alternative measure also used in research is the sum of the market shares of the largest firms in the industry, usually the largest three or four firms.

3. For a thorough description of the use of mitigating factors in antitrust analysis, see Holder (1993a).


5. To be precise, thrift institutions are currently included in the calculation of the HHI. Their market shares, however, have only a 50 percent weight (20 percent for the Justice Department’s evaluation procedure), which in any case always determines a reduction in the HHI calculated on banks only. Because of the inclusion of the thrift institutions, the 1,800/200 rule is sometimes called the 1,800/200/50 rule.

6. The Justice Department’s horizontal merger guidelines define markets with a post-merger HHI below 1,000 as unconcentrated and unlikely to present anticompetitive concerns. Markets with a post-merger HHI between 1,000 and 1,800 are defined as moderately concentrated. In such markets a variation in the HHI of less than 100 points is unlikely to present anticompetitive concerns. Markets with a post-merger HHI above 1,800 are defined as highly concentrated, and a variation of the HHI greater than 50 points is thought to have adverse competitive consequences. In the past several years, however, the Justice Department has not challenged a merger unless the post-merger HHI was at least 1,800 and the change in the HHI at least 200 points (see Litan, 1994).

7. A firm joining a collusive agreement always has an incentive to abandon the agreement (or “cheat”) and set prices and/or quantities that maximize its own profits. The costs associated with the collusive agreement are therefore expressed either in terms of the losses suffered by participants in the event that one of them cheats, or in terms of the punishment that a firm would sustain in the event it is caught cheating (for instance, all firms revert to competitive pricing forever after collusion breaks down, hence the deviating firm will no longer be able to make positive profits.)

Prager and Hannan (1998) examine a cross-section of such markets, finding that banks operating in markets where a merger produces a substantial increase in concentration have deposit rates that are lower than those set by banks not operating in such markets. They interpret the result as evidence that these mergers lead to increased market power.

REFERENCES


The 35th Annual Conference on Bank Structure and Competition

GLOBAL FINANCIAL CRISSES
Implications for Banking and Regulation
May 5–7, 1999
The Federal Reserve Bank of Chicago will hold its 35th annual Conference on Bank Structure and Competition at the Westin Hotel in Chicago, May 5–7, 1999. Attended each year by several hundred financial institution executives, regulators, and academics, the goal of the conference has been to generate an ongoing dialogue on public policy issues affecting the financial services sector.

The major theme for the 1999 conference will be an analysis of financial crises. What appears to have started as a currency devaluation for a relatively small Asian country has seemingly led to a general upheaval in financial markets throughout the world. Suddenly, global capital market integration, which had been credited with prolonging and enhancing economic growth through the 1990s, has been criticized as a major cause of the crisis. What are the implications of these crises for U.S. firms? What are the appropriate public policy responses? Finally, what are the most effective means to prevent such crises—including an evaluation of the need for a new regulatory architecture?

These events raise a number of issues for bankers—both large and small—as well as for alternative providers of credit, and public policymakers. Many of these issues will be discussed at the 1999 conference. As in past years, much of the program will be devoted to the primary theme of the conference, but there also will be a number of additional sessions on topical issues of financial structure and regulation. Some highlights of this year's scheduled program include:

- The keynote address by Federal Reserve Board Chairman Alan Greenspan.
- A discussion of the conference theme by a panel of industry experts. The participants scheduled to attend include Carter Golembe from CHG Consulting Inc., William McDonough, President of the Federal Reserve Bank of New York, John Heimann from Merrill Lynch and recently appointed to chair the Financial Stability Institute of the Bank for International Settlements, Andrew Sheng from the Hong Kong Securities and Futures Commission, and Allan H. Meltzer from Carnegie Mellon University.
- A luncheon address by John B. McCoy, President and Chief Executive Officer of Bank One Corporation, the nation's fifth largest bank holding company resulting from the 1998 merger of Bank One and First Chicago/NBD Corporation.
- A panel discussion on the need for regulatory reform and alternative perspectives on the appropriate form this should take.
- A discussion of the potential benefits and problems associated with bank mergers and reasons for the apparent lack of evidence supporting the contention that significant cost and efficiency gains should result from mergers.
- A discussion of the appropriate response to financial crises emphasizing the role of the lender of last resort and the impact of debt forgiveness.

As usual, the Wednesday sessions will showcase a wide array of technical research papers of primary interest to researchers in academia and government. The Thursday and Friday sessions are designed to discuss issues that focus on the interests of a broader audience.

If you currently are not on our mailing list or have changed your address and would like to receive an invitation to the conference please contact:

Portia Jackson
Conference on Bank Structure and Competition Research Department
Federal Reserve Bank of Chicago
P.O. Box 834
Chicago, Illinois 60604-1413
Telephone: 312-322-5775
e-mail: portia.jackson@chicfrb.org

Federal Reserve Bank of Chicago
Agglomeration in the U.S. auto supplier industry

Thomas H. Klier

Introduction and summary

The General Motors (GM) strike during June and July 1998 showed the extent to which lean manufacturing production methods, such as efforts to keep inventories low and reduce the number of parts suppliers, have taken hold in the U.S. auto sector. As observers tried to assess the ramifications of this event, it became apparent that we know much more about the spatial structure of light vehicle assembly operations and Big Three (Ford, GM, and Chrysler) owned parts plants than of the large number of independent parts suppliers. In an environment of tightly linked supply chains, it is important to understand the spatial nature of these linkages. Such knowledge would help policymakers assess the economic impact of regional shocks, such as a strike. In addition, data on individual customer-supplier linkages would facilitate the study of the geographic extension of supplier networks and offer new evidence on the ability of economic development efforts to attract suppliers to locate in the same state as a large assembly facility.

Lean manufacturing was pioneered by Toyota Motor Company in Japan during the 1950s. It has since become the standard for many manufacturing companies in Japan and around the world. This production system tries to improve on the types of mass production systems that have been prominent in the postwar period. Instead of organizing production according to a preset schedule, it operates on the premise of a so-called pull system, whereby the flow of materials and products through the various stages of production is triggered by the customer. In addition, the production process itself is subject to continuous improvement efforts.

The 1998 strike at two GM-owned parts plants in Flint, Michigan, was about issues related to production rates and health and safety. Strategically, however, it centered on issues pertinent to the implementation of new production methods—more efficient production processes that would reduce the demand for labor in the assembly plant and efforts by the assembly company to outsource more of the production of parts. The strike quickly shut down most of GM’s North American assembly operation. In turn, it caused production adjustments at many of the company’s independent suppliers.

In this article, I examine the spatial structure of the auto supplier industry and how firms in different locations interact. First, I document the extent to which plants are concentrated geographically, that is, the degree of spatial agglomeration, in the U.S. auto supplier industry. My analysis is based on information on the location of over 3,000 auto supplier plants. I find that the auto supplier industry is concentrated in five states—Indiana, Kentucky, Michigan, Ohio, and Tennessee—that constitute the so-called auto corridor, which is defined by interstate highways 65 and 75, extending south from Michigan to Tennessee. These states are home to 58 percent of the plants in the study. A closer analysis of plant locations reveals the importance of access to highway transportation to ensure timely delivery of production to customers. I find that having suppliers located in the immediate vicinity of the assembly plant is not necessary to maintain a system of tight linkages and low inventories. Comparing the spatial structure of individual assembly networks, I find them to be remarkably similar. The geographic concentration is highest for assembly plants that are located near the heart of the auto corridor, with between 70 percent and 80 percent of supplier plants located within a day’s drive of the assembly plant. This suggests a clustering of economic activity at the regional rather than local level.

Thomas H. Klier is a senior economist at the Federal Reserve Bank of Chicago. The author would like to thank Jim Rubenstein and Bill Testa for their valuable suggestions; Neil Murphy and George Simler for excellent research assistance; and seminar participants at the Federal Reserve Bank of Chicago for helpful comments.
Second, I investigate the changing nature of the geographic concentration of this industry over time. This analysis is limited by the cross-sectional nature of the data. However, there are a few cases in which the data allow a comparison of supplier networks of different vintages. In addition, I apply a location rule to a subset of all the supplier plants that allows me to use information on the location of all light vehicle assembly plants in the U.S. from 1950 to 1997. Consistently, I find evidence of increased clustering in the auto supplier industry relative to 30 or 40 years ago.

**Literature review**

Geographic concentration in U.S. manufacturing has received greater attention in recent years. Krugman (1991) suggests that Silicon Valley-style agglomerations may be more the rule than the exception and that what we can learn from them about the source of the underlying forces. Ellison and Glaeser (1997) address the question of how to properly measure industry concentration over and above the general level of concentration in manufacturing. To that end they develop a model that captures both random location effects and those caused by localized industry-specific spillovers and natural advantages. The authors develop indexes of localization and find almost all industries to be somewhat localized. In many industries, however, the degree of localization is small. The authors report that almost all of the most extreme cases of concentration are apparently due to natural advantages.

Hewings et al. (1998) analyze the 1993 commodity flow statistics, using a detailed econometric input–output model, to learn about a slightly different issue: To what extent are the states of a specific region (the Midwest) linked economically? They find very strong evidence of industry clusters at the regional level. For example, in the case of the auto industry, an initial loss of automotive production in Michigan would create secondary effects that are heavily concentrated in the Midwest. Specifically, losses in the Midwest would represent 43 percent of the secondary effect outside of Michigan.

Addressing these issues for the U.S. auto industry, several studies suggest that the assembly plants for light vehicles have reconcentrated in the Midwest and Southeast since the mid-1970s (Rubenstein, 1992; McAlinden and Smith, 1993; and Rubenstein, 1997). Rubenstein (1997) attributes this to the demise of the branch plant assembly system, whereby identical models were produced around the country at assembly plants that were located close to population centers. Developments in the supplier industry are not as clear cut. Apparently there has been a migration of especially labor-intensive parts production to the southern U.S. and south of the border; however, parts requiring highly skilled labor, such as engines, transmissions, and large stampings, have remained heavily concentrated in the Midwest. That is especially true for parts plants operated by the auto assemblers themselves (so-called captive suppliers) (see table 1).

As for the potential location effect of lean manufacturing, the prevailing anecdotal evidence suggests that the application of lean manufacturing techniques has resulted in a tiering and consolidation of the supplier base of the auto industry, as well as a higher degree of communication and interaction between suppliers and assemblers (Helper, 1991). Has this resulted in tighter geographical linkages between assembler and supplier plants? Proponents suggest that close linkages work most effectively when supplying and receiving plants are in reasonably close proximity (Estall, 1985; Kenney and Florida, 1992; Mair, 1992; and Dyer, 1994). However, there is also evidence that spatial clustering is not a necessary outcome of lean manufacturing applications. What ultimately matters is the quality of transportation infrastructure in combination with the capability of delivery management systems in ensuring predictable on-time arrival of goods. This might well be achieved with no significant increase in clustering at the industry level.

A set of studies specifically investigates the existence of effects of lean manufacturing on the spatial structure of the auto supplier industry. Rubenstein and Reid (1987) and Rubenstein (1988) analyze the changing supplier distribution of U.S. motor vehicle parts suppliers. Their thorough analysis of supplier plants located in Ohio cannot establish a clear-cut effect of lean manufacturing on plant location, yet the authors find evidence of a change in the locational pattern after 1970.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Distribution of captive parts plants (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly company</td>
<td>Share of captive suppliers in MI, IN, and OH</td>
</tr>
<tr>
<td>General Motors</td>
<td>69.8</td>
</tr>
<tr>
<td>Chrysler</td>
<td>82.3</td>
</tr>
<tr>
<td>Ford</td>
<td>84.6</td>
</tr>
<tr>
<td>Overall</td>
<td>75.6</td>
</tr>
</tbody>
</table>

Most of the existing analysis of the location effects of lean manufacturing, however, concerns Japanese-owned manufacturing establishments within the U.S. This is not surprising, as these plants generally apply lean manufacturing. In addition, most of them represent new plants established at newly developed, so-called greenfield sites. As their location decision usually does not involve a re-location, they are a preferred object of study. Woodward (1992) investigates what determines the location of Japanese manufacturing start-up plants in the U.S. The author estimates location models of the spatial behavior of Japanese-affiliated manufacturing investments undertaken between 1980 and 1989. While his observations include plants from many different manufacturing industries, he estimates a model specification at the county level for 250 observations in the Michigan–Tennessee automotive corridor. Woodward finds proximity to urban areas not to be important for these plants; however, an interstate connection linking counties to major markets appears to be crucial. Reid (1994) tests the effect of just-in-time inventory control on spatial clustering in observing the level of inputs purchased locally for a set of 239 Japanese-owned manufacturing plants in the U.S. The author performs this analysis at three different levels of aggregation—county, state, and national. He finds differences in the proportion of material inputs purchased locally between plants that use just-in-time inventory control and those that do not only at the county level. This result suggests spatial clustering effects on a very local scale. Smith and Florida (1994) test for the existence of agglomeration effects in the location decisions of over 400 Japanese-affiliated manufacturing establishments in automotive-related industries. They perform a formal analysis for all U.S. counties, as well as an automotive corridor subset, and find that Japanese-affiliated suppliers prefer to locate in close proximity to Japanese automotive assemblers. On a regional scale, they find a concentration of Japanese auto suppliers in the auto corridor.

**Spatial characteristics data**

In this article, I present evidence on the spatial characteristics of independent auto supplier plants located in the U.S., with particular emphasis on linkages between supplier and assembly plants. First, I document the extent to which plants are concentrated geographically, or the degree of spatial agglomeration, in the U.S. auto supplier industry. Second, I investigate the changing nature of this geographic concentration over time.

Publicly available data do not provide this level of detail. The obvious data source, the *Census of Manufactures*, can offer only incomplete information, because it does not distinguish between original equipment manufacturers and producers of replacement parts. In addition, because of the large variety of parts that make up an automobile, supplier plants in the auto industry are classified among 18 of the 20 two-digit standard industry classification (SIC) codes. Finally, Census data cannot establish information about linkages between supplier plants and their customers.

The basis for my analysis is the “ELM GUIDE supplier database,” a set of plant-level data on the auto supplier industry put together by a private company in Michigan. The data are for 1997 and cover 3,425 independent supplier plants in the U.S. As the database identifies customers for the individual supplier plants, I was able to categorize these plants by supplier tier: 2,008 plants are tier 1 suppliers, that is, supplier plants that ship their products exclusively to auto assembly plants and not to other supplier plants or other customers; 1,292 are mixed-tier suppliers, that is, in addition to auto assembly plants, their customers include other supplier plants and/or nonautomotive assemblers; and 50 observations were excluded from the analysis because they did not provide information on their customers.

I then added several variables to the database. For tier 1 plants, I obtained start-up year data from various state manufacturing directories and phone calls to individual plants. I added information on foreign ownership available through industry press reports and the Japan Auto Parts Industries Association. Table 2 shows an ownership breakdown of the industry.

Accounting for incomplete information on start-up year, I end up with 1,845 individual plant records, representing independent tier 1 supplier plants operational in 1997. Next, I analyze data on these 1,845 plants to test for agglomeration at the industry level, as these plants arguably represent the subset of supplier plants that is most closely linked to the auto assembly plants by way of production and delivery. In addition to the cross-sectional comparisons, information on the vintage of individual plants allows some comparison of location patterns of older and recently opened plants.

The analysis of assembly plant-specific networks draws on all the 3,137 records of independent supplier plants.

**Industry-level agglomeration**

Table 3 presents the distribution of the 3,137 independent supplier plants included in the database.
It shows the auto supplier plants and employment to be highly spatially concentrated, with almost 50 percent of all plants located in just three states—Michigan, Ohio, and Indiana. However, it is important to keep in mind that this information represents plants from rather different vintages. For example, the oldest plants in the sample date from the nineteenth century; 38 plants opened prior to 1900. In order to get a better read on recent plant location choices, I focus on the subset of supplier plants that have opened since 1980, marking when lean manufacturing arrived in the U.S. As data on the establishment year are available only for tier 1 supplier plants, I focus on the subset of 820 tier 1 supplier plants that opened in 1980 or after and were still in operation in 1997. While a pure cross-sectional data set prevents me from testing for changes in location patterns over time, concentrating on plants of recent vintage enables me to present the location choices in a lean manufacturing environment in much more detail.

Figure 1 shows the plants that opened between 1980 and 1997 and their concentration among the five states of the auto corridor. Domestic plants are shown in black, foreign-owned plants in color. A circle indicates that two or more plants are located within one zip code. In addition, stars mark the location of light vehicle assembly plants in operation at any point during this period. One can clearly see that plant openings are highly clustered in a north–south direction (in southern Michigan and the four states to the south). Figure 2 adds the grid of interstate highways to the pattern of plant openings. This exercise demonstrates the relevance of the I-65/I-75 corridor. Note, however, that interstate access plays an important role for east–west connectivity as well. For example, Toyota operates a car assembly plant in Georgetown, Kentucky, a recently opened light truck assembly plant in Princeton, Indiana, and an engine plant in Buffalo, West Virginia. All three of these are linked by Interstate 64, highlighting the importance of highway access to ensure timely delivery of shipments in an environment of just-in-time production.

Looking at the auto corridor locations more closely, figure 3 (page 23) reveals a different location pattern for domestic and foreign-owned supplier plants during 1980–97. While they are similarly concentrated among three states, foreign-owned suppliers choose to locate in the southern part of the automotive corridor (that is, Ohio, Kentucky, and Tennessee). Domestic suppliers, on the other hand, locate in the northern part, with Ohio being the only state chosen prominently by both domestic and transplant supplier plants. Does this indicate that the auto corridor is a phenomenon driven by the location of foreign-owned plants? What explains the apparent different spatial pattern in plant locations? Do foreign-owned suppliers respond differently to lean manufacturing conditions than domestic suppliers? Figure 3 and table 4 (page 23) suggest a different explanation: The difference in the spatial distribution of domestic and foreign-owned assembly plants seems to dominate the location choices of supplier plants. As a rule of thumb, between 1980 and 1993 supplier plants located close to assembly plants of the same nationality. This can be seen in figure 3, which distinguishes between domestic (gray) and foreign-owned (colored) assembly plants.

### TABLE 2

<table>
<thead>
<tr>
<th>Auto suppliers by ownership, 1997 (percent)</th>
<th>Plants</th>
<th>Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td>84.7</td>
<td>81.6</td>
</tr>
<tr>
<td>Foreign-owned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>9.6</td>
<td>11.2</td>
</tr>
<tr>
<td>Other</td>
<td>5.7</td>
<td>7.2</td>
</tr>
</tbody>
</table>

Notes: Calculations are based on 3,137 independent supplier plants open in 1997; numbers do not include captive supplier plants. Industry employment: 901,343 jobs. Source: See table 1.

### TABLE 3

<table>
<thead>
<tr>
<th>Distribution of auto suppliers, 1997 (percent)</th>
<th>Plants</th>
<th>Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois</td>
<td>6.9</td>
<td>6.8</td>
</tr>
<tr>
<td>Indiana</td>
<td>9.1</td>
<td>10.1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>4.0</td>
<td>4.1</td>
</tr>
<tr>
<td>Michigan</td>
<td>26.8</td>
<td>19.2</td>
</tr>
<tr>
<td>Ohio</td>
<td>13.2</td>
<td>11.2</td>
</tr>
<tr>
<td>Tennessee</td>
<td>4.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>3.6</td>
<td>3.1</td>
</tr>
<tr>
<td>Midwest</td>
<td>59.6</td>
<td>50.4</td>
</tr>
<tr>
<td>Auto corridor</td>
<td>57.8</td>
<td>50.4</td>
</tr>
<tr>
<td>U.S.</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Notes: Calculations are based on 3,137 independent supplier plants open in 1997; numbers do not include captive supplier plants. Industry employment: 901,343 jobs. The auto corridor comprises Indiana, Michigan, Ohio, Kentucky, and Tennessee. The Midwest comprises Illinois, Indiana, Michigan, Ohio, and Wisconsin. Source: See table 1.
FIGURE 1
Plant openings by tier 1 suppliers, 1980–97

Source: See table 1.

FIGURE 2
Importance of highway transportation for the auto corridor

Note: For some of the highways, the figure shows only the part that intersects the auto corridor.
Source: See table 1.
Focusing on relationships to primary customers only would provide more conclusive evidence. However, the data do not allow identification of the distribution of output among customers. Instead, I present information on the distribution of supplier plants that report a particular customer mix. Table 4 shows data on domestic suppliers that supply only to Big Three assembly plants, as well as data on Japanese transplant suppliers that do not supply to any Big Three assembly plants. If the nationality of the assembly plant customer was important to the location choice of the supplier plant, one would expect these two groups to be relatively concentrated in their respective halves of the auto corridor. Table 4 provides evidence

TABLE 4

Percent of supplier plants opened in auto corridor, 1980–97

<table>
<thead>
<tr>
<th></th>
<th>Domestic</th>
<th>Supplying only to Big Three</th>
<th>Japanese-owned</th>
<th>Not supplying to Big Three</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td>Overall</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>31.3</td>
<td>40.0</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>10.9</td>
<td>11.4</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>10.4</td>
<td>10.2</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>6.3</td>
<td>4.2</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>4.1</td>
<td>1.8</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>Top three</td>
<td>52.6</td>
<td>61.6</td>
<td>52.6</td>
<td></td>
</tr>
<tr>
<td>Number of plants</td>
<td>607</td>
<td>166</td>
<td>173</td>
<td></td>
</tr>
</tbody>
</table>

Source: See table 1.
of just such a "customer" effect, as each group of supplier plants with a specific customer mix is more concentrated at one end of the auto corridor.16

This simple comparison between the location choices of assembly and supplier plants, however, cannot address the issue of timing. Did assembly or supplier plants locate first?17 The data allow me to shed some light on this question for the Japanese-owned supplier plants. Table 5 shows that 55 percent of these plants opened between 1987 and 1989, well after the first Japanese auto assembly plants had started operating in the U.S.18 That pattern suggests that in the case of Japanese transplants, the suppliers followed the assemblers (see also Rubenstein, 1992). However, the initial location decision of Japanese assembly plants was undoubtedly influenced by proximity to the existing, that is, mostly domestic, supplier base.19

Network data

Next, I discuss the extent to which supplier plants locate near their assembly plant customers. As the data set includes information on customers of the individual supplier plants, I am able to construct supplier networks for specific assembly plants.20 However, my choice of assembly plants is limited to a set of essentially single-plant assembly companies as the supplier plants' customer information is provided only at the company level. I can construct networks for the following assembly plants: Honda of America, which opened its Marysville, Ohio, plant in 1982 (and added a second assembly plant in nearby East Liberty, Ohio, in 1989); Nissan, which opened an assembly plant in Smyrna, Tennessee, in 1983; NUMMI, a joint venture between Toyota and GM, operating in Fremont, California, since 1984; AutoAlliance, which started as a joint venture between Ford and Mazda in 1987 in Flat Rock, Michigan; Diamond-Star, which started production as a Mitsubishi-Chrysler joint venture in Normal, Illinois, in 1988; Saturn, GM's attempt to capture the efficiencies of lean manufacturing, which started production in 1990 in Spring Hill, Tennessee; BMW, which opened an assembly plant in South Carolina in 1994; and Mercedes-Benz, which opened a plant in Alabama in 1997.

Table 6 presents characteristics of the networks identified from the database.21 Each network includes all independent supplier plants that list the respective assembler as a customer. Not surprisingly, the networks vary in size, with Honda, the oldest, being the largest, and Mercedes-Benz, the most recently opened assembly plant on the list, the smallest. To measure the networks' spatial characteristics, I calculate the median distance between supplier and assembler and the percentage of suppliers located within both a 100-mile and a 400-mile radius of the assembly plant (table 6, column seven, ranks networks by percentage share of suppliers within 400 miles). The 400-mile radius roughly defines the boundary for a one-day shipping distance, while the 100-mile distance captures plants that locate close enough to allow multiple deliveries using the same truck.22

According to these statistics, the individual networks look more alike than different. In general, the spatial concentration increases toward the heart of the automotive corridor. The AutoAlliance and Honda networks are most concentrated within 100 miles (column five), for the 400-mile criterion, the disadvantage from being located at the fringe of the automotive corridor mostly disappears. Two cases in point are the Diamond-Star and Subaru-Isuzu networks, which are, for the larger radius, essentially as concentrated as Honda's and Toyota's. The spatial features of supplier networks reported in table 6 seem to be explained by two factors: where the assembly plant is located relative to the auto corridor and whether the assembly plant is domestic or foreign-owned.

---

**TABLE 5**

<table>
<thead>
<tr>
<th>Japanese transplant tier 1 supplier plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start-up year</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1980</td>
</tr>
<tr>
<td>1981</td>
</tr>
<tr>
<td>1982</td>
</tr>
<tr>
<td>1983</td>
</tr>
<tr>
<td>1984</td>
</tr>
<tr>
<td>1985</td>
</tr>
<tr>
<td>1986</td>
</tr>
<tr>
<td>1987</td>
</tr>
<tr>
<td>1988</td>
</tr>
<tr>
<td>1989</td>
</tr>
<tr>
<td>1990</td>
</tr>
<tr>
<td>1991</td>
</tr>
<tr>
<td>1992</td>
</tr>
<tr>
<td>1993</td>
</tr>
<tr>
<td>1994</td>
</tr>
<tr>
<td>1995</td>
</tr>
<tr>
<td>1996</td>
</tr>
<tr>
<td>1997</td>
</tr>
<tr>
<td><strong>173</strong></td>
</tr>
</tbody>
</table>

Note: Column labeled "Percent" may not total due to rounding.
Source: See table 1.
TABLE 6
Spatial characteristics of supplier networks, 1997

<table>
<thead>
<tr>
<th>Assembly company</th>
<th>Start-up year</th>
<th>Number of suppliers</th>
<th>% Domestic</th>
<th>Median distance</th>
<th>Network %&lt;100 miles</th>
<th>Industry %&lt;100 miles</th>
<th>Network %&lt;400 miles</th>
<th>Industry %&lt;400 miles</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honda</td>
<td>1982</td>
<td>507</td>
<td>65</td>
<td>251</td>
<td>17</td>
<td>9.4</td>
<td>77</td>
<td>74.8</td>
</tr>
<tr>
<td>Toyota</td>
<td>1988</td>
<td>452</td>
<td>69</td>
<td>285</td>
<td>10</td>
<td>4.2</td>
<td>76</td>
<td>75.7</td>
</tr>
<tr>
<td>Subaru-Isuzu</td>
<td>1987</td>
<td>292</td>
<td>60</td>
<td>245</td>
<td>9</td>
<td>6.2</td>
<td>76</td>
<td>71.8</td>
</tr>
<tr>
<td>Diamond-Star</td>
<td>1988</td>
<td>286</td>
<td>63</td>
<td>309</td>
<td>5</td>
<td>1.7</td>
<td>72</td>
<td>69.3</td>
</tr>
<tr>
<td>AutoAlliance</td>
<td>1987</td>
<td>360</td>
<td>71</td>
<td>242</td>
<td>29</td>
<td>24.7</td>
<td>65</td>
<td>66.4</td>
</tr>
<tr>
<td>Nissan</td>
<td>1983</td>
<td>460</td>
<td>70</td>
<td>423</td>
<td>10</td>
<td>3.8</td>
<td>45</td>
<td>36.7</td>
</tr>
<tr>
<td>BMW</td>
<td>1994</td>
<td>119</td>
<td>75</td>
<td>477</td>
<td>20</td>
<td>3.7</td>
<td>42</td>
<td>26.6</td>
</tr>
<tr>
<td>Saturn</td>
<td>1990</td>
<td>300</td>
<td>81</td>
<td>462</td>
<td>8</td>
<td>3.4</td>
<td>35</td>
<td>32.4</td>
</tr>
<tr>
<td>Mercedes-Benz</td>
<td>1997</td>
<td>77</td>
<td>68</td>
<td>610</td>
<td>8</td>
<td>0.8</td>
<td>34</td>
<td>17.5</td>
</tr>
<tr>
<td>NUMMI</td>
<td>1984</td>
<td>178</td>
<td>60</td>
<td>1,966</td>
<td>6</td>
<td>0.8</td>
<td>11</td>
<td>2.4</td>
</tr>
<tr>
<td>B.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flint (1950)</td>
<td>1907</td>
<td>126</td>
<td>72</td>
<td>192</td>
<td>28</td>
<td></td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Ford (1970–80)</td>
<td>N.A.</td>
<td>222</td>
<td>89</td>
<td>405</td>
<td>18</td>
<td></td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Ford (1983–93)</td>
<td>N.A.</td>
<td>301</td>
<td>77</td>
<td>200</td>
<td>31</td>
<td></td>
<td>66</td>
<td></td>
</tr>
</tbody>
</table>

Note: N.A. indicates not applicable.

For example, figure 4 shows how Honda’s independent supplier plants cluster around its two Ohio assembly plants. The three circles envelop the first three quartiles of the distance distribution of supplier plants in the network. The figure shows an assembly operation that is centrally located in the auto corridor. It turns out to be the most spatially concentrated network: 17 percent of Honda’s 507 suppliers are located within 100 miles and 77 percent within 400 miles of the assembly plant.

In contrast, Diamond-Star is located at the western edge of the auto corridor (see figure 5). Therefore, it is able to attract only 5 percent of its suppliers to locate within 100 miles. However, that disadvantage disappears at the 400-mile radius, which, for Diamond-Star as for Honda, includes 77 percent of its supplier plants.

The case of Saturn presents yet a different picture. Its suppliers are relatively dispersed (see figure 6). Notice the large diameter of the first quartile. Only 35 percent of Saturn’s supplier plants are operating within 400 miles of Spring Hill, Tennessee. This reflects the fact that Saturn most strongly relies on domestic suppliers, which are located at the northern end of the auto region. Its assembly plant, however, is located at the southern end of the corridor.

Alternatively, one can analyze the concentration of individual supplier networks relative to the distribution of all the supplier plants. In calculating what share of the entire industry is located within a certain radius of the assembly plant, one can then assess a network’s degree of concentration relative to the industry baseline. Table 6, panel A, shows this information for both the 100-mile and the 400-mile radius. Columns five and six show that for every single assembly plant analyzed, a greater share of suppliers is concentrated within 100 miles than the overall industry distribution would suggest. At the 400-mile radius (see columns seven and eight of table 6), one can distinguish two network groups. The supplier networks of assemblers located in the northern end of the auto corridor plus Kentucky represent very closely the industry’s overall spatial distribution. However, the five assembly plants located in Tennessee, Alabama, South Carolina, and California are far more concentrated than the industry, even at that relatively large radius. What drives that result is the large number of suppliers operating at the northern end of the auto corridor. For example,
suppliers in Nissan’s network that are located within 400 miles of the Tennessee assembly plant represent a far greater concentration of auto suppliers in the region than indicated by the distribution of all supplier plants.

The different spatial distribution of domestic and foreign-owned supplier plants across the auto corridor is reflected within the individual networks as well. Foreign-owned supplier plants are clustered much more densely around Japanese assembly plants than domestic suppliers (see, for example, Honda, Toyota, and Nissan in table 7 on page 29). Yet even for that group, less than one-third of suppliers are located within a two-hour drive, or 100 miles, of the assembly plant. This represents a considerably smaller degree of spatial concentration within lean manufacturing than previously reported in the literature. The case of Saturn represents a domestic auto assembler whose network is not very spatially concentrated. This applies to both its domestic and foreign-owned supplier plants (quite in contrast to Nissan, which is located not very far from Saturn). Finally, AutoAlliance shows the effect of being located in the heart of the traditional U.S. auto region. Its network includes by far the largest percentage of suppliers within a 100-mile radius. At 31.9 percent, that share is significantly higher for domestic suppliers than for foreign-owned suppliers (21.9 percent).

The analysis of the regional concentration of supplier networks at that disaggregate level can again be complemented by a comparison with the industry level of spatial concentration. For the 100-mile radius, table 7 (columns three and four) shows a higher degree of concentration for both domestic and foreign-owned suppliers within each network than is indicated by the overall distribution of the industry. At the 400-mile radius (columns five and six), the differences between these two measures of spatial concentration disappear in most cases. Noteworthy exceptions are the most recently opened assembly plants to the south and east of the auto corridor (Mercedes and BMW) and NUMMI. Saturn is the only domestic assembly plant in the study. Its network shows a smaller percentage of within-network foreign-owned suppliers within 400 miles of the assembly plant than the overall industry.
level would suggest. The spatial distribution of Saturn’s network presents a stark contrast to that of Nissan, the other assembly plant in Tennessee.

**Changing industry structure?**

To what extent are these observations indicative of changes in the spatial pattern of auto supplier plants? I address that question in three different ways. First, I compare the structure of different networks over time. From Henrickson’s (1951) analysis of the supplier structure of the Buick city assembly plant in Flint, Michigan, it is possible to reconstruct that assembly plant’s supplier network (see table 6, panel B, page 25) and compare it with a current network (Honda) that operates based on a different manufacturing system. It turns out that the median distance is statistically different for these two networks; however, the percentages within 400 miles are not statistically different. In other words, during the prime of the manufacturing system perfected by Henry Ford, one of its showcase plants, GM’s Buick city plant, had a supplier structure that was remarkably spatially concentrated. However, it is important to keep in mind that such a comparison is not adjusted for different degrees of vertical integration, changes in the mode and speed of transportation, as well as quality of the transportation infrastructure since 1950. In other words, a 400-mile radius in 1950 in all likelihood represented a smaller degree of spatial concentration than the same radius in 1997.

Second, I test for differences in spatial concentration for one network over time, using data on one of the Big Three assemblers, Ford. Instead of constructing networks for each of Ford’s individual assembly plants, I use Dearborn, Michigan, as the center of Ford’s assembly operations. Since 1970 there have been two decades, 1970–80 and 1983–93, during which Ford neither opened nor closed an assembly plant. Juxtaposing these two periods allows for an interesting comparison of the changing spatial pattern of Ford’s supplier network (see table 6, panel B on page 25 and figure 7 on page 30). It shows a marked
increase in concentration of Ford's supplier base around southern Michigan. During the more recent decade, 31 percent of newly opened supplier plants located within 100 miles of Dearborn (versus only 17 percent during the earlier decade). Comparing 1970–80 and 1983–93, the closures of two California plants and a New Jersey plant in the intervening years might have reduced average distances to Dearborn somewhat (for example, by reducing the percentage of plants greater than 400 miles away). However, one would not expect that alone to contribute to the simultaneous increase in plants located within 100 miles of Dearborn. Comparing 1970–80 and 1983–93, the statistical tests show all three measures of spatial concentration reported in table 6 to be different at the 99 percent confidence level, providing strong evidence of increasing spatial concentration within one of the Big Three supplier networks.

Third, I ignore the customer information provided by the database and employ a simple location algorithm, motivated by a Weberian model of plant location, to link suppliers with assembly plants. By applying a uniform location rule across time for supplier plants, I can test whether their location decisions changed over time. To perform this test, I break the sample into two periods: plants that have opened since 1980, whose location decisions were presumably influenced by lean manufacturing constraints, and plants that opened between 1950 and 1979, when supplier location decisions were influenced by the need to be close to Big Three operated parts distribution facilities. Comparing plant locations for these two samples, I can test for a change in location pattern in two directions. That is, I can ask if the pattern exhibited by the younger plants fits that of the older ones and vice versa. Specifically, for the most recent period I apply two versions of a location rule that minimizes the distance between supplier and assembly plant. This approach represents the influence of just-in-time production; supplier plants in that environment want to be located closer to the assembly plant to minimize production and transportation costs. It links the supplier to the closest operational assembly plant. I do not incorporate information provided in the database (and used above)
on actual assembler-supplier linkages. However, in applying a general location rule I am no longer restricted to the number of assembly plants listed in table 6, but can consider all light vehicle assembly plants in the U.S. A slightly different version averages the three shortest distances between a supplier and operational assembly plants. I apply the location rule to both sets of supplier plants, resulting in a distribution of distances for each sample. I then test if the median of the more recent sample is statistically different from the median of distances for the older plants. If I find no statistical difference, then the just-in-time location rule describes both time periods equally well, and there is no evidence for change in location pattern. However, if there is evidence of a difference in the pattern, I interpret this as a strong signal for a change in the location pattern, as it is established by applying the same decision rule for both periods. The test results are described in table 8, panel A (page 31). Under both versions of the just-in-time rule, median distances decrease over time. In fact, the differences in the median are significant at the 99 percent level of confidence, according to a Wilcoxon signed-rank test.

In testing for a change in location pattern in the opposite direction, I use the following rule to approximate decisions made by the older supplier plants: minimize distance to Detroit. Prior to the tiering of the supplier industry, supplier plants would usually ship their output to a regional parts distribution center operated by the Big Three, which in turn directed the parts to assembly plants around the country. In recognition of the spatial clustering of auto supplier plants in southeast Michigan, northern Indiana, and Ohio, I calculate the distance to Detroit for each plant that opened during the earlier period. These results
FIGURE 7
Increase in concentration of Ford’s supplier network over time
A. Ford tier 1 suppliers, established 1970–80
B. Ford tier 1 suppliers, established 1983–93

Note: The circles around the assembly plant indicate the closest 25 percent, 50 percent, and 75 percent of the supplier network, respectively. Source: See table 1.

are presented in panel B of table 8. The actual distances to Detroit increased from 1980 onward, which is not surprising considering the changing shape of the auto region in that period. Again, I find the median distances to be statistically different at the 99 percent level, complementing the result of the first part of the test for a change in location patterns over time. To summarize, I find symmetrical evidence for structural change in the way supplier plants locate around assembler plants. Both tests suggest an increase in the clustering of suppliers around assembly plants since 1980 relative to 30 or 40 years ago.

Conclusion
By refining a commercially available database, this article provides a detailed look at the supplier networks of some recently opened auto assembly plants in the U.S. My analysis focuses on a description of existing spatial relations between assembly plants and their tier 1 supplier plants. This study supports earlier findings about regional agglomeration of supplier plants in the I-65/I-75 automotive corridor. For supplier plants of recent vintage, the five auto corridor states, Michigan, Indiana, Ohio, Kentucky and Tennessee, represent the preferred location. Within that region, however, domestic and foreign-owned supplier plants locate in noticeably different patterns, apparently due to differences in the location of domestic and foreign-owned assembly plants.

The evidence I present on the auto industry supports the view that agglomeration economies play out at the regional level (see Hewings et al., 1998). It does not support the notion that immediate proximity to the assembly plant is necessary for operating a system based on tight linkages and low inventories. In analyzing the extent of localization of production around individual assembly plants, I find networks to be remarkably similar, with about 70 percent to 80 percent of suppliers located within one day’s drive of the assembly plant. Differences seem to be explained by the location of the assembly plant in relation to the heart of the auto corridor as well as by nationality of the assembly plant. Within individual networks, the spatial concentration differs across domestic and foreign-owned supplier plants.

This evidence on spatial agglomeration has relevance for economic development (see table 9). The economic development literature has generally reported on the effects of locating a new assembly plant on either its immediate and surrounding counties (see, for example, Fournier and Isserman, 1993) or on the host state (see, for example, Marvel and Shkurti, 1993). However, the analysis presented here
allows us to investigate the extent of the regional distribution of related upstream plant employment in much greater detail. Take, for example, the case of the Mercedes plant that opened in 1993 in Alabama. The state provided incentives worth about $250 million to attract that plant. However, the evidence presented on the spatial extension of supplier networks suggests that suppliers to Mercedes will locate not just in Alabama, but more likely in Tennessee, Kentucky, and even further north.\(^{33}\) In fact, to date only 35 percent of Mercedes’ suppliers are located within 400 miles of the assembly plant, and only 16.5 percent of its supplier employment resides in Alabama.\(^{34}\) In Mercedes’ case, attractive targets for location efforts seem to have been foreign-owned companies (see table 7 on page 29). In short, this type of analysis suggests that subsidies that are offered by a state not in the auto corridor are considerably less effective in terms of attracting a significant portion of the related supplier employment to that state.

In the case of Toyota’s Kentucky assembly plant, a comparison of my network data on the distribution of supplier jobs with forecasts projected by a 1992 study also suggests a greater degree of spatial dispersion of supplier employment than expected.\(^{35}\)

Finally, several tests address the question of structural change in the spatial pattern of supplier plant locations. While limited by the cross-sectional nature of the data available, these results suggest that the degree of spatial concentration of supplier plants around assembly plants has increased since 1980. The timing of that change is consistent with the application of lean manufacturing techniques and just-in-time production linkages. However, the order of magnitude of the increased concentration does not support the concept of a supplier base that is tightly clustered around its customers. Within the auto corridor, the existing infrastructure apparently allows for frequent deliveries to multiple customers from a single supplier plant location.

### TABLE 8

<table>
<thead>
<tr>
<th>Supplier plants opened</th>
<th>1950–80</th>
<th>1980–97</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Just-in-time location rule</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortest distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>60.4(^*) (649)</td>
<td>52.2(^*) (806)</td>
</tr>
<tr>
<td>Domestic</td>
<td>59.6(^*) (605)</td>
<td>47.1(^*) (594)</td>
</tr>
<tr>
<td>Closest three avg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>108.2(^*) (649)</td>
<td>84.3(^*) (806)</td>
</tr>
<tr>
<td>Domestic</td>
<td>105.0(^*) (605)</td>
<td>73.2(^*) (594)</td>
</tr>
<tr>
<td><strong>B. Distance to Detroit rule</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>97.2(^*) (649)</td>
<td>296.7(^*) (806)</td>
</tr>
<tr>
<td>Domestic</td>
<td>188.6(^*) (605)</td>
<td>203.0(^*) (604)</td>
</tr>
</tbody>
</table>

\(^*\)indicates that the median distances are significantly different at the 99 percent confidence level, according to a Wilcoxon signed-rank test.

Note: Numbers in parentheses indicate number of observations.

Source: See table 1.

### TABLE 9

<table>
<thead>
<tr>
<th>Company</th>
<th>State</th>
<th>State investment ($mil.)</th>
<th>1997 employment*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Entire network</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(percent)</td>
</tr>
<tr>
<td>Honda</td>
<td>Ohio</td>
<td>21</td>
<td>23.4</td>
</tr>
<tr>
<td>Honda</td>
<td>Ohio</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Nissan</td>
<td>Tennessee</td>
<td>33</td>
<td>11.2</td>
</tr>
<tr>
<td>AutoAlliance</td>
<td>Michigan</td>
<td>49</td>
<td>21.1</td>
</tr>
<tr>
<td>Diamond-Star</td>
<td>Illinois</td>
<td>83</td>
<td>12.9</td>
</tr>
<tr>
<td>Toyota</td>
<td>Kentucky</td>
<td>150</td>
<td>14.8</td>
</tr>
<tr>
<td>Toyota</td>
<td>Indiana</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Saturn</td>
<td>Tennessee</td>
<td>80</td>
<td>12.1</td>
</tr>
<tr>
<td>Subaru-Isuzu</td>
<td>Indiana</td>
<td>86</td>
<td>17.6</td>
</tr>
<tr>
<td>BMW</td>
<td>S. Carolina</td>
<td>130</td>
<td>18.1</td>
</tr>
<tr>
<td>Mercedes</td>
<td>Alabama</td>
<td>252</td>
<td>16.5</td>
</tr>
</tbody>
</table>

\(^*\)Since I do not have information on the distribution of a supplier’s output across its customers, I adjust the reported plant-level employment by dividing it by the number of customers per supplier plant. In essence, I am treating all customers as of equal importance to a supplier. The last two columns report the percentages of supplier employment in the state of the assembly plant based on these adjusted employment figures.

\(^*\)Percentages in this column refer to tier 1 suppliers only as I do not have information on the start-up year of mixed-tier supplier plants.

NOTES

1 Marshall (1920) identified three reasons for localization: An industrial center allows a pooled labor market for workers with specialized skills, an industrial center allows provision of non-traded inputs specific to an industry in greater variety and at lower cost; and an industrial center generates technological spillovers as information flows more easily locally (Krugman, 1991, pp. 36–37).

2 The authors find the largest coagglomeration for the following two upstream–downstream industry pairs: motor vehicle parts and accessories (SIC 3714) and motor vehicles, car bodies (SIC 3711); and automotive stampings (SIC 3465) and motor vehicles, car bodies (SIC 3711).

3 It identifies for each of these the address, the list of products produced as well as the production processes used, employment, and the plants’ customers (at the company level). See ELM International, Inc. (1997).

4 My analysis does not cover the so-called captive supplier plants. An earlier paper (Klier, 1995) presented a much more limited analysis of the same issues for a comparatively small set of data for independent supplier plants operational in 1993.

5 It is difficult to accurately assess the coverage of this database, since the size of the true population is unknown.


7 About 8.1 percent of the 2,008 tier 1 plant records as provided by the ELM database could not be tracked down, either in the manufacturing directories or by phone, and are therefore not included in the subsequent analysis.

8 However, this is not equivalent to a time-series analysis since the sample only contains plants operating during 1997 and not those plants that were shut down in earlier years.

9 They represent 1,845 tier 1 and 1,292 mixed-tier plants.

10 Honda opened its first auto assembly plant in the U.S. in Ohio in 1982. McAlinden and Smith (1993) refer to the 1980s as a period of significant structural change for the U.S. automotive parts industry.

11 Woodward (1992) presents empirical evidence of the importance of highway access at the county level in attracting plant openings.

12 About 63 percent of foreign-owned plants are Japanese; see table 2.

13 Automobile assembly and component plants that are fully or partly owned by foreign companies are generally referred to as transplants.

14 Smith and Florida (1994) find evidence for such an effect for their sample of Japanese-owned supplier plants.

15 In the case of Japanese assembly plants, this has been well documented in the context of corporate ties between assembly and supplier companies (see Reid et al., 1995). For example, Ohio is perceived by both Japanese assemblers and bankers as “Honda’s state” (see Rubenstein, 1992).

16 There were too few observations for the following two categories to be reported in the table: domestic suppliers not supplying to Big Three assembly plants (16 plants) and Japanese suppliers only supplying to Big Three facilities (nine plants). However, in both cases the evidence is consistent with table 4. Plants in these two categories are noticeably less concentrated in the top three states (31.2 percent for the domestic supplier plants and 33.3 percent for the Japanese-owned plants).

17 See Rubenstein (1997) on the reconcentration of auto assembly plants in the Midwest and Southeast.

18 Only two of the Japanese-owned assembly plants in the study opened after 1987—Diamond-Star and Toyota, both in 1988.

19 Reid et al. (1995) suggest that was one of the ways Japanese assemblers minimized risk and uncertainty related to their direct foreign investment in the U.S.

20 The vast majority of supplier plants (over 90 percent) ship to multiple customers.

21 The tables and maps refer only to supplier plants located in the U.S. The overwhelming majority of independent suppliers located in Canada are concentrated in southwestern Ontario, between Windsor and Toronto. These plants are well connected to assembly plants in Canada and the northern end of the U.S. auto corridor via route 401.

22 All distances are calculated between the respective coordinates of a plant’s zip code; they are not adjusted for actual travel routes.

23 For example, Kenney and Florida (1992) report data on approximately 70 Japanese-owned auto supplier plants in the auto corridor and show 41.4 percent of plants within 100 miles of their respective assembly plants. In contrast, the highest concentration of Japanese-owned suppliers around Japanese assemblers I can find applies to the Honda network, with 29.3 percent of Japanese-owned suppliers within 100 miles of the assembly plant, followed by Toyota (22.9 percent) and Auto Alliance (22.5 percent).

24 The Buick plant in Flint was at the time one of the largest integrated automotive plants in the world. It employed about 22,000 people and produced about 2,000 cars a day. Henrickson’s data include both independent and captive suppliers of metal auto parts, tire and tube supplies, and mechanical rubber goods.

25 Between 1980 and 1983, three Ford assembly plants closed (two in California and one in New Jersey). In 1992, a body plant for Ford’s large vans in Avon Lake, Ohio, added an assembly line for the production of the Mercury Villager/Nissan Quest.

26 As the opening of the Avon Lake assembly line in 1992 could possibly explain some of the increase in supplier plants locating close to Dearborn, Michigan, I checked for robustness of my results by shortening the second time period to and in 1991. The exercise leaves the spatial distribution of Ford suppliers that opened during the 1980s essentially unchanged. This strongly suggests that the opening of the Avon Lake assembly line is not driving the reported reconcentration.

27 I would like to thank David Marshall, who suggested using this technique.
In calculating these distances, I consider only assembly plants that were operational when the supplier plant opened.

From 1950 to 1979, that corresponds to 77 light vehicle assembly plants; for the later time period, there are 76 plants.

The Wilcoxon signed-rank test is a nonparametric test that can be used to test whether the median of a set of observations equals some prespecified value. The test is based on calculating, ranking, and signing the differences between the actual observations and the constant. In panel A of table 8, I report the results from testing whether the median of the more recent distances between assembly and supplier plants (52.2 miles in the case of all observations) is different from the median of the distribution of distances for the older set of observations (60.4 miles). The test statistic \( T \) , which is distributed approximately normally, is obtained by taking the differences \( D_i = x_i - \text{median}_{x_i} \), where \( x_i \) represents the actual distances observed in the older data set. These differences are then ranked and signed; the test statistic \( T \) represents the sum of the signed ranks. The null hypothesis states that the median difference \( D_i \) equals zero. If it cannot be rejected, it follows that the median distances for both data sets are equal.

I would like to thank Jim Rubenstein, who suggested this approach.

See Reid (1994), Mair (1993), and Kenney and Florida (1992), who seem to suggest the need for very close proximity between assembler and suppliers.

Elhance and Chapman (1992) find similar evidence in analyzing the labor market of the Diamond-Star assembly plant in central Illinois. They find that the labor market for that plant covers a large geographical area, stretching over 15 states. They take this as evidence to suggest that the benefits of incentive packages intended to attract large manufacturing plants will not remain within the communities or states providing such incentives.

As table 9 shows, the percentage of supplier employment residing within the state of the assembly plant tends to increase if calculated for the set of supplier plants that opened after the respective assembly plant. In a couple of cases the percentages increase dramatically, but it is important to point out that these changes are in reference to only a small number of supplier plant openings. In Mercedes’ case, no supplier plant opened during 1997.

The Center for Business and Economic Research’s (1992) analysis of the economic impact of Toyota’s assembly plant on the other auto corridor states plus Illinois employs a specific input-output model (RIMS II) and its multipliers. Comparing the distribution of jobs associated with the production of inputs for Toyota’s assembly plant, information from my network data shows the overall network employment at about 36 percent of that estimated in the earlier study. However, one needs to point out that the numbers are not directly comparable, as the ELM database does not include purchases of raw materials and production equipment. With that caveat, a comparison of the distribution of employment by state suggests that Toyota’s supplier network might actually be more dispersed than originally estimated. Specifically, based on information presented in this article, the share of jobs in Michigan and Indiana is lower than estimated, while Illinois, Ohio, and Tennessee report a relatively higher share.

REFERENCES


Database Publishing Co., 1997, Arizona and California Manufacturers Register, Phoenix, AZ.


Fournier, Stephen E., and Andrew M. Isserman, 1993, “Putting it all together: The effects of the Honda plant on its host county and rural hinterland,” West Virginia University, Regional Research Institute, research paper, No. 9305.


Henrickson, G. Rex, 1951, Trends in the Geographic Distribution of Suppliers of Some Basically Important Materials Used at the Buick Motor Division, Flint, Michigan, Ann Arbor, MI: University of Michigan, Institute for Human Adjustment.


McAlinden, Sean, and Brett Smith, 1993, “The changing structure of the U.S. automotive parts industry,” University of Michigan, Transportation Research Institute, research paper, No. UMTRI 93-6.


Ward’s Communications, 1950–98, Ward’s Automotive Yearbook, Overland Park, KS.

The new view of growth and business cycles

Jonas D. M. Fisher

Introduction and summary

Two central concerns of economic policy are growth and business cycle stabilization. There is considerable interest in devising government policies and institutions to influence prospects for economic growth and mitigate the distress associated with economic downturns. Proper evaluation of the benefits and costs of a given policy proposal requires knowledge of the determinants of growth and business cycles. This is one reason for the considerable body of research aimed at understanding these phenomena.

The last two decades have seen considerable advances in this research. Recent empirical evidence, however, brings into question two of its basic assumptions—first, that technological change is homogeneous in nature, in that it affects our ability to produce all goods symmetrically, including consumption and investment goods; and second, that business cycles are driven by shocks which affect the demand for investment goods.

In this article, I document the key evidence that challenges the conventional views of growth and business cycles. I then discuss the plausibility of alternative theories that have been advanced to meet the challenge. To date, the evidence seems to support a new view of growth and business cycles, one that is based on technical change biased toward new investment goods like capital equipment.

The key evidence involves two observations on the behavior of the relative price of business equipment over the last 40 years. First, in almost every year since the end of the 1950s, business equipment has become cheaper than the previous year in terms of its value in consumption goods. This means that if one had to trade restaurant meals for a piece of equipment that makes the same number and quality of, say, bicycles, one would forgo fewer meals in 1998 than in 1958. Second, this relative price tends to fall the most when the economy, and investment expenditures in particular, are growing at relatively high rates, that is, it is countercyclical.

The first piece of evidence is striking because it suggests that much of post-WWII economic growth can be attributed to technological change embodied in new capital equipment. This conflicts with conventional views on what drives economic growth. A piece of capital equipment is a good that is used to produce another good, such as a crane or a computer. An improvement in capital-embodied technology is the invention of equipment that takes the same amount of labor and preexisting equipment to produce as the old equipment but that produces more goods when combined with the same amount of labor as before. If a new production process yields the same units of capital equipment with less factor inputs, then this has the same economic implications as if the capital equipment produced were itself more efficient. Hence, an equivalent interpretation of what constitutes capital-embodied technical change is that it involves an improvement in the technology that produces capital equipment.

To understand the relationship between capital-embodied technical change and the trend in the equipment price, suppose the technology for producing consumption goods is fixed. With improvements in technology embodied in equipment, the supply of (quality-adjusted) investment goods increases relative to consumption goods, so the equipment price falls. Greenwood et al. (1997) build on this insight to show that a large fraction of economic growth can be attributed to capital-embodied technical change. This conflicts with the conventional view that most

---

Jonas D. M. Fisher is a senior economist at the Federal Reserve Bank of Chicago. This article has benefited from conversations with Larry Christiano. The author thanks Judy Yoo for excellent research assistance.
growth is due to disembodied technical change, or multifactor productivity. Improvements in disembodied technology, usually measured as the Solow (1957) residual, make it possible to produce all kinds of goods, not just capital goods, with less capital and labor. If this were the dominant source of growth, then we should not have seen such a large drop in the price of equipment over the last 40 years.

The second piece of evidence runs counter to standard views of the business cycle. Standard theories hold that the business cycle is driven by shocks which affect the demand for investment goods. For example, consider the IS–LM model, which summarizes much of what is often called Keynesian macroeconomics. This model is the focus of most textbooks on macroeconomics and underlies much of the discussion of macroeconomic policy in the media. In this model, business cycles are due to shocks to aggregate demand, such as monetary and fiscal disturbances. For example, expansionary monetary policy stimulates demand for investment goods through lower interest rates. If there is an upward sloping supply schedule for investment goods, we would expect the relative price of investment goods to rise. The same holds for expansionary fiscal policy, if government spending does not fully crowd out investment. Another view of business cycles, often attributed to Keynes, is that they are primarily investment cycles driven by variation in animal spirits, that is, changes in confidence about future growth prospects. With the same assumptions on investment supply, we would expect investment prices to be high when investment is high. In summary, traditional Keynesian views of business cycles imply that investment good prices should be procyclical, that is, be high when overall economic activity is relatively high.

In recent years, an alternative view of business cycles, based on “fundamentals” that influence aggregate supply, has gained credence. This real business cycle view says that business cycles are driven in large part by disturbances to multifactor productivity. Just as the shocks to aggregate demand which are central to Keynesian theories, these disturbances influence business cycles through their effect on the demand for investment goods. Hence, if there are costs in terms of forgone consumption of expanding investment good production, that is, if the supply schedule of capital is upward sloping, these models also predict the relative price of investment goods to be procyclical (Greenwood and Hercowitz, 1988).

Since the relative price evidence contradicts the major schools of business cycle thought, it poses a challenge to our understanding of business cycles. There are two leading hypotheses that could reconcile the theory and evidence. One, the embodied technology view, is built from the real business cycle tradition and takes into account the trend evidence on equipment prices. Falling equipment prices are compelling evidence of capital-embodied technological progress over long horizons. Perhaps changes in the rate of such technological progress occur over shorter horizons as well. Suppose the business cycle were driven, to a large extent, by these disturbances. An increase in the rate of capital-embodied technical change would lead to an outward shift in the supply schedule for investment goods. With stable investment demand, investment would rise and equipment prices would fall. This new view of business cycles, which complements the new view of growth suggested by the long-run evidence on investment prices, has been explored by Christiano and Fisher (1998), Fisher (1997), and Greenwood et al. (1998).

The other leading theory is more easily understood in the context of traditional Keynesian views of the business cycle. If shocks to aggregate demand occur with a downward sloping investment supply curve, then the price of equipment could fall in a boom. A downward sloping investment supply curve would arise if increasing returns to scale played an important part in the production of capital equipment, so this is called the increasing returns view. This view has been advanced by Murphy, Shleifer, and Vishny (1989).

Below, I document the trend and business cycle evidence on equipment prices. There is no reason to expect that capital-embodied technological change is unique to equipment. Equipment is one of many investment good aggregates, that is, types of capital. Moreover, for simplicity most economic models assume only one or two types of capital. Therefore, in addition to equipment prices, I analyze other investment good aggregates. Next, I discuss research that sheds light on the plausibility of the alternative views, including some new evidence. To date, the evidence seems to support the new view of growth and business cycles based on capital-embodied technical change.

If growth and business cycles are originating from changes in capital-embodied technology, then the models we use for policy analysis have to incorporate this and, consequently, policy recommendations could change. For example, to the extent that technological change is embodied in capital equipment, government policies that affect equipment investment could have a dual impact on growth via the quality and quantity of capital goods. This could mean, for example, that investment tax credits directed toward improvements in the efficiency of capital equipment could have a significant impact on growth.
The implications for stabilization policy of the embodied technology view are less obvious. The fact that it seems to supplant the increasing returns view means that the arguments for interventionist stabilization policy that this view lends support to are less compelling. For example, increasing returns could provide scope for policy intervention, as it either involves externalities or is inconsistent with perfect competition. Moreover, it makes models based on animal spirits more plausible, which also has implications for stabilization policy (see Christiano and Harrison, 1999). The embodied technology view is more in line with the real business cycle tradition, in which policy interventions are counterproductive.

Evidence on investment good prices

To study the trend and business cycle properties of investment good prices, we need two things—a way to extract real prices and quantities from data on nominal investment expenditures; and a precise definition of what we mean by the business cycle component of the data. Below, I address these issues. Then, I introduce the data and present the results characterizing the trend and cycle behavior of investment good prices.

Measuring prices and quantities

This section describes how relative prices and real quantities of investment goods are measured. My measures of prices and quantities are based on measures published in the “National income and product accounts” (NIPA) of the U.S. Bureau of Economic Analysis (BEA).

The basis of the BEA procedure is to construct a price deflator. To be concrete, a given nominal quantity of expenditures on some good \( i \), \( X_i \), is decomposed into a price deflator, \( P_i \), (which measures the nominal price of the good) multiplied by a quality-adjusted index of the real quantity of the good, \( q_i \).

The BEA measures \( P_i \) and \( q_i \) for different goods using a so-called chain-weighting procedure, which is summarized in box 1. My measure of quantity is simply \( q_i \), measured in units of 1992 dollars. My measure of the real price, alternatively the relative price, of good \( i \) at date \( t \), \( p_i^t \), is the real quantity of consumption goods that would need to be sold in order to purchase one unit of good \( i \) at time \( t \). It is defined as the price deflator for good \( i \) divided by the price deflator for consumption of nondurables and services. The rationale for this measure is described in box 1.

Measuring the business cycle component of the data

In the introduction I described how the price of producer durable equipment (PDE) varies over the business cycle. Below, I provide a brief description of how I measure the business cycle component of the data. A detailed discussion of the procedure is given in Christiano and Fitzgerald (1998).

Figure 1 illustrates the basic idea behind the procedure. The colored line in panel A of figure 1 displays real 1992 dollar chain-weighted gross domestic product (GDP). The reported data are the logarithm of the raw data. The advantage of using the logarithm is that the resulting movements correspond to percent changes in the underlying data. The deviations between the data and the trend line (graphed in panel B) contain the rapidly varying, erratic component, inherited from the choppy portion of the data that is evident in panel A. The colored line in panel B is my measure of the business cycle component of real GDP. This measure excludes both the trend part of the data and the rapidly varying, erratic component. It includes only the component of the data that contains fluctuations in the range of two to eight years. According to this approach, the economy is in recession when the business cycle measure is negative and in prosperity when it is positive.

Figure 1 also compares this measure of the business cycle with the one produced by the National Bureau of Economic Research (NBER). This organization decides, based on an informal examination of many data series by a panel of experts, when the economy has reached a business cycle peak or trough. The start of each shaded area indicates the date when, according to the NBER, the economy reached a business cycle peak. The end of each shaded area indicates a business cycle trough. Note how real GDP falls from peak to trough and then generally grows from trough to peak. An obvious difference in the two business cycle measures is that the measure used in this article is a continuous variable, while the NBER’s takes the form of peak and trough dates. As a result, my measure not only indicates when a recession occurs, but also the intensity of the recession. Apart from these differences, the two measures appear reasonably consistent. For example, near the trough of every NBER recession, my measure of the business cycle is always negative. However, the two measures do not always agree. According to my measure, the economy was in recession in 1967 and 1987, while the NBER did not declare a recession then. In part, this is because there must be several quarters of negative GDP growth before the NBER declares a recession. The procedure I use only requires a temporary slowdown.

The data

I consider a broad variety of investment goods, as outlined in table 1. The broadest measure of investment
The U.S. Bureau of Economic Analysis (BEA) uses the chain-type Fisher index to measure real output and prices. For a thorough discussion of the procedures the BEA uses, see Landefeld and Parker (1997), which this box draws on. This index, developed by Irving Fisher, is a geometric mean of the conventional fixed-weighted Laspeyres index (which uses weights of the first period in a two-period example) and a Paasche index (which uses the weights of the second period). The Laspeyres price index for period \( t \) constructed using base year \( t-1 \), \( L_t \), is given by

\[
L_t = \frac{\sum_{i=1}^{N} P_{t-1}^i \times q_{t-1}}{\sum_{i=1}^{N} P_{t-1}^i \times q_i^t}
\]

The Paasche price index for period \( t \) constructed using base year \( t \), \( S_t \), is given by

\[
S_t = \frac{\sum_{i=1}^{N} P_t^i \times q_t}{\sum_{i=1}^{N} P_t^i \times q_i^t}
\]

Here \( N \) is the number of goods whose prices are being summarized by the index, \( P_t^i \) is the date \( t \) dollar price of the \( i \)th quality-adjusted good, and \( q_t^i \) is the quality-adjusted quantity of good \( i \) purchased at date \( t \). The Fisher price index at date \( t \), \( F_t \), is

\[
F_t = \sqrt{L_t \times S_t}
\]

From this definition we see that changes in \( F_t \) are calculated using the “weights” of adjacent years. These period to period changes are “chained” (multiplied) together to form a time series that allows for the effects of changes in relative prices and in the composition of output over time. Notice that a quantity index can be computed in a manner analogous to the price index. A nice feature of the Fisher index is that the product of these two indexes equals nominal expenditures. Landefeld and Parker (1997) discuss several advantages of this index over previously used fixed weight indexes.

To measure relative prices we need to choose a numeraire. In the introduction the term “value in consumption goods” was used. Implicit in this statement is the assumption that consumption goods, specifically nondurable and services consumption, is the numeraire. Define the price deflator for nondurable and services consumption as \( P_{t}^{n} \). Then the relative price of the good \( i \) at time \( t \), \( p_{t}^{i} \), is defined as

\[
p_{t}^{i} = \frac{P_{t}^{i}}{P_{t}^{n}} = \frac{\frac{\text{time } t \text{ dollars / good } i}{\text{time } t \text{ dollars / consumption good}}}{\frac{\text{consumption goods}}{\text{good } i}}
\]

Notice that the units of the price are what we require. The BEA does not provide a measure of price deflator for nondurable and services consumption. To construct the consumption deflator used in this article, I applied the chain-weighting methodology outlined above, treating the NIPA quantity and price indexes for nondurable consumption and service consumption as the prices and quantities in the formulas.

is total private investment (TPI). This measure includes all private expenditures on capital goods and consumer goods designed to last more than three years. This is a broader measure of investment than the conventional NIPA measure of investment, private fixed investment (PFI), which excludes expenditures on consumer goods. Within TPI, I define two main components, nonresidential and residential. Nonresidential has two main subcomponents, structures (NRS, for example, factory buildings and office buildings) and producer durable equipment (PDE, for example, auto-assembly robots and personal computers). Similarly, residential is broken down into residential structures and equipment (RSE, for example, single family homes and refrigerators) and consumer durables (CD, for example, televisions and vacuum cleaners). These four major subcomponents of TPI are then broken down further. The “Nominal share” and “Real share” data provide information on the relative magnitudes of expenditures on the different measures of investment, as well as a preliminary indication of interesting trends in relative prices. The nominal and real shares for TPI are calculated as the ratio of nominal and real TPI relative to nominal and real GDP, respectively. For example, in 1958 nominal TPI expenditures were 22 percent of nominal GDP and real TPI expenditures were 16 percent of real GDP. The remaining shares are calculated using TPI as the base for the share calculations. For example, PDE expenditures accounted for 24 percent of nominal TPI and 20 percent of real TPI in 1958.
real and nominal shares for many of the components of investment listed in table 1, suggesting that trends in relative prices are exhibited by many of the subcomponents of TPI. Second, the difference between the real shares of TPI and PFI (the former is a fraction of GDP, while the latter is a fraction of the former) is seen to be due to the increasing quantities of consumer durables being purchased. Third, the much talked about “information age” manifests itself here as the huge increase in the fraction of TPI that has been due to expenditures on information and related equipment since 1960. In 1960 this type of investment accounted for less than 1 percent of real TPI. By 1995, its share had grown to 13 percent. Finally, note that both residential and nonresidential structures account for less of TPI in 1998 than in 1958.

**Trends in investment good prices**

In this section, I explain two main findings relating to the long-run behavior of relative prices for the various components of investment listed in table 1. First, the relative price of TPI has fallen consistently since the mid-1950s. Second, there is considerable heterogeneity in the long-run behavior of the prices of the subcomponents of TPI. Generally, the behavior of the price of TPI is dominated by dramatic drops in the prices of PDE and CD, which are also evident in the prices of most of the main subcomponents of these investment aggregates. The prices of RSE and NRS and their subcomponents, while exhibiting trends over subsamples of the period studied, have not fallen as consistently and their changes over time are much smaller than those of PDE and CD.

Figure 2 displays the relative price trend evidence. The black lines in figure 2 are measures of the (natural logarithm of the) relative price of each of the investment components listed in table 1 over the period for which data are available. The colored lines are the trends calculated in the same way as the trend of real GDP displayed in figure 1. The first column of panels in figure 2 displays prices and trend lines for the main aggregates. The remaining columns display prices and trends for the four broad categories of TPI and their main subcomponents.

Figure 2 shows that the relative prices of different components of investment have behaved quite
differently in the postwar era. The price of the broadest investment measure, TPI, has been falling consistently since the early 1950s. Since the plot of the relative price of TPI is in natural logarithms, one can take the difference between the prices for two years to calculate the percentage change. This procedure indicates that the price of TPI in terms of consumption goods fell about 42 percent between 1958 and 1998.

Studying the other plots in figure 2, we see that this large drop in the price of TPI can be attributed to strong downward trends in PDE (particularly information and related and transportation equipment) and CD (all three types). The drop in the relative price of information equipment is particularly dramatic, at almost 200 percent since 1961. The prices of NRS and its components were generally rising until the late 1970s, were falling for most of the rest of the sample period, and have started to rise again in the 1990s. RSE and its components display a similar pattern. Generally, the long-run changes in structures prices have been much smaller than in PDE and CD prices. When the investment components are aggregated into nonresidential and residential, the strong downward trends in PDE and CD prices dominate the changing trends in structures.

**Prices of investment goods over the business cycle**

My objective here is to determine the extent to which investment good prices are generally procyclical, countercyclical, or acyclical (do not display any distinctive pattern over the business cycle). I find that, generally speaking, prices of PDE, NRS, and their components are countercyclical, prices of RSE and its components are procyclical, and prices of CD and its components are acyclical. There is some sample period sensitivity, as outlined below.

In table 1, the column headed $\sigma_q / \sigma_\rho$, indicates the relative volatility of the different investment components over the business cycle. This is the standard deviation of the business cycle component of the indicated real quantity series divided by the standard deviation of the business cycle component of real GDP. We see that TPI varies almost three times as much as GDP. The most volatile components of

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Measures of investment used in the analysis</th>
<th>Nominal share</th>
<th>Real share</th>
<th>Business cycle volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total private investment</td>
<td>0.2184</td>
<td>0.2678</td>
<td>0.2378</td>
</tr>
<tr>
<td>Nonresidential</td>
<td>0.4165</td>
<td>0.4496</td>
<td>0.4656</td>
</tr>
<tr>
<td>Structures</td>
<td>0.1730</td>
<td>0.1510</td>
<td>0.1226</td>
</tr>
<tr>
<td>Nonresidential buildings</td>
<td>0.0975</td>
<td>0.0821</td>
<td>0.0906</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.0418</td>
<td>0.0399</td>
<td>0.0238</td>
</tr>
<tr>
<td>Mining exploration, shafts, &amp; wells</td>
<td>0.0233</td>
<td>0.0255</td>
<td>0.0115</td>
</tr>
<tr>
<td>Producer durable equipment</td>
<td>0.2438</td>
<td>0.2985</td>
<td>0.3430</td>
</tr>
<tr>
<td>Information &amp; related</td>
<td>0.0358</td>
<td>0.0690</td>
<td>0.1153</td>
</tr>
<tr>
<td>Industrial</td>
<td>0.0799</td>
<td>0.0783</td>
<td>0.0734</td>
</tr>
<tr>
<td>Transportation &amp; related</td>
<td>0.0599</td>
<td>0.0782</td>
<td>0.0871</td>
</tr>
<tr>
<td>Residential</td>
<td>0.5850</td>
<td>0.5504</td>
<td>0.5344</td>
</tr>
<tr>
<td>Residential structures &amp; equipment</td>
<td>0.2188</td>
<td>0.2175</td>
<td>0.1782</td>
</tr>
<tr>
<td>Single family structures</td>
<td>0.1289</td>
<td>0.1203</td>
<td>0.0897</td>
</tr>
<tr>
<td>Multifamily structures</td>
<td>0.0228</td>
<td>0.0212</td>
<td>0.0122</td>
</tr>
<tr>
<td>Other structures</td>
<td>0.0627</td>
<td>0.0715</td>
<td>0.0721</td>
</tr>
<tr>
<td>Consumer durables</td>
<td>0.3643</td>
<td>0.3329</td>
<td>0.3562</td>
</tr>
<tr>
<td>Motor vehicles &amp; parts</td>
<td>0.1453</td>
<td>0.1539</td>
<td>0.1414</td>
</tr>
<tr>
<td>Furniture &amp; household equipment</td>
<td>0.1659</td>
<td>0.1223</td>
<td>0.1443</td>
</tr>
<tr>
<td>Other</td>
<td>0.0534</td>
<td>0.0567</td>
<td>0.0705</td>
</tr>
<tr>
<td>Private fixed investment</td>
<td>0.6357</td>
<td>0.6671</td>
<td>0.6438</td>
</tr>
</tbody>
</table>

Notes: See box 1 for a description of the notation. For total private investment and gross domestic product, $Y$, nominal shares in the first row are $P^0q^0/\sum(P^0q^0)$. Nominal and real shares for investment good $i$ in the other rows are given by $P^i q^i / \sum(P^i q^i)$. Real shares for total private investment and gross domestic product are $q^i / \sum q^i$. Real shares for investment good $i$ in the other rows are given by $q^i / \sum q^i$.


---

*1959 data.
*1995 data.
*1960 data.
Notes: Relative price (black line) is a measure of the (natural logarithm of the) relative price of each of the investment components listed in table 1 over the period for which data are available. The trend (colored line) is calculated in the same way as the trend of real GDP displayed in figure 1. Panels A through D show prices and trends for the main aggregates. Panels E through T show prices and trends for the four broad categories of total private investment, along with their main subcomponents.
Source: See figure 1.
investment are single family structures, multifamily structures, and consumer expenditures on motor vehicles and parts. The least volatile components are NRS, furniture and household equipment, and the “other” component of CD. The column headed $\sigma_{p}/\sigma_{y}$ indicates the relative volatility of the prices of different investment components over the business cycle. This is the standard deviation of the business cycle component of the indicated relative price series divided by the standard deviation of the business cycle component of real GDP. The prices are much less volatile than the quantities. With one exception (mining exploration, shafts, and wells), all the prices are less volatile than real GDP over the business cycle.

As a preliminary look at the cyclicality of investment good prices, figure 3 displays the business cycle components of the prices (colored lines) and quantities (black lines) of seven of the broadest measures listed in table 1, along with the business cycle component of the deflator for consumption of nondurables and services. The latter price is used in the denominator of all the investment relative prices, so its business cycle dynamics will influence all the relative price measures discussed here.\(^{11}\)

Notice first that the consumption deflator rises in all but one recession, 1981 Q3–82 Q4 (see shaded areas in figure 3). This is a force for procyclicality of investment good prices. For example, if the price deflator for an investment good were constant, then the real price of that good would be procyclical. As expected, the quantities are generally procyclical, although the peaks and troughs do not exactly coincide with the NBER dates. The prices do not display as consistent a pattern as the quantities. For example, sometimes the price of TPI moves with the quantity of TPI (1950s, 1960s, and 1990s) and sometimes it moves in the opposite direction (1970s and 1980s). More distinct patterns emerge when TPI is decomposed into nonresidential and residential. In the 1950s and 1990s, the prices and quantities of nonresidential appear to move closely together. In the 1960s, 1970s, and 1980s, prices and quantities of this investment measure generally move in opposite directions. Prices and quantities of residential show more evidence of moving together. The most striking pattern to emerge among the subcomponents of nonresidential and residential is in PDE. With the exception of the 1950s, almost every time the quantity of PDE moves up, the price of PDE moves down. This suggests countercyclical behavior in the real price of PDE.

For a more formal examination of how the prices of investment goods vary with the business cycle, I use a cross-correlogram. A cross-correlogram is a diagrammatic device for describing how two variables are related dynamically. For example, it provides a measure of whether, say, movements in one variable tend to occur at the same time and in the same direction as movements in another variable. It can also be used to measure whether, for example, positive movements in a variable tend to occur several quarters ahead of positive movements in another variable.

The basis for the cross-correlogram is the correlation coefficient, or correlation. A correlation is a measure of the degree to which two variables move together and always takes on values between −1 and 1. If a correlation is positive, then the two variables are said to be positively correlated. Similarly, if a correlation is negative, the variables are said to be negatively correlated. Larger absolute values in a correlation indicate a stronger pattern of moving together. A correlation for two variables measured contemporaneously is a measure of how much two variables move together at the same time. A correlation can be computed for two variables measured at different times. For example, we can measure the correlation between variable $x$ at time $t$ and variable $y$ at time $t−k$, where $k$ is a positive integer. This would measure the degree to which variations in $y$ occur before movements in $x$. A cross-correlogram plots these correlations for various values of $k$.

Figure 4 displays cross-correlograms (along with a two-standard-deviation confidence interval, a measure of how precisely the correlations are estimated) for various business cycle components of real investment and GDP, $-6 \leq k \leq 6$. For example, panel A of figure 4 displays the correlations of real nonresidential investment at date $t$ and real GDP at date $t−k$ for the various values of $k$. The fact that the correlation for $k=0$ is positive and close to 1 for all the plots in figure 4 shows that all the components of investment displayed are strongly positively correlated with GDP contemporaneously. This confirms the impression given by figure 3 that real expenditures on these investment goods are strongly procyclical. Notice that the largest correlations for nonresidential and its two main subcomponents, NRS and PDE, are for $k>0$. This says that these components of investment tend to lag GDP over the business cycle. Another way of saying this is that movements above trend in GDP tend to occur before movements above trend in these measures of investment. On the other hand, the largest correlations for residential and its main subcomponents, RSE and CD, are all for $k<0$. This says that these components of investment lead output over the business cycle. Because the correlations in figure 4 are mostly positive, this figure shows that the main components of investment are generally procyclical. (If they had been mostly
FIGURE 3
Business cycle components of investment good prices and quantities

A. Consumption deflator
logarithm

B. Nonresidential
logarithm

C. Nonresidential structures
logarithm

D. Durable equipment
logarithm

E. Total private investment
logarithm

F. Residential
logarithm

G. Residential structures
logarithm

H. Consumer durables
logarithm

Notes: Each business cycle component has been scaled by its standard deviation, and all data are quarterly. The colored lines represent the business cycle component of the price series for the indicated variable and the black lines represent the business cycle component of the quantity series for the indicated variable. Shaded areas indicate recessions as determined by the National Bureau of Economic Research. Source: See figure 1.
negative, then this would have been evidence of countercyclical. If the correlations were mostly close to zero, this would have been evidence of acyclical.

Figure 5 displays cross-correlograms (with standard errors) for the prices of the broadest measures of investment and real GDP. The plots in figure 3 indicate that there may be some sample period sensitivity in the estimation of the underlying correlations, so figure 5 displays cross-correlograms based on two sample periods. The first column of panels in figure 5 is based on the sample period 1947:Q1–98:Q3 and the second column is based on 1959:Q1–98:Q3. Notice that none of the correlations for the TPI price based on the longer sample are significantly different from zero. This means that the price of the broadest measure of investment is essentially acyclical. There is some evidence of countercyclical movements in this price for the shorter sample, although the correlations in this case are generally not very large in absolute value or statistically significant.

The cyclical behavior of prices for the narrower investment aggregates displayed in figure 5 reveals
that the lack of any distinct cyclical pattern for the price of TPI masks interesting differences between the prices of nonresidential and residential goods. Over the longer sample, the nonresidential price is estimated to be essentially acyclical, but the residential price is clearly procyclical. Over the shorter sample the nonresidential investment price is clearly countercyclical and the residential price remains procyclical. The difference in the estimated cross-correlogram for nonresidential over the two sample periods turns out to be due to differences in the behavior of the price of PDE in the 1950s compared with the later sample period (see figure 3).

The evidence in figure 5 suggests two things. First, the cyclical behavior of investment good prices depends to some extent on the sample period examined. Second, considering a broad investment aggregate masks potentially interesting cyclical characteristics of more narrowly defined investment good prices. Figures 6 and 7 try to uncover whether the cyclical behavior of nonresidential and residential prices also masks different cyclical behavior among the subcomponents.

**FIGURE 5**

Business cycle correlations between investment prices \( (t) \) and output \( (t-k) \)

Sample period: 1947–98

Sample period: 1959–98

**Note:** Black lines are point estimates of correlations for the indicated series; colored lines are a two-standard-error confidence band.

**Source:** See figure 1.
of these broad investment aggregates. These figures display price-output cross-correlograms for the main subcomponents of nonresidential and residential. Due to data availability, the sample period for estimating the correlations is 1959:Q1–98:Q3.

The first column in figure 6 pertains to NRS and its main subcomponents, nonresidential buildings, utilities, and mining. The price of NRS is significantly countercyclical. This appears to be mainly driven by the price of utilities and mining. The second column of figure 6 pertains to PDE and its main subcomponents, information and related equipment, industrial equipment, and transportation equipment. There are two observations to make here. First, the price of PDE is strongly and significantly countercyclical. The contemporaneous \( k = 0 \) correlation is \(-0.63\) with a standard error of 0.03. The largest correlation in absolute value is for \( k = 2 \), indicating that this price lags output by about two quarters, about the same as the quantity of PDE (see figure 4). The second observation is that the prices of the main components of PDE behave almost identically: They are strongly and significantly negatively correlated with output and lag output by about two quarters. The behavior of the industrial equipment price is particularly striking, given that the long-run behavior of this price is so different from that of the other two subcomponents of PDE (see figure 2).

Figure 7 is constructed similarly to figure 6, with RSE and its subcomponents in the first column and CD and its subcomponents in the second column. This figure shows that prices of RSE are generally procyclical and prices of CD goods are mostly acyclical. The behavior of RSE is driven mostly by the cyclicality of single and multifamily structures. Interestingly, despite the fact that investment in RSE tends to lead output over the business cycle, the real price of RSE and its components lags output. The real price of CD is driven mostly by motor vehicles and other. Of the subcomponents of CD, only the furniture price displays significant countercyclical.

**Summary of the evidence**

The key features of the evidence presented in this section can be summarized as follows. First, there is strong evidence of a downward trend in the price of investment goods in terms of consumption goods. This downward trend is concentrated among components of PDE and CD. Second, the broadest category of investment, TPI, displays little distinct cyclical variation over the sample period 1947:Q1–98:Q3, but is moderately countercyclical in the later period, 1959:Q1–98:Q3. If we are willing to abstract from the 1950s, say because of the dominating influence of the Korean war, then it seems reasonable to say that the price of the broadest component of investment is weakly countercyclical. Certainly it is difficult to make the case that this price is procyclical, regardless of the sample period considered.

Many components of TPI display distinct cyclical characteristics, even if we include the 1950s. The prices of the two main components, nonresidential and residential, behave differently. The former is significantly countercyclical and the latter is significantly procyclical. The behavior of the nonresidential price is dominated by the PDE price. The PDE price is strongly countercyclical, as are the prices of all its subcomponents. The price of NRS is mildly countercyclical, but this pattern is not shared by all its subcomponents. The behavior of the residential price is dominated by RSE prices, which are strongly procyclical. CD prices are acyclical or weakly countercyclical.

**Implications for growth and the business cycle**

How does the trend and cycle behavior of investment goods prices presented above challenge conventional views about growth and business cycles? Next, I discuss various attempts to reconcile theory with the evidence and some empirical work that sheds light on the plausibility of competing theories.

**Growth theory**

Recent years have seen an explosion of theoretical and empirical research into economic growth. On the theoretical side, two leading classes of models of the determinants of economic growth have emerged. The first is based on the accumulation of *human capital* and follows from the work of Lucas (1988). Human capital consists of the abilities, skills, and knowledge of particular workers. The basic idea behind this view of economic growth is that it is fundamentally based on improvements in the stock of human capital of workers over time. This view of growth holds that, other things being equal, the larger is the stock of human capital of workers, the more productive they are. This means that one expects an improvement in the stock of human capital to increase the amount of output of any good that can be produced for a fixed quantity of workers and capital. In this sense, growth due to the accumulation of human capital has a homogeneous impact on the economy’s ability to produce goods.

The second leading class of models focuses on research and development. Pioneering work along these lines includes Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992). One of the key insights of this literature is that growth can emerge if there are nondecreasing returns to produced
FIGURE 6

Business cycle correlations between nonresidential prices ($t$) and output ($t-k$)

A. Nonresidential structures
B. Nonresidential buildings
C. Utilities
D. Mining
E. Producer durable equipment
F. Information
G. Industrial
H. Transportation

Note: Black lines are point estimates of correlations for the indicated series; colored lines are a two-standard-error confidence band.

Source: See Figure 1.
FIGURE 7

Business cycle correlations between residential prices (t) and output (t–k)

A. Residential structures

E. Consumer durables

B. Single family structures

F. Motor vehicles

C. Multifamily structures

G. Furniture

D. Other structures

H. Other

Note: Black lines are point estimates of correlations for the indicated series; colored lines are a two-standard-error confidence band.
Source: See figure 1.
factors of production (such as knowledge or capital, but not labor). The bottom line of this theory is similar to that of the human capital models. Improvements in technology due to research and development usually increase the productivity of all factors of production. Consequently, if there is such an improvement in technology, more of all goods can be produced with a fixed quantity of capital and labor. Again, technological change is assumed to have a homogeneous impact on produced goods.

The evidence on trends in investment good prices, particularly the trend in the price of PDE, challenges these views of growth, because it strongly suggests that there have been substantial improvements in technology that have affected one kind of good but not another. Specifically, the data suggest that the quality and technology of capital goods production have advanced almost nonstop since the end of World War II. Why do the data suggest this? Assuming that the prices and quantities of PDE are correctly measured, the real price of PDE measures how many (constant quality) consumption goods need to be sold in order to raise the funds to purchase one (constant quality) unit of PDE. If this price has been falling, then fewer and fewer consumption goods are needed to buy a unit of PDE. This suggests that the supply of PDE has grown relative to the supply of consumption goods. One way the supply of PDE can rise in this way is if the technology for producing capital goods improves at a faster rate than that for producing consumption goods. In this case, the same amount of capital and labor applied to producing PDE or consumption goods will yield more PDE than consumption as time passes. That is, the supply of PDE will grow relative to consumption goods. The basic logic of supply and demand then dictates that the price of PDE in terms of consumption goods must fall. Greenwood et al. (1997) build on this intuition to show how the trend in the relative price of PDE and the associated increase in the share of PDE in aggregate output (see table 1) can be accounted for in a growth model in which most growth is due to capital-embodied technical change. In addition, the authors argue that other potential explanations for the price and quantity trends are implausible or boil down to essentially the same explanation.

Greenwood et al. (1997) apply their model of growth to reevaluate conventional estimates of the importance of technological change in improving standards of living. This line of research is called growth accounting. The effects of technical change using standard models, like the ones briefly described above, can be summarized by multifactor productivity, which is also called the Solow residual. Multifactor productivity is an index of the quantity of aggregate output that can be produced using a fixed quantity of (quality-adjusted) capital and labor. The higher the multifactor productivity, the more output can be produced. Traditionally, most of growth is viewed as being due to improvements in multifactor productivity. Greenwood et al. (1997) use their model to show that approximately 60 percent of all improvements in productivity can be attributed to capital-embodied technical change, while the multifactor productivity index accounts for the rest. This says that capital-embodied technical change is a fundamental part of growth.

**Business cycle theory**

To assess the cyclical evidence on relative prices, we need to understand how various shocks to the economy might influence the cost of investment goods compared with consumption goods. Figure 8 displays a production possibilities frontier (PPF) for consumption and investment goods. The PPF depicts the various quantities of consumption and investment goods that can be produced if capital and labor are fully employed and used efficiently. The shape of the frontier reflects the fact that, holding fixed the quantity of labor and capital employed in producing goods, it is costly to shift production toward either producing more consumption goods or more investment goods. This is reflected in the figure by the increase in the (absolute value of the) slope of the frontier as one moves from the upper left to the lower right. In a competitive equilibrium, the slope of the frontier equals the relative price of the goods. Hence, as more investment goods are produced, the relative price of investment goods rises.

The PPF summarizes the supply side of the economy. The actual price in a competitive equilibrium is determined by the interaction of the demand for consumption and investment goods with the supply. Suppose that the demand for consumption and investment goods dictates that the quantity of consumption goods and investment goods actually produced is given by $C_t$ and $I_t$ in figure 8. Now, suppose a Keynesian demand shock—for example, an increase in the money supply which lowers interest rates—increases the demand for investment goods relative to consumption goods. Since this is a demand shock, the PPF in figure 8 does not change. The change in demand leads to a movement down the frontier, say to a point where consumption and investment are given by $C_t$ and $I_t$. Since the slope of the frontier is steeper at this point, the relative price of investment goods must rise. If aggregate output is driven by shocks to investment
demand, then the price of investment goods is predicted to be procyclical.

An aggregate supply shock has a similar implication. The conventional assumption about these kinds of shocks is that they raise multifactor productivity and influence all produced goods symmetrically. This is shown in figure 9 as a proportional shift out in the solid line PPF to the dashed line PPF. The dashed line PPF has been drawn so that its slope is identical to the slope of the solid line PPF along a straight line from the origin. This means that if the ratio of consumption to investment goods produced before and after the technology shock is constant, then the relative price of investment goods will be unchanged. However, this is not what is predicted in standard models. These models say that when a good technology shock arrives, which raises the productivity of all factors of production, the optimal response of individuals is to smooth consumption. That is, not have consumption change too much in the short run. The result of this is that investment rises more than consumption. In figure 9, this is represented by consumption and investment changing from $C_o$ and $I_o$ before the productivity shock to $C_i$ and $I_i$, after the shock. It follows that the price of investment goods must rise in this case as well. Since output also rises with a positive technology shock, the price of investment goods is predicted to be procyclical.\footnote{In view of the cyclical evidence presented earlier, these model predictions are problematic. They are consistent with the behavior of residential investment, but inconsistent with the behavior of the other major components of investment and the broadest measure, TPI. Why are investment goods prices not procyclical? The two leading explanations involve assumptions about the technology for producing investment goods. One is based on increasing returns to scale in the production of investment goods (but not consumption goods). The other is based on a variation in the rate of capital-embodied technical change. The increasing returns view assumes that the more investment goods that are produced, the less costly it is to produce a unit of investment goods. One way to represent this is shown in figure 10, which displays a pseudo-PPF.\footnote{Notice that the shape is different from figures 8 and 9. Now when more investment goods are produced relative to consumption goods, the price of investment goods falls. In this case, both aggregate technology shocks and Keynesian demand shocks can lead to countercyclical relative prices.}

To understand the embodied technology view, consider an increase in the productivity of producing investment goods that has no direct impact on the production of consumption goods. This could take the form of improvements in the efficiency of producing investment goods. It could also take the form of an improvement in the quality of investment goods produced so that a given quantity of capital and labor can produce a higher quantity of quality-adjusted goods. Either way, we can represent the change in technology as in figure 11. The improvement in technology is shown by the shift from the solid to the dashed frontier. Along the dashed frontier, for each quantity of consumption goods produced, more investment goods can be produced. Moreover, along any straight line from the origin, the slope of the dashed frontier is flatter than the solid frontier. That is, for any fixed ratio of consumption to investment goods, the investment goods are cheaper in terms of consumption goods after the change in technology. Now, after the increase in technology, there will be a shift in favor
of the production of investment goods. If this shift is strong enough, the movement along the dashed frontier could in principle raise the investment good price. In practice, this does not happen. Since aggregate output rises after this kind of technology shock, if business cycles are in part driven by this kind of disturbance, then investment good prices could be countercyclical.

Evaluating the theories

Beyond the work of Greenwood et al. (1997), little has been done to evaluate the plausibility of the capital-embodied technological change theory of the trend evidence on investment prices. However, more work has been done to evaluate the differing views on the cyclicity of investment good prices.

Generally, the empirical evidence seems to go against the increasing returns interpretation of the cyclical evidence on prices. Harrison (1998) examines annual data on capital, labor, and value added in various industries in the consumption good sector and the investment good sector. She finds some empirical support for increasing returns associated with capital and labor in the production of investment goods. However, she does not find a sufficient degree of increasing returns to generate increasing returns in the factor of production, labor, that is variable in the short run. Consequently, the work does not support the increasing returns view. Other research on measuring increasing returns focuses on the manufacturing sector. Basu and Fernald (1997), Burnside (1996), and Burnside, Eichenbaum, and Rebelo (1995) have overturned previous empirical claims of increasing returns in the manufacturing sector, including capital equipment industries.

Other empirical work attempts to address a key implication of the increasing returns view—that the supply curve for investment goods slopes down. That is, holding other things constant, the cost of investment goods is diminishing in the quantity of investment goods produced. Shea (1993), in a study of many sectors of the economy, uses instrumental variables econometric techniques to distinguish supply shocks from demand shocks to trace out the slope of supply curves. The author’s main conclusion is that, broadly speaking, supply curves slope up. Goosbee (1998) focuses specifically on the supply of capital goods and uses a series of “natural experiments” (involving periodic changes in federal laws providing for investment tax credits) to identify a disturbance that affects the demand for investment goods but not the supply. He finds clear evidence of an upward sloping investment supply curve. To summarize, empirical work on the sign of the slope of the investment good supply schedule finds that it is positive.

Other research assesses the plausibility of the embodied technology view. Christiano and Fisher (1998) and Greenwood et al. (1998) evaluate business cycle models in which a major driving force for fluctuations is variations in capital-embodied technical change. They test the embodied technology view by examining the ability of their models to account for various business cycle phenomena. Both studies find that their models do about as well as other business cycle models in accounting for business cycle phenomena. As a measure of the importance of capital-embodied technical change as a driving force for business cycles, Greenwood et al. (1998) find that about 30 percent of business cycle variation in output can be attributed to this kind of shock. Christiano and Fisher (1998), in a very different model, find that about three-quarters of output fluctuations are due to this shock. Either way, the evidence suggests that variation in the rate
of technical change embodied in capital equipment accounts for a significant proportion of business cycle variation in output.

New evidence

Some new research attempts to distinguish the increasing returns view from the embodied technology view of the cyclical behavior of investment good prices. This evidence is based on two econometric procedures designed to identify disturbances to the aggregate economy that influence the demand for investment goods, but leave supply unchanged. The specific shocks considered are an exogenous increase in government purchases (that is an increase in government purchases that is unrelated to developments in the economy) and an exogenous monetary contraction.

In the government spending case, the idea is to investigate how particular investment quantities and prices respond to an exogenous increase in government purchases. The exogenous increase in government spending takes the form of a large military buildup (specifically the Korean war, the Vietnam war, and the Carter-Reagan buildup.) The methodology is identical to that employed by Eichenbaum and Fisher (1998). Figure 12 displays the estimates, which are based on quarterly data for 1947:Q1–98:Q3. The first row of figure 12 plots the response to an exogenous increase in government purchases of real investment in PDE and RSE (solid lines) along with a 68 percent confidence band (colored lines). The second row plots the corresponding relative price responses. Interestingly, PDE investment rises and RSE investment falls. Under the increasing returns view, we would expect the PDE price to fall and the RSE price to rise. The second row of plots indicates that the RSE price response is inconsistent with the increasing returns view, while the PDE price response seems to confirm it.

The monetary shocks case examines how quantities and prices of PDE and RSE respond to an estimate of a contractionary monetary disturbance. The methodology is standard and has been summarized by Christiano (1996) (see also Christiano, Eichenbaum, and Evans, 1999). The estimated responses (along with a 95 percent confidence interval) are presented in figure 13. Looking at the quantities in the first row of plots, notice that both PDE and RSE fall after an

**FIGURE 12**

Response of investment quantities and prices to exogenous increase in government purchases

<table>
<thead>
<tr>
<th>Quantity/Price</th>
<th>Time (Quarters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Equipment quantity</td>
<td>0-15 quarters</td>
</tr>
<tr>
<td>B. Residential quantity</td>
<td>0-15 quarters</td>
</tr>
<tr>
<td>C. Equipment price</td>
<td>0-15 quarters</td>
</tr>
<tr>
<td>D. Residential price</td>
<td>0-15 quarters</td>
</tr>
</tbody>
</table>

Notes: Black lines are point estimates of the responses; the colored lines are a symmetric 68 percent confidence band.
Source: Author’s calculations from data extracted from BIS Basic Economics database, 1959–98.
exogenous monetary contraction. Under the increasing returns view, one would expect the prices of both investment goods to rise. Studying the second row of plots, we see that the PDE price response is not significantly different from zero and the RSE price drops significantly.

Taken together, the evidence on the responses of RSE prices and quantities to government spending and monetary shocks goes against the increasing returns view. It conforms to a standard neoclassical view of investment, in the sense that it is consistent with the discussion of the production possibilities frontier in figure 8. Of course, the increasing returns view is really intended to apply to PDE investment. The responses of PDE prices and quantities provide mixed signals. The responses to a monetary shock provide evidence neither for nor against increasing returns, since the quantity falls but the price response is not very precisely estimated and could be either positive, negative, or zero. The responses to a government spending shock might be viewed as evidence in favor of increasing returns. However, one interpretation of the PDE price response in this case is that it is dominated by the Korean war military buildup. This occurred just after World War II, when military spending had fallen from very high levels. The increasing returns that could support a lower price with higher investment might conceivably be due to the resumption of large-scale production at facilities that had been operating far below minimum efficient scale. If this is true, it seems more like a special case than an enduring feature of the U.S. economy.

**Conclusion**

In this article, I have presented evidence on trends and business cycle variation in the prices of investment goods relative to nondurables and services consumption. This evidence seems to go against conventional views of both business cycles and growth. How can one reconcile theory with the evidence? The leading views include one based on increasing returns to scale in the production of investment goods and another based on capital-embodied technical change. While some of the evidence I presented could be viewed as supporting the increasing returns view, generally, there is little empirical support
for increasing returns. At this point, then, the leading candidate to reconcile theory with the data appears to be the one based on capital-embodied technical change, that is, the embodied technology view.

This conclusion has implications for our understanding of growth and business cycles, future research on these subjects, and policy. The prospect of a comprehensive theory of growth and business cycles is appealing because of its simplicity. Disembodied technical change has gained credence for its supposed ability to account for growth and business cycles. Yet, the theory of business cycles based on disembodied technology has always been problematic because the shocks are hard to interpret. The growth accounting results of Greenwood et al. (1997) bring into question the growth implications of this theory as well. In the search for a comprehensive theory of growth and business cycles, then, advances in capital-embodied technology seem to offer a promising alternative. In addition, they provide a much more tangible notion of growth. These considerations suggest that future research on growth and business cycles that emphasizes capital-embodied technical change may be fruitful.

If growth and business cycles are originating from changes in capital-embodied technology, then the models we use for policy analysis have to incorporate this and, consequently, policy recommendations could change. To the extent that technological change is embodied in capital equipment, government policies that affect equipment investment could have a dual impact on growth via the quality and the quantity of capital goods. This could mean, for example, that investment tax credits directed toward improvements in the efficiency of capital equipment could have a significant impact on growth. More research is required to uncover the full implications of this.

The implications for stabilization policy of the embodied technology view are less obvious. The fact that it seems to supplant the increasing returns view means that the arguments for interventionist stabilization policy that this view supports are less compelling. For example, increasing returns could provide scope for policy intervention, because it either involves externalities or is inconsistent with perfect competition. Moreover, it makes animal spirits models more plausible, which also has implications for stabilization policy (see, for example, Christiano and Harrison, 1999). The embodied technology view is more in line with the real business cycle tradition, in which policy interventions are counterproductive. Real business cycle theory says that the business cycle is largely the result of optimal behavior by individuals in the economy interacting, for the most part, in perfectly competitive markets. Any policy interventions in such an environment tend to reduce overall welfare. To the extent that the embodied technology view is more compelling than previous incarnations of real business cycle models, it lends greater support to the argument that interventionist stabilization policy cannot improve the well-being of any individual in the U.S. economy without hurting some other individual. Of course, this still leaves open the possibility that equity considerations might be used to defend interventionist stabilization policy.

NOTES
1 Equivalently, higher quality goods of all kinds can be produced with the same amount of capital and labor. As described in more detail below, new models of endogenous growth have reduced forms, which have similar implications for growth accounting to those of models written in terms of exogenous disembodied technical change.
2 Examples of textbooks that emphasize the IS–LM model are Abel and Bernanke (1997), Gordon (1998), Hall and Taylor (1997), and Mankiw (1997).
3 For a survey of theories based on animal spirits, see Farmer (1993).
4 A good summary of this view is Prescott (1986). For a discussion of how this view can be used to explain the 1990–91 recession, see Hansen and Prescott (1992).
5 This section relies heavily on Christiano and Fitzgerald (1998, pp. 58–59).
6 This is the empirical counterpart to investment as it is usually defined in the real business cycle literature.
7 The aggregation in this table is identical to the aggregation used by the BEA, except for “residential,” which is calculated as the chain-weighted aggregate of “residential structures and equipment” and “consumer durable.” See box 1 for the chain-weighting procedure.
8 For TPI and GDP, y, the nominal shares in the first row are \(P_1^{PI}\), and the real shares are \(q^{PI}/q^{PI}\). Nominal and real shares for investment good i in the other rows are given by \(P_1^{P_i}q^i/(P_1^{PI}q^{PI})\) and the real shares are \(q^i/q^{PI}\).
9 In the notation used above, the black lines are (the natural logarithm of) \(P_i^{PI}\) corresponding to the 20 types of investment listed in table 1 over the period for which data are available.
10 Many of the trends evident in figure 2 are not apparent in the NIPA fixed-weighted constant 1982 dollar and earlier NIPA data. In a very influential book, Gordon (1989) argued that the conventional BEA treatment of investment good quality severely underestimated the degree of quality change in investment goods. His analysis was the first to show that there is a substantial downward trend in the prices of PDE and CD. The BEA now incorporates many of the adjustments for quality change advocated by Gordon (1989).
11 The procedure used to extract the business cycle component of the relative price data involves the application of a linear filter. This, combined with the fact that this filter is applied to the natural logarithm of the relative prices, implies that the business cycle component of each relative price is the business cycle component of the relevant investment deflator minus the business cycle component of the consumption deflator.
REFERENCES


