Summary of the Papers

THE FOURTH CONFERENCE of the microeconomics meetings of the Brookings Panel was held in December of 1990 and papers were in two groups, consisting of three data-intensive empirical analyses and three papers that provide an assessment of the field of industrial organization. The first two papers were studies of individual industries. Richard E. Caves, Michael D. Whinston, and Mark A. Hurwitz looked at the pharmaceutical industry and the entry of generic drugs when patents expire, and Fred Mannering and Clifford Winston studied brand loyalty in the auto industry. Following this was the study by Steve J. Davis and John Haltiwanger of the increase in wage inequality in U.S. manufacturing. The three papers on aspects of industrial organization consisted of two review essays, one by Franklin M. Fisher and the other by Alvin K. Klevorick, of the recently published *Handbook on Industrial Organization* and an analysis of the antitrust guidelines by Robert Willig.

Caves, Whinston, and Hurwitz on Entry in the Pharmaceutical Industry

Both government and private industry have become increasingly concerned about the rising cost of health care, and one component of cost that is seen as a possible area for relief is the cost of ethical drugs. The pharmaceutical industry has devoted huge amounts of money to the development of a stream of new drugs, which are then sold at prices that are well above the marginal cost of producing them. The industry argues that this gap between price and production cost is necessary in order to recoup the costs of the research and development (R&D) needed to develop the drugs, while industry critics argue that the prices charged are excessive. One way in which government and private health insurers have attempted to reduce the cost of drugs has been to encourage the use of generics instead of brand-name pharmaceuticals. And in this paper, Richard E. Caves, Michael D. Whinston, and Mark A. Hurwitz explore the extent to which generic drugs enter the market and compete with established brand-name drugs. As well as its importance for health care costs, the authors argue that studying the entry of generic drugs is interesting because the drug industry provides a good case study of competitive behavior in response to entry or potential entry.

To explore these issues, Caves, Whinston, and Hurwitz use a panel of thirty drugs that lost patent protection during the period 1976–87. They have information on prices, quantities sold, and market shares for the brand-name drugs and the competing generics, and they also have information on the advertising expenditures of the brand-name suppliers.

The authors start with a review of the structure of the industry. They point out that the demand for drugs depends heavily on the behavior of the doctors that prescribe them, rather than on the preferences of the patients using them. Peter Temin has shown that doctors typically do not have well-organized information on the comparative effectiveness and riskiness of substitute drugs. Their decisions are based largely on the customary behavior of other doctors. In addition, the fact that insurance frequently provides reimbursement for the cost of the drugs means that neither the patient nor the doctor is very sensitive to the economic costs of alternative drugs. They are much more sensitive to the effectiveness of the overall treatment and will avoid generic substitutes if there is any question about the quality or therapeutic equivalence of the generics.

Another factor that limited the use of generics in the past was that some states traditionally had laws that prevented pharmacists from substituting generics for brand-name prescriptions. At this time, these laws have mostly been repealed. It is now up to the pharmacist and the patient to decide whether or not to use a generic. In fact, pharmacists may prefer to prescribe generics because they have larger markups for the pharmacists than brand names.

The end result of this combination of incentives and lack of incentives for patients and pharmacists to switch to generics is that in 1989 when prescriptions were written that permitted substitution for a brand name, only 29 percent ended up using the generic drug. This does, however, represent a major increase compared with the figure of 5 percent for 1980, an increase that has been driven in part by pressure from insurers, including government reimbursement schemes.

Prescriptions filled through pharmacies represent 82 percent of all prescriptions; the remainder involve drugs used in hospitals. Physicians are encouraged to use lists of approved drugs prepared by the hospital, and these lists generally suggest generic drugs when they are available. Hospitals pool information concerning drug effectiveness and costs.

On the supply side, the industry consists of many different firms, including many small firms. The industry as a whole is not terribly concentrated, although the number of firms producing any particular drug or close substitutes for it is small. Economies of scale do not seem to be significant for production costs. Much of the output of the industry is sold through drug wholesalers, with additional sales to hospitals, health maintenance organizations, and pharmacy chains.

The 1962 (Kefauver) amendments to the Pure Food Act had a major impact on the industry, making it harder for firms to get drugs approved, raising the development cost per drug, and reducing the period of effective patent protection by delaying the time of approval. The amendments, therefore, had some effects that raised costs and reduced returns for the R&D-intensive companies. At the same time, the amendments also reduced the ability of generics to compete once the patent expired because the generics were required to duplicate the testing procedures for the original drug. This requirement was eliminated in 1984 by the Waxman-Hatch Act, thereby allowing easier entry for generics once a patent had expired. However the act also increased the duration of patent protection for the innovating companies.

At this time, the primary barrier to entry faced by potential producers of generic drugs following patent expiry consists of the accumulated goodwill of the established brand-name and the concerns physicians and patients may have about the quality of the generics.

The drug companies spend substantial sums in promoting the drugs that they develop, using particularly an approach called detailing, which consists of visits to physicians by manufacturers' representatives to provide information on the new drugs and how they are to be used. In 1989 detailing represented 74 percent of total promotion costs. Observers have generally concluded that this sales promotion reflects a mixture of information and persuasion.

The combination of promotional costs, R&D costs, and profits makes up the bulk of the total cost of drugs. The direct production costs may be as low as 5 percent. In examining aggregate data on prices and input costs, Caves, Whinston, and Hurwitz note that drug prices have risen much more rapidly than labor costs since 1982. There seems to have been some change in pricing behavior in the industry.

The thirty drugs in the authors' study were obtained by consulting a variety of drug lists. All of them had clear patent expiration dates over the period considered. The drug prices that they used reflected wholesale prices, based on the average transactions prices, and the authors were able to distinguish sales to pharmacies and to hospitals. On average, the drugs in their sample had sales of \$67.3 million a year (1982 dollars) and made 87 percent of those sales through pharmacies. Sales promotion expenses are about 6 percent of sales by the year of patent expiry but are greater in earlier years. The drugs evaluated experienced rising sales over their period of patent protection, considered as a whole, but often had declining sales in the period approaching expiration. In their sample, seven had patents that expired before 1980, sixteen between 1980 and 1984, and seven after 1984, the year of the Waxman-Hatch Act. The authors found that the generics do not all enter on the date that the patent expires; rather they come in over time. There was a change in the pattern of entry of generics apparent in the data after the 1984 act.

The authors note one important omission in their analysis, namely, the extent to which there are substitute drugs available before the expiration of a patent. Some drugs represent the only effective treatment for a medical condition and have no good substitutes, while other patented drugs are only one of several similar drugs available. The authors were not able to construct a simple measure of the availability of substitutes for the drugs in their sample.

The empirical analysis that the authors use first involved making estimates of the prices of branded drugs sold to all purchasers together. The price of a drug is assumed to depend on the number of years since the drug was first introduced (a nonlinear specification was used to capture pattern of price change over the life cycle of the products) and the availability of generic substitutes. The estimates used instrumental variables to capture the fact that the number of generics that enter a market will depend on the price charged by the brand name. The results indicated a small (but statistically significant) effect of the entry of generics on the prices of brand names. Each generic reduces the price of the brand name by 0.8 percent, reflecting about a 2 percent decline for the average number of 2.5 generics. Excluding the drugs with no generic competitors entering the market, the average branded price decline was still only 4.5 percent.

The authors then extend their analysis to see whether the share of sales to hospitals had an effect on pricing. Concentrating on those drugs with less than 20 percent of sales going to hospitals, they find that a typical branded drug's price falls 2 percent following the first entrant, 15 percent with ten entrants, and 22 percent with twenty generic producers. Somewhat surprisingly, however, they found that when they split their sample into sales to hospitals and sales through pharmacies they did not obtain results that were very different for the two parts of the market. The price responsiveness for the hospital sales of those drugs with less than 20 percent of sales to hospitals did show greater price responsiveness than the total of all drugs, however.

The authors then turn their attention to the generics themselves and the extent to which their own prices are affected by the number of competing generics. They found that the generics sold at a very substantial discount— on average, the generic price is about 60 percent of the brand-name price when there is only one generic. And the entry of new generics actually depresses the prices of existing generics by more than the price of the brand-name product. The result is that on average generics sell for less than half of brand-name prices. These results suggest that the entry of generics has only a small effect on the market share of the established brand names, and the authors do indeed confirm this result. The brand names retain about three-quarters of the market even with five generic competitors. For all markets with generic competition, the generics capture only 36 percent of the total market on average.

Caves, Whinston, and Hurwitz then look at the pattern of advertising expenditures by the brand names around the time of patent expiration. They find that advertising declines as a result of impending and actual entry of generic competitors. They note that this suggests that the advertising is designed primarily to expand the total size of the market, rather than to persuade doctors to avoid the generic competitors. And finally, the authors look at the pattern of total sales in the markets where entry is occurring. There is a normal pattern of decline in sales for drugs as they get older, but the authors find that even after controlling for this normal life-cycle effect, there is a differential effect in markets where there is entry. Total sales are declining in the year of patent expiration and continue to decline subsequently. And there is little evidence that the entry of generics sustains the size of the overall market. The authors explain these findings as the result of two offsetting effects that occur with generic entry—lower overall price and reduced advertising.

The authors reflect on the implication of their analysis of this industry for market behavior. Their results suggest the great importance of the stock of goodwill built up by an established producer and that marketing effort plays an important role in sustaining the market for particular drugs. There was no evidence in their results of behavior by firms to deter the entry of the generics. Their results do show that generic competition makes a difference to average prices. In those markets where there are generics, they have a market share of about 40 percent, and the generics sell for half or less of the prices of the brand names. Generic entry reduces costs for consumers and reduces the returns to the firm with the brand-name drug.

The implications of their findings for public policy are somewhat surprising, however. Instead of seeing sales increase as a result of the generics, they see the opposite. This is because sales are not increased much by the lower prices and are reduced by the reduction in the advertising of the brand names. The overall effect of the generics on economic welfare, therefore, depends on whether the advertising was providing valuable information about the drug or was simply a way of persuading people to buy it.

In his discussion of the paper, Peter Temin suggested that the authors' model was fairly simple and that it would be helpful to impose a bit more structure on the data and analysis. In particular, he pointed to the importance of competing brand names, drugs that have their own patents but are close substitutes for the original drug. Ariel Pakes raised two concerns about the data analysis. First, the separate equations that explain drug store prices, hospital prices, and advertising may all be interrelated, he says, and probably should have been estimated together. And second, he argued that a separate analysis of whether or not generics enter a particular

market would have been helpful. It is possible that a generic that looked unprofitable in one period might look profitable in the next period.

Mannering and Winston on Brand Loyalty in the Auto Industry

The U.S. auto companies have experienced considerable difficulty in competing effectively against the Japanese producers. After concentrating for a number of years on building market share with exports from Japan, the Japanese companies are now engaged in building production capacity here in the United States. By the mid-1990s the transplants are expected to be able to supply nearly 25 percent of U.S. sales, while at the same time, the U.S. companies are closing plants. The nature of competition in this industry is gradually changing. Instead of U.S. plants competing against imports from Japan, increasingly the competition is now between U.S.-owned plants and Japanese-owned plants based in the United States.

Fred Mannering and Clifford Winston argue that an important element in the competitive struggle in this industry is a concept that they identify as "brand loyalty." The U.S. companies had built up substantial brand loyalty among their customers over many years, and this loyalty helped to sustain their share of the auto market in the face of the increase in foreign competition in the 1970s. Brand loyalty provides only temporary protection against competition, however. If the customers of the U.S. firms are dissatisfied with their purchases, they will eventually switch and they may then build up a new brand loyalty to the Japanese cars. And Mannering and Winston find that this is exactly what has been happening. In order for the domestic companies to regain their market shares or prevent further erosion of shares, they will have to rebuild the brand loyalty that they once held.

Mannering and Winston say that many firms consider high market share to be a key to high long-term profitability. Brand loyalty is related to developing, maintaining, and protecting market share. The authors' analysis of brand loyalty starts from the number of repeat purchases by a given customer of a particular make of automobile. But brand loyalty is not just repeat purchases. A consumer who is buying the same brand each time he purchases is not doing so only because of brand loyalty. He or she will be comparing current prices and qualities of the available brands and making a choice based partly on current information. Brand loyalty reflects the extent to which the probability of buying a particular brand of automobile is increased by the number of past repeat purchases of that brand, *holding other determinants of the purchase decision constant*. The other determinants include such things as price, weight, reliability, fuel efficiency, *and current brand preference*.

The authors describe a model of consumer choice that will be the basis of their empirical analysis. Those consumers that have decided to purchase a vehicle in a given year decide whether to buy a new or a used car and then select a particular make, model, and vintage. The two choices are then studied together, and so is the decision about the frequency of purchase. As their data base, Mannering and Winston have obtained information on 488 complete vehicle ownership histories for randomly selected households. In this data base there are households that have purchased only one car in their lives and households that have purchased as many as ten. There is considerable detail in the data set on the nature of the auto purchases and on various socioeconomic characteristics of the households (including residential location, income, and age).

In some preliminary analysis, the authors found that vehicle choice decisions looked somewhat different in the 1980s than they were in years prior to that, so they make separate estimates for the two periods. For the pre-1980 period there was insufficient data on purchases of Japanese brands to construct meaningful measures of brand loyalty for these brands by separate manufacturer. A single loyalty variable was therefore constructed for all the Japanese brands combined.

Looking at the results for the pre-1980 period, the authors find that, holding current characteristics constant, prior repeat purchases from the same manufacturer had a major impact on the probability of buying a new or used car from that same manufacturer—indicating, they say, the importance of brand loyalty. The results indicate that the Big Three U.S. manufacturers all maintained strong brand loyalty, with the largest effect for Chrysler, then GM, and finally Ford. As well as brand loyalty, there are several other significant determinants of auto choice. There are differences in brand preference irrespective of brand loyalty, and such factors as price and reliability affect decisions in the ways that are to be expected. The results for used cars are broadly similar to those for new cars.

By 1980 the U.S. auto market had changed. The Japanese manufacturers, Nissan, Honda, and Toyota, had gained sufficient market shares over enough years that Mannering and Winston were able to estimate separate brand loyalty for these companies. The results are given in table 3 of the paper, and they show that the Japanese manufacturers have significantly higher brand loyalty than the U.S. firms as measured by the coefficients on the repeat purchase variable in the vehicle choice equations. In addition, the brand loyalty of the U.S. manufacturers had sharply declined compared with the pre-1980 period. The authors point out that these results about brand loyalty are quite distinct from what they find with respect to brand preference. Holding equal the observed characteristics of the different makes and the different individuals, U.S. consumers would rather buy U.S. brands than Japanese brands. Americans want to buy American products. But the experience of past ownership of a given brand has a much greater positive effect on the chances of someone buying that brand if the past experience was with a Japanese car than if it was with a U.S. manufacturer. And the positive effect of past ownership has declined over time for the U.S. makers.

The authors then tackle an econometric issue that was discussed at length in the meeting. They interpret their results about the effect of past repeat purchases on current purchase decisions as indicative of brand loyalty. But Zvi Griliches suggested in his comments on the paper that it is possible that these same results could have arisen because of characteristics of the consumers and the manufacturers that had changed only slowly over time and had predisposed consumers toward a particular make. For example, suppose that there are some young professional households in the sample and one manufacturer has models that are particularly suited to such households. We will then observe repeat purchases by these households of that particular make. Now suppose that new models are introduced into the marketplace by other manufacturers that cater to this same group. We will then observe the young professional households spreading their purchases around more, and the pattern of repeat purchases will be reduced. The repeat purchases were never really indicative of brand loyalty, and the apparent loss of brand loyalty simply reflects the increased choices now available. This issue in econometrics is one of distinguishing state dependence (the brand loyalty case) from heterogeneity (the unobserved differences in household and auto characteristics).

Mannering and Winston use several methods of testing their results, and they find that these tests confirm strongly that their interpretation of the empirical findings is the correct one. First, the heterogeneity model assumes that there are unobserved characteristics of makes and households that are mimicking brand loyalty effects. The authors report that adding additional characteristics to their estimating equations had no effect on their results. Second, they performed three standard statistical tests designed to show whether unobserved heterogeneity is at work, based on the ideas of James Heckman and Gary Chamberlain. All of the tests failed to reject their original results.

The authors then turn to look at the implications of their results. They note that their findings for General Motors indicate that before 1980, the purchase of a GM car would increase the chances of a subsequent purchase by over 6 percentage points. After 1980 this had dropped to only 3 percentage points. Another way of looking at these results is that after 1980, Toyota could charge at least \$1,000 more for a comparable vehicle purely on the basis of its higher brand loyalty. The authors note that this finding is confirmed by the price differential that existed between a Toyota Corolla and the identical Chevy Nova (actually that differential was \$1,700, even more than the authors estimate).

Mannering and Winston also look at how brand loyalty affects market share. They ask what the U.S. manufacturers' market shares would have been if they had retained the same brand loyalty after 1980 that they had enjoyed before 1980. They find that General Motors lost 3.7 points of market share as a result of the loss of its brand loyalty, about a third of the total loss of share that the company experienced. Chrysler, they estimate, lost 2.4 percentage points of market share as a result of the decline in brand loyalty. The company was able to offset this by other factors, but has to be concerned that in the future these other factors may not be able to be a source of overall increase in market share. Ford suffered the least decline in its brand loyalty and did not have a decline in its market share. But all three of the U.S. makers, say the authors, have to be concerned about the strong brand loyalty that has been established by the Japanese manufacturers. This will make it a slow and difficult business for the U.S. firms if they are ever to regain their dominant positions in the U.S. auto market.

In his discussion of the paper at the meeting, Richard Schmalensee wondered if the authors were exaggerating the importance of brand loyalty in explaining what has been happening in the U.S. auto market. Brand loyalty is presumably important for a new purchase because past ownership of a particular brand reflects information people have about that brand.

xviii

Schmalensee noted that people obtain information about cars from many other sources, such as friends, neighbors, and auto test reports.

Davis and Haltiwanger on Wage Dispersion

Steve Davis and John Haltiwanger take as their starting point the large increase in wage inequality that has occurred in the U.S. economy since the 1960s. The gap between the highest-wage and the lowestwage workers has been widening. In this paper, Davis and Haltiwanger argue that we can increase our understanding of the rise in wage inequality by looking at data on individual manufacturing plants. They use the longitudinal research data base (LRD) prepared by the Center for Economic Studies at the Bureau of the Census that covers all plants in the periodic censuses of manufacturing and all plants surveyed in the Annual Survey of Manufacturing.

This data set gives information from the employer side—based on the jobs people hold—to supplement what has been learned in the past from studies of individuals and families, based on data sources such as the Current Population Surveys (CPS). Moreover, the authors are able to link their plant data with the data from the CPS in order to provide an overall decomposition of the variance of wages.

An important issue in understanding the rise in wage inequality is the extent to which it was associated with increases in inequality *between* groups of workers or, alternatively, the extent it is associated with increases in inequality *within* groups of workers. In the former case, for example, we might see that the gap between college graduates and high school graduates had widened. We would then interpret the increase in inequality as being the result of an increase in the return to education. Another example would occur if the gap between men's and women's wages had widened.

Davis and Haltiwanger first look at the wage data for manufacturing from 1975 to 1988 using the CPS, and they find, consistent with earlier studies, that within-group inequality has increased substantially over this period. The group characteristics being considered are education, experience, and sex. They confirm that these observed characteristics of the CPS manufacturing sample do not explain much of the rise in wage inequality. Nor do other characteristics within the CPS add much. This failure of the observed characteristics of individuals to explain the growth in wage inequality contrasts with the findings on the effects of observed plant characteristics. The authors find that these do have the potential to explain wage differentials, at least in a statistical sense. The size of plants, their capital intensity, and other differences are associated with wage differences. This means that changes in the nature of manufacturing plants could provide an explanation of the changes that have been found in the pattern of wages.

What are the changes that have been going on within manufacturing that could be affecting manufacturing plants? One of the most important changes, the authors argue, is that there has been a major change in the size distribution of plants since 1967. Plant size can be looked at in two ways. The first way is to say, How large is the average plant? This involves taking total employment and dividing by the number of plants. This approach says that if we pick a plant at random, then the average plant size reflects the expected value of the employment in the plant. However, this concept of average plant size is not the one that is most applicable to the study of the effect of size on wages. An alternative way is to compute what Davis and Haltiwanger call the "coworker mean." This says: Pick a worker at random and ask how many coworkers he or she has. The coworker mean is more relevant for studying the effect of plant size on wages because it talks about a typical worker rather than a typical plant. And Davis and Haltiwanger find that this mean size has fallen from 1,139 in 1967 to 665 in 1985. They also report that there has been a reduction in the standard deviation of plant size. The typical worker is working in a plant that is smaller, and the plants are becoming more uniform in size.

Another change in the nature of manufacturing plants is that they are becoming more specialized. Results by Gollop and Monahan for the LRD indicate that there has been an increase in product specialization between 1967 and 1982, with the biggest changes in the smallest plants. Davis and Haltiwanger investigate this idea in their own data and confirm the rise in specialization through 1986.

And finally, another important change is that there have also been shifts within the occupational and skill mix of workers within manufacturing. There has been a shift away from workers classified as operatives and toward managerial and professional workers. This has accompanied an increase in the proportion of manufacturing workers with some college education.

Davis and Haltiwanger then proceed to their first main empirical effort, which is to decompose the variance of hourly manufacturing wages into three components: between-industry variance, between-plant variance, and within-plant variance. This is accomplished by estimating the total variance from the CPS survey and then estimating the between-plant and industry variances from the LRD. Then the part of total variance that cannot be accounted for by the between-plant and between-industry variances must be the within-plant variance. They make this decomposition separately for production and nonproduction workers. The authors acknowledge that this leaves the estimate of within-plant variance as a residual, but they spend some time looking at the sensitivity of their estimates to potential errors in the data. They find that, while data errors cannot be ruled out, their overall conclusions are pretty robust, particularly the conclusions they reach about changes in wage inequality over time.

In table 2 the authors report their results. About 55 percent of the variance of manufacturing wages is accounted for by the dispersion of mean wages across plants. The mean wage gap between production and nonproduction workers accounts for about 7.5 percent, and the remaining 37.5 percent of the variance is accounted for by within-plant variance. They also find that the within-plant dispersion of wages is much greater for nonproduction workers than for production workers. Nonproduction workers account for 40 percent of hours worked in manufacturing but 63 percent of the variance of overall wages.

Looking between industries, the authors report that between-industry wage dispersion is similar for production and nonproduction workers, but the *relative* importance of the between-industry component is much greater for production workers (about 25 percent, compared with only 7 percent for nonproduction workers).

Davis and Haltiwanger then turn to the key question of the contributions of these alternative sources of variance to the overall increase in wage inequality. Notably, they find that the increase in betweenplant variance accounts for 48 percent of the growth in wage variance, and the increase in within-plant variance accounts for 41 percent. The rising wage gap between production and nonproduction workers accounts for the remainder. This overall pattern breaks down quite differently when production and nonproduction workers are considered separately. For nonproduction workers, 56 percent of the increase in wage inequality was within plants. By contrast, only 9 percent of the increase for production workers was from this source. This contributes a minuscule 1.7 percent to the growth in overall wage inequality.

Davis and Haltiwanger note that those who believe either that social norms have changed or that unions have diminished in strength, thereby allowing greater wage differentials to be paid, need to explain why this has not affected production workers. By contrast, these results are consistent with the hypothesis of skill-biased technical change. Hours worked by nonproduction workers rose from 33 percent to 40 percent of total hours between 1975 and 1986, and the average wage gap between the groups increased by nearly a third. This suggests an increase in the demand for high-skill workers.

The authors also suggest that the small rise in variance among production workers is consistent with the idea that production technologies may differ in the scope that they offer for individual differences in ability to manifest themselves as differences in productivity. Those technologies that use production workers intensively arguably offer less scope for ability and skill differences to generate productivity differences. A shift in the nature of technology toward production methods that are more skill dependent would then help explain the increase in the share of nonproduction hours in total hours and the increase in the wage dispersion among nonproduction workers.

Davis and Haltiwanger then go on to examine the extent to which the observed characteristics of plants can explain the pattern of wages. They focus on three questions: What are the basic patterns of variations that link plant characteristics to wages? How much of the between-plant dispersion of wages is accounted for by these plant characteristics? And what do time-series changes in the plant-wage structure tell us about rising wage inequality? The data base has information on industry, size, age, region, ownership type (single or multiunit firm), energy intensity, capital intensity, and product diversification. They find that mean wages are higher at larger plants, older plants, multiunit plants, more energy intensive plants, more specialized plants, and more capital intensive plants. Plant size has the biggest effect. The average wage gap between plants of over 5,000 and plants with 20 to 49 workers is \$4.92 an hour for production workers

and \$3.60 for nonproduction workers. There are also more nonproduction workers at large plants. Another interesting fact revealed by these differences is that a very high fraction of all hours worked are in plants older than ten years (about three-quarters) and in multiunit firms (about the same proportion).

When the authors look at the growth of wage inequality over time, they report that the decline in the proportion of the work force that is in large plants (most of which occurred by 1972) and the increase in the wage differential between large and small plants account for 36 percent of the increase in production-worker wage inequality. They note that size-wage differentials have increased steadily since 1967. Davis and Haltiwanger believe that this contribution of plant size to the observed increase in inequality of wages is an important new fact that must be confronted by any explanations of the pattern of wages.

To investigate further the role of observable plant characteristics on the wage distribution, the authors use the method proposed by Juhn, Murphy, and Pierce. This takes the increase in wage inequality over time and breaks it down into three parts. First, the authors look at the changes over time in the observed characteristics of the plants, holding fixed the wage differentials associated with those characteristics. They find that these changes had little impact on wage inequality. Second, they look for the changes in the wage differentials associated with the plant characteristics and find that these account for two-thirds of the rise in inequality. In fact the increase in the plant-size wage differential by itself accounts for a substantial fraction of the increase in inequality. And the third component of the increase in wage inequality reflects the role of unobservable changes in wages and characteristics. The authors are also able to use this same methodology to show that most of the increase in inequality comes about because of the increase in high-wage jobs. There is no increase in the gap between the average plant and the low-wage plants. The inequality is not increasing because of an increase in "bad jobs."

In their concluding analysis, Davis and Haltiwanger investigate further whether or not unionization and international trade variables can help explain the pattern of wages by size of plant. They use data on trade flows by four-digit industry and data on unions by two-digit industry to see if these can explain the changes in the structure of wages among plants. The coefficients reflecting deunionization and changing trade flows are generally statistically significant, but they are small. For example, the estimates indicate that an increase in the import penetration ratio as large as 10 percentage points would increase the size wage differential by only twelve cents an hour. A 10 percentage point increase in union density would reduce the wage differential between large and small plants by only one cent an hour.

The authors argue, therefore, that changes in trade flows, such as the increased imports of steel and autos into the United States, would not have induced the observed increase in the large-plant wage differential. Instead, the decline in the large steel and auto plants would have reduced the demand for high-skill workers and hence reduced the skill differential. They conclude that skill-biased technical change is a more likely explanation of the pattern of rising wage inequality than changes in trade flows.

In their discussions of the paper, Lawrence Katz and Robert Topel raised some concerns about the results. Katz argued that there was some inconsistency between the LRD and the CPS because of employees in nonmanufacturing establishments of manufacturing firms. These workers are excluded from manufacturing in the LRD, but included in the CPS. Katz also questioned the method that Davis and Haltiwanger used to impute the wages of nonproduction workers in the LRD. He claims that the authors' wage estimates for these workers from the LRD are 10 to 30 percent higher than CPS wage estimates. Topel argues that the decline in plant size that Davis and Haltiwanger report may reflect data errors and changes in the way establishments are defined. He also notes that most of the decline in plant size took place between 1967 and 1972, whereas the increase in wage inequality began in the 1970s and accelerated in the 1980s. Haltiwanger responded that in their view the decline in plant size is less important than the change in the quality of the average labor force in large plants.

Franklin Fisher on the Handbook

The field of industrial organization has undergone an explosion of research and interest in recent years, prompting the publication of the massive *Handbook of Industrial Organization*, edited by Richard Schmalensee and Robert Willig. For this issue of *Micro BPEA*, we asked two distinguished participants in the field of industrial organization, Franklin M. Fisher and Alvin K. Klevorick, to review the *Hand*-

book and give their own reflections on the field, on what we have learned and where research should be heading.

Fisher's paper focused on parts two and three of the *Handbook*, the sections dealing with the analysis of markets from a theoretical and an empirical perspective. Fisher notes at the outset that the *Handbook* is an excellent book, well-written and full of information, but it is only partially successful in achieving its goal of providing a survey of recent developments in the area. Some of the authors in the theory section provided too many details of their own work rather than giving a full survey. And in the empirical section, the authors have difficulty in providing an adequate critical guide to the literature. This is a particular problem, says Fisher, because of the gap between the theoretical and empirical literatures—a theme that he will return to.

Another concern that Fisher expressed is that two important areas were not well covered: empirical studies of individual industries and public policy. He points out that the main practical use of the analysis of market behavior is in antitrust policy, and there is no coherent treatment of this topic in the *Handbook*. Fisher suggests that had these two areas been covered, this would have provided an opportunity for the authors to give more linkage between theory and evidence. Or perhaps it would have revealed more clearly the gap between the theory and the actual practice in real industries or real antitrust cases.

Having looked at the *Handbook* as a book, Fisher then turns to consider what it reveals about the way that the field of industrial organization has developed, and he starts with a rather nihilistic perspective. Industrial organization has no organizing principles, he says, and the main result of the theory is that anything can happen. This is particularly true of game theory, which has come to dominate the field. The theorists construct interesting examples that are stripped of all but their essential features, and these examples are then used to demonstrate that some result can happen for rational reasons, usually a result that is the opposite of what had been thought before. Expanding the possibilities is a legitimate role for theory, says Fisher, but there is no general theory that is emerging as a result, so the field is left with a laundry list of possibilities.

The problem, says Fisher, is that the stripped-down examples take the realism away from the analysis, so that the theory is not providing a guide to actual practice. Stripped-down theory can be useful, he notes, but tractability is no advantage in deriving inapplicable results. This problem with theory is made worse by the fact that many theorists have a rather casual attitude toward empirical verification. And by the fact that much empirical work is performed with little reference to the available theory. For example, empirical studies often assume that competitive firms do not make supernormal profits, but this is not correct, he says. Supernormal profits are only bid away in long-run equilibrium, and even then there must be an adjustment for risk.

Fisher is particularly vehement in criticizing empirical studies that make use of accounting rates of return, which have been shown to be inappropriate. And another example where empirical work ignores theory arises in studies of innovation. Fisher says that he and Peter Temin showed some time ago that most empirical studies of the relation between innovation and markets do not in fact test what they purport to test. Despite the fact that the Fisher-Temin results are well known, most of the empirical analysis in this area has proceeded in ignorance of them. One area where there has been some recognition of the advances of theory in empirical work is in studies of industries with market power, the literature surveyed by Timothy Bresnahan. But the problem here is that the empirical results that have been found are not terribly helpful. Still, notes Fisher, these studies are far ahead of those in other areas.

Since he is so critical of the research agenda that has in fact been pursued, what does Fisher think the agenda should be? He says that the current approach to industrial organization theory is wrong. The strategy of constructing a simple model, assuming rational behavior and then deducing the conduct and performance of the economic agents, ignores the context in which the economic decisions are being made. As an example, Fisher points out that theoretical models of the pricing behavior of groups of firms find that the outcome should be the same for a thousand firms as it is for two. This is absurd, he says, and provides no guide to studies of the effect of industry concentration. We know that the two cases just described are different because if there are only two firms, they will find it much easier to cooperate in raising prices than will a thousand firms. Moreover, in any realistic setting, the number of firms represents only one of the factors that determine the industry outcome. Theory has shown that there are many possible equilibrium outcomes, and the appropriate research agenda for industrial organization is to determine which one of these is chosen. This will require

detailed studies of particular industries, studies that provide the contextual setting in which decisions are made.

Fisher also believes that detailed studies of particular industries and the antitrust cases that have arisen in them will also prove helpful in the application of industrial organization to antitrust policy. In order to determine whether a particular behavior was indicative of collusion, one should ask, *given the context of the industry*, whether it is plausible to assume that such behavior could only have been the outcome of such behavior or whether it could have resulted from purely individual decisionmaking.

Fisher concludes his study by stressing that he is not against exemplifying theory and not against theory that sets out all of the possible outcomes. What he does want to see is theory and empirical work moving more closely together. The theorists need to study real industries in depth and use models rich enough to serve as the basis for crossindustry comparisons.

Klevorick on the Handbook

Klevorick starts his review of the *Handbook* by pointing to the greater weight given to theory than to empirical work. The second volume is about applied issues, and it is much shorter than the first. And even in the second volume there is much theoretical material. This predominance of theory in the survey volumes accurately reflects the way in which the field has developed. The volumes' editors say in their introduction that they have been motivated by "new waves of work in industrial organization," particularly theoretical work. And they suggest that these have carried the field beyond the limitations inherent in traditional industry studies.

Klevorick has concentrated his attention on the sections of the *Handbook* that deal with the determinants of firm and market organization, international issues, and government intervention. He notes that different authors in the text use different approaches in their tasks, but one problem with all of these alternatives is that they are written by experts who are participants in the areas that they survey. Thus, while the authors try to give balanced assessments and to talk about problems, necessarily they are all enthusiasts for their approaches. The survey chapters might have

benefited from critical comments from observers who are more skeptical of the value of developments in the areas. Klevorick also notes that there are some serious gaps in the coverage, probably caused by authors who did not complete their tasks. In particular, there is no systematic treatment of empirical work on the contractual nature of the firm. And, as Fisher also laments, there is no single chapter giving an overall analysis of antitrust policies and their enforcement.

Klevorick looks next at the important trends in the literature that are identified in the *Handbook*. The first of these is that the theorists now consider the information available to the different participants in economic transactions, particularly the treatment of asymmetric information. For example, the literature looks at the mechanisms that are used by firms to write contracts with suppliers and with customers. A firm wants to contract for raw materials or components from a supplier on the best terms, but it does not know the technology or the costs faced by the supplier. Another example is that of the government agency charged with regulating an industry. In setting electricity rates, for example, the regulators do not have complete information about the firms that they are regulating, and these firms may have an incentive not to reveal all the information that they have.

Another important trend in the literature is that ideas or results from industrial organization have been crossing the lines and affecting other fields. For example, traditional models of international trade assumed that there was perfect competition in product markets. Modern trade theory stresses the importance of product differentiation and imperfect competition. And policy interests in several areas have stimulated the development of new ideas. For example, concerns about mergers and corporate reorganizations have prompted developments in the theory of organizations.

Klevorick points to the fact that the concept of the firm is a much richer concept today than it was in the industrial organization of some years back. The firm is no longer a lifeless vehicle for maximizing profit, but instead is a set of interrelations among the firm's owners, its managers, its workers, and its suppliers. Real people make decisions using the limited information that is available to them. This trend also carries over into the analysis of regulation. Regulatory policies are no longer seen as government fiats imposed on the firms. Instead, policies are the result of interactions between the regulators and the industries that they affect.

xxviii

As it has extended the concept of the firm, the recent literature has become more interdisciplinary, for example, as economists have incorporated information about the legal system and about accounting methods into their work. And political science and economics have been linked in efforts to understand the regulatory process. But in one respect the interdisciplinary effort has not gone far enough. The insights of individual psychology have not been brought into mainstream industrial organization. In particular, the psychologists have learned a lot about how people gather and process information, but the conventional utility maximizer of economic theory is still the central model for economic models, ignoring studies that find that people depart systematically from the expected utility model when they make decisions under uncertainty.

On a more positive note, however, Klevorick sees the fruitful interaction between law and economics as a key contribution of the recent literature. The theorists' studies of incomplete contracts and the importance of asymmetric information are tied to the legal concept of property rights. What things can be contracted for? What aspects of economic activity are covered by legal contracts and what aspects employ other institutional arrangements? To what extent are precedent or other implicit or indefinite contracts enforceable in law?

To illustrate the importance of the legal structure to the new economic theory, Klevorick gives the example of the two conditions that are necessary before there can be a viable contract, namely, observability and verifiability. To ensure the verifiability of a contract, some procedure must be available to determine whether the terms of a contract have been met. And clearly this involves the legal framework. But, in addition, legal rules will also determine what can or cannot be observed. What records are parties required to keep and who can have access to them? Indeed, Klevorick notes, the asymmetry of information is itself a product of legal rules that determine who has access to what information.

This same idea extends to the analysis of regulation. The extent to which a regulated monopoly is earning an adequate rate of return on its assets can be determined by the regulators only if the legal rules have granted them adequate and verifiable information. The legal system is also important in restraining opportunistic behavior either by the regulators or by the regulated firms. And this points to another important economic and legal interaction. When legislation is passed that regulates some aspect of industrial behavior, such as pollution control, the legislators may be unwilling to specify their goals and the methods to be used in order to achieve them. They are unwilling to deal with decisions that are conflicted and have adverse political fallout. They may instead devise a statute that puts the burden on the bureaucracy and the legal system to settle the most difficult issues.

Klevorick now turns to comment on the imbalance between theory and empirical work that he noted at the outset. He sees a particular problem in that much of the most interesting new work involves ideas about the firm, whereas there is little empirical work done in this area. He does single out Paul Joskow's studies of electricity contracts as exceptions, examples of empirical work that have evaluated the institutional and legal framework under which the contracts are determined. Another difficulty in empirical applications of the recent developments in theory is that it is hard to find quantitative ways to measure the informational asymmetries that are important in the theory. In fact, says Klevorick, the study of asymmetric information may lead us to reject traditional interpretations of standard results in regulatory analysis. Was the regulation inefficient, or did the regulators simply have incomplete information when they made decisions? (One hears an echo of Howard Baker at the Watergate hearings.)

Klevorick concludes with two implications for policy from the new industrial organization literature. First, policy initiatives must pay more attention to the legal environment in which they will operate. This will affect the way in which firms respond to any policy. Second, once we see the nature of the firm not as a fixed but as a changing entity that will change its nature when economic conditions change, then we must also take this into account in making policy. For example, in the antitrust area, the courts are considering whether or not a company can have an illegal conspiracy with another company that it partially owns.

There was a very lively discussion of the two papers on the *Handbook*, and Fisher's forceful criticisms of the field provoked much response. Joseph Farrell argued that the discovery that "anything can happen" is often very valuable. It avoids a mistaken belief that we know what the answer is in some situation. Both Farrell and Timothy Bresnahan pointed to areas where they thought that recent theory had made major contributions. For example, says Bresnahan, in 1975 the prevailing view was that cartels must break down. Subsequent developments showed how first-mover advantages or entry barriers can create permanent competitive ad-

vantages. Paul MacAvoy commented on Klevorick's paper, but ended up sharing Fisher's skepticism about what has been learned. He saw a little but not much in the *Handbook* that was helpful in dealing with regulatory policy. In his comments on Klevorick, Sam Peltzman looked at the empirical analyses of regulation in the *Handbook*, and he says that there has been no significant movement toward a more sophisticated conceptualization. For example, in looking at the effect of regulation, empirical studies have compared regulated and unregulated firms. But the test usually consists of including a dummy variable for regulation or testing some very simple hypothesis, such as the view that utilities tend to be overcapitalized. There is no attempt to capture the actual complexity of regulatory procedure.

Robert Willig on the Merger Guidelines

Most economists and policymakers judge that vigorous competition among U.S. companies is essential for maintaining the international competitiveness of the economy. However, the inappropriate use of antitrust policy can hurt consumers and retard growth if firms are penalized just because they are large or successful, rather than because they abuse market power. Where should the line be drawn?

In 1982 and 1984 a set of guidelines was developed at the Justice Department that specified when the department would challenge business mergers on the grounds that they were anticompetitive. The new guidelines developed a framework for merger analysis that was consistent with the prevailing economic theory of competitive behavior, but given the importance of mergers in today's economy, it seems appropriate to examine these guidelines in order to see how they are working and whether there have been more recent developments in theory that suggest the need for fine-tuning. This is the task that Willig takes on in this final paper, and he starts by sketching the steps that are given in the 1984 guidelines for the analysis of horizontal mergers.

These steps are delineation of the relevant product and geographic market, identification of the firms participating in the market, calculation of the concentration of these producers, assessment of ease of entry and other factors, and assessment of efficiencies. The most novel and central of the steps is the delineation of the market. The guidelines define a market such that if there were a monopoly supplier to that market, then the monopolist would find it advantageous to impose a "small but significant and nontransitory increase in price." This has become known as a "SSNIP," and it is usually taken to be a 5 percent price increase or more. The relevant market is generally defined by looking for *the smallest group of products and locations for which the SSNIP criterion holds*.

The participants in the market consist of all of the firms currently supplying the products included in the relevant market, plus firms that would be induced by a 5 percent price rise to supply the products within a year and without the need to commit to new facilities. Once the list of participants is found, the market shares are determined and hence the degree of concentration is calculated. Concentration is measured by summing the squares of the percentage market shares, thereby forming an index known as the Herfindahl-Hirschman index. If this index is below 1,000, the market is considered unconcentrated and the merger is permitted without further analysis. If it is in the range 1,000 to 1,800, mergers will be scrutinized, particularly if the proposed merger increases the index by 100 points or more. For values of the index above 1,800, any proposed merger that raises the index by more than 50 points is considered "a matter of significant competitive concern," and in such cases the guidelines call for further analysis of other competitive factors and potential efficiency gains from the merger.

Willig takes up a simple model in order to understand the rationale for these guidelines. This model ignores potential entry and any gains in efficiency that might result from the merger, but with this simplification, it provides a clear validation of the approach to merger policy that is set out in the guidelines. The potential loss of economic welfare that results from a merger can be related to the concentration of the industry and the way in which the sellers in the industry behave. With a given mode of competitive behavior, the model shows that the welfare loss depends on the level of concentration and its change.

Turning to the possibility of efficiency gains, the guidelines say that "clear and convincing evidence" of significant gains will be taken into account when applying the guidelines. Such evidence could be used to counter the "competitive concern" that might otherwise cause the department to oppose a given merger. Willig says that the difficult problem of assessing efficiency gains is not one that he will focus on in his

xxxii

study, but he notes that efficiency gains do play an important role in the way in which the guidelines have been structured. In most cases, the guidelines will allow firms to achieve gains through mergers without interference from the Department of Justice.

Willig notes that in the guidelines there are several other factors that are cited as relevant when a determination is made about the anticompetitive effects of mergers, but he notes that most of the discussion of such effects focuses on the likelihood of tacit or explicit collusion among firms. This, he says, has led to a neglect of the detailed factors that are key to analyzing other anticompetitive consequences from horizontal mergers. He will explore some of these, therefore, and he starts this process by dividing competitive effects into two types. The first is where a merger allows the merged firm to profit from anticompetitive unilateral action, and the second is where the merger increases the chances of tacit collusion in the industry.

Willig starts his discussion of the unilateral case by looking at Cournot models. These assume that firms will act unilaterally and take the actions of other firms in the industry as given when setting their own output. Willig points to results derived by Farrell and Shapiro and examines his own version of their model. The lesson from these Cournot models, says Willig, is that mergers between firms with significant market shares may lead to price increases because of the additional profits that can be obtained by curtailing output. There may be some offsetting gains, but these are hard to predict. And the models also show that where the firms that are not directly involved with the merger are not concentrated or where they can readily expand their own output, then this will protect consumers against the adverse effects of mergers.

The Cournot models can give interesting results, but their assumption about the homogeneity of products gives them limited applicability. These models do not seem applicable to the cases often found in practice where there are differentiated products. In such cases, Willig argues, Bertrand models are more applicable. In Bertrand models, firms set the prices of their products by weighing the benefits from higher prices against the losses from reduced sales as customers switch to the products of other firms. Following a merger, there is a change in the incentive for price increases because some of the diversion of sales following any increase in the price of a product sold by one merger partner will go to the other merger partner. In principle, this analysis suggests the need to determine the extent to which differentiated products in a market are good or bad substitutes for each other. In some cases there might be enough data available to do this and when there is not, it may be possible to use market shares instead. The models suggest, says Willig, that using concentration as a basis for merger evaluation can still be valid because market shares provide an indicator of the general appeal of different products. This result is somewhat fragile, however, and would change if the underlying assumptions were changed. In particular, inferences from market shares are valid only if the products are neither very similar nor very different. Otherwise, the market shares can understate or overstate the expected competitive effects from a merger between their sellers.

Willig concludes his discussion of the case of markets with differentiated products with a caution. The characteristics of products can often be changed. If this is the case, then a merger that raises the prices of the products of the merged company will likely cause other companies in the industry to reposition their own products, and this will reduce the anticompetitive effects of the merger.

Willig turns next to a discussion of entry. The possibility of entry is considered in the guidelines in two forms. First, the group of firms that are considered market participants includes "production substituters," which are those firms that would be induced to enter the market within a year in response to a SSNIP without having to make a significant investment. And then, second, the guidelines treat entrants differently from substituters. A full-blown entrant is defined in the guidelines as a firm that would take longer than a year or require a larger commitment of resources to respond to a 5 percent price rise.

Willig argues that theory supports the idea of including production substituters among market participants, and this means that if there are many production substituters out there, then the measured index of concentration will be low.

Willig argues that entry requiring commitment should be considered as a possible response to the outcome that is the matter of concern in the particular case. If there is concern that the merger will lead to higher prices as a result of diminished competition, the question is whether there will then be entrants who see higher returns in that industry over an extended period. The availability to potential entrants of the technology and other necessary factors of production will obviously be

xxxiv

relevant. The likelihood, timeliness, and sufficiency of the entry induced by any initial reduction in competition will then determine whether or not potential entry will be effective in deterring an anticompetitive merger in the first place, or whether entry can adequately replace the competition lost as a result of the merger.

In his conclusion, Willig says that the 1982–1984 guidelines may need some fine-tuning, but basically they reflect much of what has been learned by theorists and in applied work. The most important area for further research is the impact of horizontal mergers on competition for innovation.

In his comments on the paper, Steven Salop notes that the guidelines needed revision in the way they treated entry and the way they treated unilateral rather than collusive behavior. He says that Willig has made a provocative start on these problems, but that more work is needed, particularly for entry, where the guidelines fail to provide a detailed analysis. F. M. Scherer criticized Willig's analysis because of its reliance on Bertrand models. These assume that each firm will take the prices set by other firms as given when setting its own price. Scherer argues that in practice firms will anticipate the reactions of other firms to any price changes that they make. He says that merger policy should focus on the conditions under which firms start to act collusively rather than refining models of unilateral action.