Summary of the Papers

THIS SECOND microeconomic issue of the *Brookings Papers on Economic Activity* begins with three studies of economic regulation. A main economic policy initiative of the past eight years, one that was intended to improve the efficiency and productivity of the American economy, has been to reduce the role of government in the regulation of private industry. The first of the three studies examines the validity of the economic theory of regulation, the theory that has provided the intellectual basis for the deregulation movement. The second examines how deregulation has worked in practice in the airline industry—a success story in the eyes of many economists but a disaster in the eyes of some members of the public. And the third looks at the electric power industry, one that many people believe is ripe for partial or complete deregulation.

Next comes a paper examining the pattern of wages across industries. The authors argue that there is a solid empirical basis for the idea that some industries provide good jobs and some bad jobs. And policy should encourage the growth of good jobs.

The issue concludes with two studies of patents and technology development. Concern has grown about the flow of inventive activity in the economy, whether it may have slowed down and whether adequate protection is being provided for intellectual property. These papers look at recent trends in patenting and explain a striking new method that has been developed to value the protection that patents provide.

Peltzman on Deregulation

In the traditional view, regulation occurs because of a failure in the way the private market operates, so that regulators must step in, for example, to prevent natural monopolies from raising prices and exploiting market power. This traditional perspective is what Sam Peltzman

has called the normative as positive view of regulation, meaning that actual regulation is said to be based on improving the public interest.

This normative view of regulation has been challenged by economists like George Stigler, Gary Becker, and Peltzman, who have developed an alternative positive theory of regulation, applying certain principles of economics to explain the behavior of politicians and regulators in practice. In their "economic theory of regulation," regulation is a self-interested exercise by politicians and regulators who use their power over industries to win reelection or to keep lucrative jobs.

Although this way of characterizing regulation has been controversial, it won considerable support and encouraged advocates of deregulation. But ironically, the very success of the deregulation movement in bringing about change has brought the new theory into question. Deregulation has greatly reduced the power of regulators in many industries. If the theory is really correct, and the regulations were introduced in the first place just to serve the interests of politicians, then why did deregulation happen? Why were the politicians and the industry groups that often supported them willing to allow and even encourage deregulation?

In his paper Peltzman reviews the economic theory of regulation, showing how it developed and the kind of evidence used to support it. He argues that in principle the deregulation movement was not inconsistent with the economic theory of regulation, which does have a basis for predicting deregulation. This theory, Peltzman maintains, should be judged on whether deregulation has occurred in the industries where the theory predicts it should occur. He looks at several deregulated industries to see if they fit these predictions.

The assessment of alternative ideas about the nature of regulation is important in understanding the pressure for further deregulation or the counterpressure for reregulation—to restore old procedures in some industries. Peltzman's paper and the two that follow, on airlines and electric power, will help us to determine what has happened and the right strategy for regulation and deregulation in the future.

According to Peltzman, George Stigler developed the economic theory of regulation because the normative view of regulation did not fit reality. In the normative view, regulation is motivated by market failures, and the regulators take steps to correct those failures. Evidence was accumulating from industries where prices were regulated, Peltzman

says, showing that when the market worked—for example, in industries with many small companies (such as trucking)—regulation raised prices. And where the market was not providing competitive discipline on prices—for example, in public utilities that have a monopoly—regulation did not succeed in keeping prices below the level they would have reached without it.

This failure of the normative view had generated ideas for alternative theories of regulation even before Stigler's contribution. In particular, some observers claimed that the regulatory process was simply taken over by the companies or groups being regulated. Regulation therefore became a legally sanctioned device to capture the potential returns from pushing up prices or otherwise intervening in the market. Stigler developed this idea into a behavioral model, an economic *theory* of regulation, by specifying what motivates the regulators and what constraints the customers will face. He argued that politicians design regulatory policies to maximize the votes they receive and pointed out that voters who want to organize to protect their interests will incur costs. Drawing on the theory of rational ignorance developed by Anthony Downs, Stigler recognized that consumers will rarely find it worthwhile to organize to oppose regulations that raise prices or interfere in the market in some other way. No specific regulation affects consumers much, but most regulations affect a few producers a great deal. Stigler's theory, therefore, predicted precisely the situation the empirical evidence seemed to be showing; namely, that groups of well-organized producers determine the outcome of regulation by contributing money or other support to the politicians who agree to regulate an industry in the way the companies in that industry want.

Stigler's theory has had a tremendous influence on economists and on many policymakers, but it was evident even as he presented his ideas that things were more complicated than his theory allowed for. Richard Posner criticized it, pointing out that regulation in practice contains many features that do not seem to fit well with the view that regulation is simply taken over by producers' interests. For example, regulations often involve the subsidization of certain groups of consumers. Before Amtrak was established, there was extensive regulation of passenger rail transportation, including the enforced provision of service for some customers along routes where the companies were anxious to get rid of

the service. Here was a clear case, among many others, in which producers' interests did not fully control the way an industry was being regulated.

Peltzman's own contribution to the developing theory was to specify that politicians will fine-tune their regulatory strategy by taking some profits away from producers to give benefits to certain groups of consumers—in return for winning *their* votes. Peltzman carried the methods of economics further into the analysis by deriving efficient conditions for the regulators. Each vote has a "price," and specific regulations will be introduced to favor selected groups of consumers or producers if their votes can be obtained at a price worth paying, measured by what other voters have to give up. For example, the provision of passenger train service was valuable, at least to some voters, and the railroads did not oppose a politician who kept such service going if he favored other rail regulations they wanted.

Peltzman's analysis opened the door slightly for several ideas of the normative view of regulation. In some markets regulation can improve efficiency, and in all markets certain ways of regulating are more efficient or less inefficient than others. Presumably it is in everyone's interest to reduce "dead-weight losses," which occur when regulation hurts some group without helping another. Becker developed this idea further. In his analysis, dead-weight losses act to limit the inefficiency that regulation can tolerate. To benefit producers, for example, regulators will have to raise prices to consumers or take something else away from them in another way. Each successive increase in price also decreases efficiency, so that the value of what producers gain is less than the value of what consumers give up. At some point, the political pressure from the losers builds up, and the regulator must provide protection for consumer interests to avoid being voted out of office. Becker's analysis suggests that the inefficiency involved in regulation has a limit.

Becker's approach also provides an answer to the question why regulation occurs in some industries and not in others. Regulation is more likely to take place when there is a specific market failure that could be corrected by the regulation. Since market failures are rare, so regulation will be rare. Also, it is difficult to change an existing market or regulatory situation. Most competitive industries do not have price regulation, because political pressure would build if regulators came in and started raising prices. This idea can then explain how regulation in

practice can have some of the characteristics described by the old normative view. In industries where the market is not working well, regulators can come in and provide benefits to producers without incurring large costs to consumers. Of course, important differences remain between the two theories. The Becker analysis still predicts that well-organized producers will be able to influence regulation for their own purposes.

Peltzman turns next to the key question of his paper. How does the experience of deregulation fit with our understanding of the causes and consequences of regulation? The normative view suggests that deregulation should take place when market or technological conditions change in ways that eliminate the market failure which prompted the regulation. The economic theory of regulation predicts that deregulation will take place when there have been changes either in political influence or in market conditions that have made the payoff to the participants too small. Peltzman's discussion focuses on the economic aspects rather than on possible political changes.

Railroads. How regulation was introduced into this industry fits with the ideas developed in the economic theory of regulation, says Peltzman. Producers were given protection from competition, but they shared these gains with others in the form of subsidies, such as the one to passenger service that producers had to provide. A regulatory equilibrium developed in railroads at the end of the nineteenth century and in the early years of this century.

This equilibrium ran into trouble as alternative methods of transportation developed, particularly trucks. The immediate response was also to regulate trucks, thereby preserving the railroads' stake in regulation, even though their benefits from it were declining. But the benefits to the railroad companies declined even more as the trucking companies and then the Teamsters union fought for greater returns for themselves. Over time, regulation actually became unprofitable for the railroads; Peltzman gives the early 1970s as the turning point, marked by a spate of railroad bankruptcies. The railroads then chose to support deregulation, and they got it.

Peltzman argues, therefore, that the railroads fit the economic theory of regulation very well. Regulation began as a way to generate benefits for producers, and when those benefits had eroded, the regulation was abolished.

Trucking. The regulation of trucks yielded its biggest returns to unionized trucking workers. Peltzman notes that estimates of the wage premium paid to these workers was more than 30 percent and was growing over time, not declining. Thus the deregulation of trucking does not fit with the economic theory of regulation. Regulation was providing a continuing or even a growing return to producer interests (in this instance the workers), and yet it was abolished.

Nevertheless, even here the theory may have some validity, says Peltzman, because the trucking *companies* were not doing so well under regulation. Regulation had led to costly competition over market shares, and the deregulation of the railroads had allowed them to compete more effectively against trucks. But that explanation is still not enough to make truck deregulation consistent with the economic theory.

On the other hand, the story of trucking regulation does not provide much support for the normative view of regulation either. There was never much evidence of market failure in the industry when the regulation was introduced, nor is there evidence that the technology had changed when deregulation occurred.

Airlines. The regulation of airlines was similar in structure to that of trucking, but it did not generate a continuing flow of benefits for producers in the same way. Until the late 1960s regulation provided returns to the airline companies and their employees of about 10 percent of revenues. Then this return began to erode because the airlines engaged in competition in providing service. Most important, they put on flights even when the flights were far from full. Average airline load factors fell from 63 percent in the 1950s to 48 percent in the early 1970s. This kind of service competition had become more important after the introduction of jet aircraft, which widened the scope for nonstop service. The Civil Aeronautics Board had responded to this change in technology by allowing more airlines to compete on long-haul routes. By 1978 regulation was no longer providing benefits to the airline companies and was abolished.

Peltzman notes that the air transport unions may still have been obtaining benefits from regulation at the time of deregulation, but in his judgment the erosion of returns to the companies was the decisive reason for deregulation, and therefore the case of airlines does support the economic theory of regulation.

Long-distance telephone service. Until the 1960s long-distance tele-

phone service was provided by AT&T, and this was thought to be an industry where the technology required one monopoly provider of service. Microwave technology was developed after World War II, and by the 1960s long-distance telephone service could be operated more cheaply by this means than by cable even if the volume of telephone calls was fairly small. Thus private companies could operate their own long-distance service to supply their telephone needs independently of AT&T. From a technological viewpoint, these companies could then rather easily connect their own long-distance telephone systems into the local telephone exchanges and sell long-distance service to other customers. To do so was illegal until 1969, but in that year MCI, an owner of its own microwave system, obtained permission to provide public long-distance service. This was the beginning of the end of the old regulated monopoly structure.

Does the telephone case fit the economic theory of regulation? Perhaps it does, for the regulators may have seen by 1969 that the new technology was bound to undermine the benefits of regulation. To run alternative long-distance service and avoid paying the regulated telephone rates was going to be too easy. Peltzman is doubtful if that story can be sustained, however. It suggests that the regulators were farsighted enough to realize as early as 1969 that the game was up for getting a return out of regulation. On balance, Peltzman judges that the normative view of regulation explains better what happened. The microwave technology eliminated the market failure that had existed with the old cable technology because now it became possible to have many competing companies.

Stock brokerage and banking. As far back as the eighteenth century, members of the New York Stock Exchange formed a cartel to set brokerage fees. The cartel was regulated after the 1930s by the Securities and Exchange Commission, but this body allowed the setting of minimum rates until 1968; full-scale deregulation did not take place until Congress took action in 1975. The deregulation of rates in this industry was preceded by an increase in stock trading by large institutions. These large traders were very desirable clients for the brokerage houses when fees were set without discount. A large block of shares could be bought or sold at a much lower cost per share to the brokerage house, and hence at a much greater than normal profit. The brokerage houses therefore began to compete with each other for these clients, using service

inducements such as providing research on the companies whose shares were being traded. At the same time, the institutional traders were circumventing the regulated fee structure by arranging to trade shares without going through the New York Stock Exchange. Between these two forces the returns to the brokerage houses from regulation were being eliminated.

In the banking area, regulation was introduced in the 1930s, and the interest rates payable on different accounts were set. Through the 1950s and early 1960s the restrictions on rates were not an important constraint on the banks; regulated banking offered other advantages valuable both to the banks and to the home-building industry. In the late 1960s there was a tremendous increase in both the level and volatility of interest rates. This created problems within the banking industry as S&Ls and commercial banks started drawing money away from each other, depending on how the different rate ceilings were set. And when unregulated institutions began to offer interest rates that were several hundred basis points above the rates banks were paying, disintermediation occurred as funds were pulled out of the banking system altogether. At this point it was clear that the regulation of interest rates was hurting rather than helping the banks, and in 1980 and 1982 rates were deregulated.

Peltzman argues from these events that the experience of both the brokerage and the banking industries supports the economic theory of regulation. When regulation became counterproductive to the interests of the banks and brokerage houses, it was eliminated.

Oil. Some writers have looked at the oil industry and argued that it is inconsistent with the economic theory of regulation. Maximum prices were set on oil in the early 1970s, and the price increases initiated by OPEC pushed world prices well above the U.S. prices. The government eliminated price ceilings in 1980 and instituted a windfall profits tax that levied a special tax on the oil companies based on the gap between the market price of oil and a base price. After 1985, however, the market price remained below the base price, so there has been no effective excess tax since then. But before 1985, producers' interests were apparently being hurt by the regulation.

Peltzman argues that the key to understanding regulation in this industry is to look back at the way regulation worked before the energy crisis. Until the 1970s the industry had been regulated in a way that

clearly provided profits to the producers. There were quotas on output and quotas on imported oil. This regulation had come about in the 1930s as prices became depressed. Until OPEC succeeded in raising prices, the increased availability of foreign oil had threatened the profits of domestic oil companies, and they had responded with political action.

OPEC's dramatic price increases upset the political balance and led inevitably to pressure for a redistribution of benefits toward consumers. Price controls were therefore initiated, followed by a windfall profits tax. The oil companies were still left with larger profits as a result of the OPEC price increases; they were simply giving up part of their returns in order to restore the political balance.

In Peltzman's view, therefore, oil industry regulation does not contradict the economic theory of regulation. The thrust of regulation had been to generate returns for the oil producers, but the OPEC shock meant that the distribution of the returns to domestic oil production had to be reassessed.

In concluding his review of the evidence, Peltzman maintains that the economic theory of regulation has itself evolved. It now emphasizes the need to balance the pressures coming from competing interest groups, and it recognizes that when regulation starts to generate inefficiencies that are large or outweigh the returns the regulation can generate, pressure will build for deregulation. The analysis of deregulation is an extension of the economic theory of regulation. Peltzman finds that most of his case studies of deregulation are consistent with the economic theory and that only one gives much support to the alternative normative view of regulation. He acknowledges, however, that not all the cases fit the same pattern and that important unresolved questions remain about how regulatory and deregulatory policies are determined.

At the meeting, Roger Noll criticized Peltzman for neglecting many contributions to the theory of political behavior that lay behind the new theory of regulation, particularly work by Mancur Olson, and for neglecting alternatives to the normative approach, particularly work by Kenneth Arrow and others on voting behavior, which has spawned a large literature on regulation. Noll argued that Peltzman was not allowing several worthy runners to compete in his race between theories.

Noll and Michael Levine also questioned Peltzman's conclusions from the evidence. The railroads had been losing money for years before deregulation took place. The payoffs to regulation in the telecommuni-

cations industry and the airline industry were both rising over time not falling when deregulation took place, Noll and Levine contended.

Other participants in the meeting commented on the history of the introduction of regulation in many industries. Companies were often very opposed to the onset of regulation, which does not fit with the idea that the producers desired regulation. Many participants, however, agreed that the key issue in regulation is whether it is driven by a public-interest effort to deal with market failures or whether the regulatory process is taken over by producer interests.

Morrison and Winston on Deregulation of Airlines

One of the important examples of deregulation considered by Peltzman is the deregulation of airlines. In the next paper Steven Morrison and Clifford Winston examine the air passenger transportation system in the United States to see how it was affected by deregulation. They then ask how public policy could be used to enhance the performance of the system.

Almost all economists favored the deregulation of the airlines, but the flying public is not enthusiastic about the recent outcome of the deregulation effort. Morrison and Winston have argued in previous work that deregulation has brought substantial benefits, and they develop new findings here which support that result. They then go on to argue that the problems with the system have arisen because public policy was not changed appropriately when deregulation occurred. For example, the public often blames deregulation for flight delays, but in fact flight delays have been caused by poor policy decisions that resulted in inadequate ground facilities, such as runways and air traffic controllers, and in the failure to set landing fees appropriately. Morrison and Winston contend that deregulation should be accompanied by advisory bodies to coordinate public policy with the effects of the deregulation. If such steps are not taken, they warn, the public may apply pressure for reregulation.

The bulk of the paper deals with the empirical analysis that provides the basis for the conclusions just described. The authors start by looking at how passengers choose among the different flights that serve a particular route. By studying these choices, the authors can estimate the values passengers place on different aspects of a plane journey—for

example, how much value passengers place on shorter flight times or how much passengers will pay to avoid increased waiting time on the ground. This method will allow them to make an overall assessment of the effects of deregulation, for if they can figure out the ways in which deregulation has changed the system, they can then value these changes as passengers would value them.

Morrison and Winston put together their own data set to estimate how passengers choose the airlines and routes they will fly. They use the *Official Airline Guide,* and for each origin-destination pair they consider possible alternative routes, looking at the fares, at the different airlines, and at the different stopovers that may be made. Passengers are assumed to be influenced by the fare, total travel time, time taken to transfer within an airport, and on-time performance as well as safety records. Promotional offerings, particularly frequent flier programs, and frequent service are also regarded as advantageous to travelers.

The authors' estimates of the impact of these various factors on passenger choice are based on a random sample of flights from their sample of routes in the third quarter of 1983. Their results strongly support the framework they are using. Average fare, travel time, schedule delay, within-airport transfer time, safety, on-time performance, complaints, frequent flier miles, and other airline performance indicators all showed up as having a significant effect on passenger choice. Morrison and Winston found, for example, that passengers will choose, say, a four-hour flight rather than a five-hour flight if the fare for the shorter flight is no higher than $35 more than the fare for the longer flight. Passengers are willing to pay even more to reduce the time spent in transferring within airports. The on-time performance of an airline was also found to be important: a 25 percentage point improvement in the proportion of a carrier's planes that arrive on time is about equivalent to a fare reduction of $30. Also important is an airline's safety record: a recent fatal accident suffered by one of an airline's planes will discourage passengers about as much as a fare increase of $70.

The authors draw some policy implications from their results. Since passengers, on average, will choose convenience (shorter travel times) over low fares (up to a point), the system should reflect this preference. Moreover, their results show that the market provides strong incentives for safety, quite apart from what is done through safety regulation.

Morrison and Winston also use their model of passenger choice to

look at the issue of airline mergers. The case for deregulation hinges on the presumption that competition among airlines will keep fares down and encourage the appropriate level of service, given passengers' preferences. The invisible hand of competition should do a better job of regulating an industry than the regulators do. Since deregulation, a number of airlines have merged, causing some concern that mergers have reduced competition. In particular, it has been suggested that the airlines have carved up the national market by developing operations in different cities and that these hub operations then form local monopolies, allowing the carrier that dominates flights out of a particular city to raise fares.

To assess the effects of airline mergers, Morrison and Winston consider six mergers approved during 1986–87. Their model of how passengers choose flights and the value they put on various attributes of a journey allow the authors to estimate the costs and benefits of those mergers. Specifically, they look at how average fares, travel time, transfer time, frequency of service, frequent flier programs, and other characteristics were affected by the mergers. They selected a sample of 115 routes for which at least one of the merger partners provided service. An important element included in their analysis was whether each airline was operating a flight out of a hub.

The authors do find that the mergers influenced fares. The number of competing airlines on a route does have a measurable effect on the fares charged. For example, if a route drops from two competing airlines to only one airline, the fare goes up 9 cents a mile, or about $90 for a flight of 1,000 miles (roughly the average flight distance), an increase of about 30 percent. But a reduction in the number of competing airlines from three to two has a much smaller impact. Fares rise only about 1 cent a mile. The effect of losing a competitor from a route was found to be influenced by whether hub operations were involved. Depending on the arrangement of the merging airlines in hubs, the loss of one airline could raise fares as much as 55 percent. These results indicate, therefore, that airline mergers can increase fares, and raise concern that mergers may erode the benefits of deregulation by limiting competition. In other respects, however, they can benefit passengers. Mergers, the authors find, can help passengers by combining frequent flier programs. And they can also help if they permit improvements in route structures that reduce travel time or airport waiting time.

The authors' findings on mergers are summarized in table 6 of the paper. It shows that half the mergers provide a net benefit to passengers and half a net cost and that the overall net benefits somewhat outweigh the net costs. This result, however, depends on the existence of frequent flier benefits. And an important reason why passengers value these programs is that business travelers can receive a tax-free fringe benefit from their employers. Part of the value of mergers to airline passengers is therefore being paid by the U.S. Treasury. When this tax effect is factored in, the costs of mergers to passengers outweigh their benefits.

The authors voice another concern about frequent flier programs besides their tax treatment. Since an airline established in a hub will have many passengers who have accumulated frequent flier miles, another airline coming into the hub to compete would have to offer lower fares to attract customers. Frequent flier programs may discourage competition from taking place.

Morrison and Winston next ask how deregulation has affected the problems of congestion and delay. Many people cite delays as an important reason for their opposition to the changes that have taken place in the industry. The hub-and-spoke mode of operation fostered by deregulation has put stress on airport capacity. In this arrangement an airline feeds many of its flights into an airport at about the same time, and then many of the passengers change planes. Once the airport musical chairs is over, the planes all take off again at about the same time. This concentration of incoming and outgoing flights increases congestion, which has also been exacerbated by the tremendous growth in air travel—a success of deregulation.

Morrison and Winston argue that the main causes of congestion are the failure to set landing fees that reflect the scarcity of landing space and the failure to invest adequately in runways and air traffic control. As regards landing fees, the authors note that when a plane lands in a crowded airport, it imposes a cost on other planes whose passengers have to wait in order to land. Landing fees have traditionally been determined by the weight of an aircraft. As a result, a small jet carrying only a few people may land in a busy airport and pay a fee that disregards the fact that it is holding up a jumbo jet with several hundred people on it. The persistence of this inefficient pricing system reflects the political power of the people who fly in the small jets, the kind of thing Peltzman's paper talks about.

Morrison and Winston find that if landing fees were set to reflect the real congestion costs a landing imposes, the welfare of airline passengers would be enhanced. If the improvement in pricing system were combined with an increased investment in runways and air traffic control, the total improvement in the air transport system would be even greater.

In discussing the policy implications of their findings on congestion, Morrison and Winston point out that increasing the availability of runways is a key element in the political feasibility of a shift to more efficient pricing of landing fees. Without extra runways, efficiency will be achieved by pricing the small planes (part of general aviation) out of the airports during peak times, a policy change that would be very hard to enact. Unfortunately, the authors say, increasing the number of runways at many airports may also be difficult because of the opposition of people living around airports to any increase in air traffic. All in all, there are serious obstacles to any of the solutions, and the authors suggest that a strong federal initiative is needed to make much progress. It is imperative to take such steps, they say, to relieve the airport congestion problems plaguing the system.

Morrison and Winston turn next to the issue of air safety, noting that results already reported show that the private market does provide substantial incentives to maintain safe airline operations. And these incentives have surely encouraged airlines to stress safety and have contributed to the fact that the number of air fatalities per passenger-mile flown has declined steadily since air travel began. This downtrend in the accident rate is due to the introduction of radar and improved landing and navigational equipment and to the better training of pilots using simulations of real-life dangers.

Despite the improvement in the safety record of air travel, the public concern about air safety has increased sharply in recent years. And the authors claim that this concern has some basis: the Federal Aviation Administration (FAA) and other parties to airline safety regulation have not provided the investments in new technology and the levels of trained personnel needed to handle the increased volume of traffic that followed deregulation.

However, the authors argue, there is no evidence that deregulation has had an adverse effect on safety *so far*. They reach this conclusion by looking at the causes of accidents. If deregulation had been a source of more accidents, they would have expected to see a larger fraction of accidents caused by maintenance problems and by inexperienced pilots.

But as reported in table 16 of the paper, the causes of accidents do not seem to have been affected by deregulation.

If the reduction in the accident rate is a trend that is to continue, the FAA should work to reinforce the market forces that encourage safety. Such work has become more difficult, Morrison and Winston argue, because of interference in the agency's operation by the Department of Transportation and Congress. This has led to poor performance in areas relevant to avoiding weather-related accidents and pilot error. In particular, there have been delays in introducing a new computer-oriented air traffic control system. The solution is to establish a separate tactical branch of the FAA concerned with traffic control, and a strategic branch concerned with long-term safety issues.

Having looked at the deregulation of airlines from the standpoint of how passengers value the elements of air travel, at the problems of airport congestion, and at the effect of mergers on airline competition and safety, the authors then assess their findings. Deregulation is a problem in that public policy has not evolved hand in hand with the changes induced in the private sector. The authors believe their proposed policy changes provide the basis for a plan of reform for the air transportation system. In particular, there must be an initiative at the level of the U.S. Department of Transportation to institute the changes in landing fees and runway construction needed to overcome the current congestion problems.

Elizabeth Bailey and Alfred Kahn, two important participants in the effort to deregulate the airlines, provided commentary on the Morrison-Winston paper at the meeting. Both expressed broad agreement with the paper's conclusions, though they had some concern about the reliability of specific statistical estimates of the impact of changes in the system. Kahn argued that the decision by passengers to choose different fare classes was an important element that had been ignored. Bailey pointed to the effect of the firing of the air traffic controllers on congestion and delays. Both applauded the authors, however, for their efforts to assess landing-fee changes in concert with an extension of runway capacity.

Joskow on Regulation of the Electric Power Industry

In the first two papers the authors consider issues raised by the revolution in deregulation; in his paper Paul Joskow examines an industry

that has remained heavily regulated. Prices and operating conditions are still set by state and federal regulatory agencies. Even so, Joskow says, the industry is undergoing some fundamental changes, most important the growth of a more competitive wholesale market in electric power for resale to the regulated utilities. In addition, certain regulatory changes have led to greater competition in the provision of future generating capacity.

That fundamental changes are being made in the industry reflects problems that have arisen in the last twenty years. In the 1950s and 1960s rapid productivity growth in the industry permitted the price of a kilowatt-hour of electricity to fall. This encouraged rapid growth in electricity demand, while allowing comfortable profit margins for the utilities. The situation started changing in the late 1960s, and the energy crisis of 1973 greatly exacerbated the problems. Productivity growth disappeared, fuel costs and interest costs rose, nuclear power turned out to be much more expensive than had been anticipated, and then, when the price of electricity went up, the growth of demand for electricity slowed down dramatically, leaving utilities with excess capacity for many years and much higher costs than expected.

Since electricity rates were regulated, these cost increases squeezed margins. When the utilities then filed for increases in rates, there was considerable opposition from customer groups, reflected in the regulatory commissions' resistance to rate increases. Introducing fuel-cost adjustment clauses mitigated the effect of rising fuel costs on utilities' margins but did not eliminate it. By the late 1970s the system was in crisis; many utilities faced difficulties in meeting heavy interest payments and earning a reasonable return on equity.

During the 1950s and 1960s economists had argued that the pricing practices or operating procedures of the utilities were inefficient in several ways, but as long as the industry was working smoothly these complaints were not often acted on. Once the industry was in trouble, regulators and politicians became much more receptive to possible ways of reorganizing the industry to promote efficiency and lower costs. This new concern has led to the changes in organization of the past few years.

To understand the actual and proposed changes taking place in the industry, Joskow says we must understand the structure of the industry and how it is regulated. His paper focuses on the investor-owned utility

sector of the electric power industry, a sector that accounts for 75–80 percent of the power supplied in the United States. Investor-owned utilities have essentially an exclusive franchise to supply electricity to retail customers in a given region. In return for this franchise, the companies must charge the regulated rates and agree to supply reliable service to any customer that wants it. Genuine full-scale competition in the retail supply of electricity is very rare in the United States.

Utilities also sell power to other utilities, or buy power from other utilities. The terms of these wholesale transactions are regulated by the Federal Energy Regulatory Commission (FERC). Such transactions may involve short-term arrangements to cut costs or increase reliability, or they may involve long-term supply contracts when one utility finds it more economical to purchase power long term than to add to its own capacity. The transactions may also involve "wheeling," in which a third party provides transmission service to carry power from one utility to another. Although FERC can regulate the prices and other terms in wholesale contracts, it cannot normally require any company to engage in them. Thus the terms of these wholesale contracts reflect the competitive market value of the power being traded.

Not all customers buy their electricity from the utilities; some companies generate their own. This source of supply was important in the early days of electricity generation, but as the economies of large-scale production grew in importance, this source diminished until by 1978 only 3 percent of generation was taking place outside utility generating facilities.

Federal control over electricity rates is primarily limited to the rates FERC sets for wholesale transactions; most of the regulatory power is held by the states, which set rates and issue certificates for new facilities. A utility must submit a request for a retail rate increase to its state commission, which holds a hearing and then fixes a tariff. The basic principle used in rate-setting is that the average price charged should reflect the average cost of providing the electricity. The prices charged to different groups of customers should be set to reflect the differential costs of supplying these different customers.

A utility's costs for purchased inputs are determined directly, while its capital cost is set to cover depreciation and a "fair" return on investment. Most regulatory bodies set this fair return by constructing a rate base for the company's fixed capital, using the original dollar cost

of the assets less accumulated straight-line depreciation. If applied exactly and continuously, the procedure would provide a return to the utility that would keep its market value of equity equal to the book accounting cost value of its equity. In practice, however, rates are slow to be adjusted up or down, so that utilities will end up earning above the fair return if costs are falling (as in the 1950s) or below the fair return if costs are rising (as in the 1970s). But the adjustment delay does create an incentive to the utilities to cut costs. If they can do so, they can gain a temporary addition to profits before rates are adjusted.

A utility's actual return can also differ from its cost of capital because regulators are not required to include as part of their calculations all the costs a utility incurs. For example, if a utility builds capacity that is considered to be too expensive, the excessive costs associated with this capacity may be excluded from the rate base.

The wholesale transactions regulated by FERC are supposedly also set on the basis of the accounting cost of providing the power, but in recent years this has often not been done in practice. The market for wholesale power has become competitive, with different sellers competing with one another to supply power to utilities. Although the rates have to be justified to FERC on the basis of costs, the agency has been willing to accept negotiated rates in practice. This represents an increase in regulatory discretion that has encouraged efficient wholesale transactions.

After describing the regulatory system, Joskow appraises the rationale for it and the way it has performed in practice. The basic rationale for the regulation of electric power is that the technology makes it much more efficient to have a single monopoly supplier operating on a fairly large scale. The technology, it is argued, creates a "natural monopoly" that is not consistent with a competitive market. Is this a valid argument today?

Most observers agree that the *transmission* and distribution of electricity is indeed a natural monopoly, especially considering the importance of coordination and reliability. It is much less clear that the *generation* of electricity is a natural monopoly. Certainly efficiency is gained from large generating facilities, but that does not necessarily rule out there being several competing companies, each with separate generating facilities that supply a given geographic region. The case for a monopoly provider may reappear, however, when one considers the

combined activity of generation and distribution of electricity. The system as a whole needs a level of cooperation among the generating and distributional facilities that might be difficult to achieve in a fully competitive environment. Thus *a key question for policy in this industry is the extent to which the benefits of competition can be introduced without creating coordination or reliability problems.* The available evidence, says Joskow, shows that *power distribution should remain a regulated monopoly for the foreseeable future, but that the growth of the wholesale market for power can provide a more competitive environment in generation.*

The efficiency with which an industry operates depends not only on whether competition exists, but also on the way in which the regulators set the rules of the game. Studies that have looked at the electric power industry have concluded that companies do not minimize their costs. Some studies find that utilities overinvest in capital, and some find that companies are wasteful or inefficient. But since estimates of the importance of these inefficiencies vary a good deal from study to study, it is hard to draw firm conclusions. The estimation problem arises because of the similarity in regulatory procedures among states, even for those that look rather different on paper. We simply do not observe the effect of radically different systems of regulation.

One aspect of the regulatory environment that may have affected the efficiency of operations is the lag in the adjustment of rates. As noted earlier, the regulatory lag has effects that have operated rather differently over time. There is some evidence that the lag encouraged cost cutting in the 1950s and 1960s (when price declines lagged behind cost declines) and that the lag has become so punitive in recent years (as rates have failed to reflect cost increases) that utilities have cut back too far on the construction of new facilities.

Joskow next discusses changes that are under way in the industry or are likely to continue in the future. Independent cogeneration and small power plants, he says, should contribute a share of new generating requirements over the next few years. And some independent production facilities are being considered that would supply the wholesale power market. These generators would be freed from many regulations that apply to the regulated facilities. At present, these independent suppliers do not exist, but there is clearly a growing environment that favors independent sources of supply which are not subject to the traditional

forms of economic regulation. The wholesale trade in power, including power purchased from Canada, has increased more rapidly since 1973 than total electricity consumption, and this trend is likely to continue.

One reason for the increase in independent sources of power is that the regulated utilities are unwilling to build major new facilities themselves. The utilities learned from the post-1973 period that they would not necessarily recover the cost and a normal return from such investments. Today the expected return on investments in new generating plants is perceived to be below the cost of capital. There is concern that the gap created by this unwillingness to invest may not be filled by the wholesale power market, or may be filled only with higher electricity rates or with a less reliable supply.

The most important reason for the development of an independent power sector is that in November 1978 Congress passed the Public Utility Regulatory Policy Act (PURPA) as the Carter administration's response to the energy crisis. This act has already strongly influenced the industry and will continue to be important in determining future trends. Under the act utilities must purchase power from companies that install cogeneration equipment and from small independent producers that use renewable energy sources. The rates and conditions under which the utilities must buy the power are set by the states, using general guidelines issued by FERC, with the rates that the utilities should pay being determined by what it would cost them to provide power from alternative sources—either their own facilities or those in the wholesale power market. The evidence suggests that the act has increased the amount of power being supplied by nonutility sources. But the full impact of the act has been masked by the long-term decline in the amount of power generated by companies for their own use, a trend that dominated the general picture until 1983. But since 1983 the fraction of power coming from nonutility sources has been rising and seems clearly attributable to the act.

Independent generation has been forthcoming in the market and can potentially increase competition in the industry, but is it appropriate for the rates to be regulated? And, if so, is the current rate-setting procedure the correct one? Joskow argues for some regulation of rates and purchase obligations. If the rates for purchased power were left completely to negotiation between the utilities and the suppliers, some regulated utilities might favor their own facilities over those of the independent suppliers even if the latter were cost efficient.

Requiring utilities to buy available power at the rate the power would cost them using their own facilities or at the rate they would pay in the wholesale market—the avoided-cost principle—has both a strength and a weakness. It provides an efficient pricing system in principle, since it reflects the opportunity cost of power. But it is hard to implement in practice, because the true "avoided cost" is hard to measure. Is a utility's capacity taken as given? Are there differences in the reliability of the different sources? And so on. A serious problem in practice arose in California when prices were set too high, too much capacity was supplied, and prices did not change flexibly when fuel prices or other economic conditions changed.

Joskow favors a system in which independent suppliers have to bid competitively against alternative sources of supply in the wholesale market. The utilities would specify the terms under which the power would be supplied by the outside sources, just as companies in other industries invite bids from subcontractors. Utilities would be allowed to construct new generating facilities themselves only if they could convince regulators that the cost of the electricity would be competitive with the cost of additional outside supply, given the same reliability and other conditions that were being required from outside suppliers.

Some states have already experimented with competitive bidding of the kind just described. So far, all utilities that have introduced competitive bidding have found abundant supplies available on attractive terms. In Massachusetts the utilities are not planning to build new facilities themselves despite the relatively tight capacity in the state. They are looking to the wholesale market for additional supplies. But independent suppliers will build generating facilities only when given long-term supply contracts, and these contracts can create their own problems of price adjustments, as the California example illustrates. Finding the form for long-term contracts that will maintain incentives, as well as provide the security necessary to allow for large fixed investments, remains a challenge. It may not be possible to support construction of very large generating projects using this approach.

In concluding, Joskow is fairly optimistic about current developments in the electric power industry. PURPA was passed to deal with the energy crisis, but it has opened the door to a more competitive generation market. The potential gains from such a market are large, though not without some significant dangers. First, the reliability of supply may deteriorate. So far this has not happened, but reliability needs to be

carefully monitored. Second, the wholesale power market may not remain fully competitive and thereby dilute the benefits of the market system to consumers. To have meaningful competition, several suppliers must be in a position to bid on supply contracts. So far, competitive bidding has taken place with several competitors bidding, but there is concern about maintaining access to the transmission system in order to sustain this kind of competition. Some creative solutions to this problem of maintaining access are being explored, and these solutions need to be developed.

Third, there are potential barriers to the efficient choice of independent suppliers of generation arising from the way in which states set retail electricity rates. Using a cost-of-service approach does not encourage the utilities to use the lowest cost sources of supply. Joskow discusses alternatives that have been proposed in New Jersey and Massachusetts. These alternatives have promise, but implementing them in practice may be difficult.

As Joskow points out, the regulatory cat is out of the bag, and he sees no reason to try to stuff it back in. Many legitimate questions remain, however, about the way to handle the current industry, with its hybrid mixture of regulation and competition.

Joskow's discussants commented on the depth and thoroughness of his paper. Douglas Bohi said Joskow might have laid more stress on the tension between the traditional goals of regulation and the requirements for an actively competitive market. He also felt that access to transmission lines by small suppliers was important in creating a competitive market environment. Frank Gollop would have welcomed more discussion of ways to combine the benefits of a competitive wholesale market with rate regulation by FERC, and in his comments he looked at some alternative proposals for achieving that.

Katz and Summers on Wage Differentials and Industrial Policy

There has been much public concern in recent years about the alleged decline in the availability of "good jobs" and whether policies should be instituted to increase their availability. But, as Katz and Summers note, this concept of a good job does not fit with standard competitive economic analysis. According to that analysis, workers are free to move

to the best job opportunities they themselves can find, so that the differences in wages or working conditions we observe reflect differences in people's skills or abilities. There are no good jobs or bad jobs, as such, only more skilled people and less skilled people.

Katz and Summers argue cogently that in fact good jobs and bad jobs do exist. Many workers receive higher wages than others simply because they work in specific industries. And, say the authors, this may justify an industrial policy on grounds of increased employment in the high-wage industries. Policies that increase the number of good jobs need not be justified on distributional grounds but can actually improve the efficiency of the labor market. They caution, however, that any economic case for activist policy must be tempered by the recognition of the formidable difficulties involved in successfully implementing structural policies.

Even though the authors focus mainly on wage differentials, they turn first to a discussion of the returns to capital, emphasized in the antitrust and strategic trade policy literature. It is assumed, for example, that specific trade policies increase the rate of profit of the companies in the industries being affected, or, in the antitrust case, that the absence of adequate competition in an industry leads to excess profits in that industry. The authors examine these possibilities directly by looking at the rates of return to capital in twenty manufacturing industries. They find that the owners of capital in the United States do not earn large monopoly profits. The average rate of return to capital is a modest 6 percent, and on average the valuation of the companies in these industries by financial markets is below the replacement cost of the companies' physical assets. The return to capital, they also point out, is simply not a large share of the total value of output, so that even if variations in rates of profit are in fact larger than they believe, then the resulting inefficiency will not greatly affect the allocation of capital resources.

Having disposed of profit rate differentials as an important source of economic inefficiency, Katz and Summers turn to their main task of analyzing wage differentials. They use the Current Population Surveys for data on wages, industry of employment, personal characteristics, and so on, combining all the monthly surveys for 1984. The CPS is the standard source for the monthly unemployment reports and other labor force estimates. In table 2 the authors report the proportionate difference between the wage earned by workers in specific industries and the

average for all industries, and find substantial differentials across industries. When the personal characteristics of the workers in the sample are controlled for, these differentials fall, showing that one reason some industries pay higher wages than others is that the people they employ are not identical. But even after controlling for these individual characteristics, the industry wage differences remain large and are distributed across industries in a similar way. This means, the authors say, that *people with the same personal attributes (such as education or occupation or sex) earn substantially different wages depending upon the industry they work in.*

Katz and Summers examine several possible explanations for the observed industry wage differentials. First, these differentials could be offset by fringe benefits or other nonwage forms of compensation. The authors find this is not true. Adding in an adjustment for fringe benefits shows that high-wage industries actually pay more for such benefits, not less.

Second, the wage differentials could be associated with differences in occupation by industry that were not controlled for by the rather crude occupational groupings available in the authors' data source. To test this possibility, Katz and Summers look at a subgroup of people within their basic sample, consisting of workers who had one of four more narrowly defined occupational characteristics—managers, secretaries, janitors, and laborers. Even within these occupational groupings they find substantial wage differentials: *secretaries or janitors earn different wages depending on which industry they are with.* The authors also cite data from the Boston area wage survey to support this idea that the same occupation can earn a different return in a different industry.

Third, the industry wage differentials could be the result of unionization. Many people would argue that the reason janitors earn more in some industries than in others is that one janitor's job is unionized and another's is not. Katz and Summers check this idea out by rerunning their analysis using only the people who do not belong to unions. They find that the industry wage differentials are somewhat differently distributed across industries for the nonunion workers, but the basic magnitude of the differentials is the same in the two groups. Unionization does not seem to be the reason for the pattern they have observed.

Fourth, the wage differentials could arise from different working conditions in the different industries. Some jobs are dirtier or riskier or

more physically demanding than others and so require a higher rate of pay. The authors argue that such "compensating" wage differentials do not explain the industry differences they have observed. They refer to an earlier study by Krueger and Summers on the impact of working conditions which found that industries which paid higher wages had in fact somewhat better working conditions than those which paid lower wages. The observed differentials were certainly not compensating for adverse working conditions. The authors also note that the industry wage differences apply across the occupations studied earlier. They wonder if differences in working conditions can really explain why secretaries in some industries earn more than secretaries in others.

Fifth, the wage differentials could result from large unobserved differences among people in the data sample. Workers who look the same in the Current Population Survey data—who have, for example, the same occupation or education—may be quite different in their actual productivities and hence be paid very different wages. Even though the authors controlled for individual differences in their sample, the information on these differences was not all that good. In terms of the *industry* differences in wages, the hypothesis must be that high-wage industries systematically choose people who are higher in skill or ability than the average within any occupation or educational group.

The authors acknowledge the importance of unobserved differences among people, an issue stressed by Robert Topel in his discussion of the paper in the meeting. But they argue that even though this hypothesis has some validity, unobserved differences among workers cannot be the main reason for the industry wage differentials. They point out that when they use the information in the data set (the observed characteristics) as controls in assessing the industry wage differences, the most important factors are such things as occupation, region, and sex. Low-wage industries, for example, disproportionately hire female operatives and are located in the South. Taking account of education or experience does not make as much difference, even though education is very important in determining any specific person's wage. High-wage industries do not use educated workers disproportionately, so the fact that educated people earn more money does not explain why some industries pay higher wages than others. This suggests to the authors that high-wage industries are not hiring a disproportionate fraction of high-ability workers, because of the known or *observed* characteristics of the

individuals; there is therefore no reason to argue that these same industries are also hiring a disproportionate fraction of high-ability workers because of the factors *not being observed.*

Having concluded that all these explanations of the industry wage differentials do not in fact explain them, Katz and Summers look for some additional evidence to support their idea that the differentials are associated with the industry in which employees work. The place they look is longitudinal evidence: what happens to people who change jobs? They review four studies that have examined this kind of data. Three of the studies found that when workers move from one industry to another, their wages change by an amount similar to the amount of the industry differences in wages. Workers who move to a high-wage industry from a low-wage one receive, on average, an increase in wages that is 60 to 100 percent of the wage differential between the two industries. One of the four studies, by Murphy and Topel, found that industry switchers receive only about one-third of the industry wage differential. The reason for this lower estimate, Katz and Summers argue, is that Murphy and Topel compare annual earnings from one year to the next. This could mean that the earnings figures may not match exactly with the jobs the people held if someone changed jobs in the middle of a year. The authors conclude, therefore, that the longitudinal evidence supports their finding that the industry of employment matters.

Katz and Summers then turn to a second piece of evidence that supports their conclusion. If some industries provide good jobs and some bad, one would expect that workers who have jobs in high-wage industries would be unlikely to quit those jobs. The authors find that indeed there is a clear negative relation between the wage paid by an industry and the quit rate and that this relation is associated with the industry premium and not with other observed characteristics of the workers. A wage premium of 20 percent, they find, reduces the quit rate by about two-thirds.

The authors next examine the question why some industries pay more than others even if they are not compelled to do so either by unions or by the need to hire high-quality workers. One idea prominent in the literature is the so-called efficiency-wage hypothesis. Some companies pay more in order to increase the motivation of their workers—to get them to work harder—and to discourage quits, which may be costly to the companies. Katz and Summers judge that this idea is important but does not really explain the wage differentials they have found. The

efficiency-wage models have been criticized on the grounds that companies could find cheaper ways to discourage quits or to encourage work effort.

The prime reason for the existence of the wage premiums, the authors argue, is that labor markets are not competitive and that when a particular industry has a technology or a market situation which allows the companies in it to earn more than the normal competitive return from their production activities, and which makes it very costly if the companies are subject to a work stoppage, the companies and workers in the industry share the available excess return (or rent, in economic terms) that has been generated. Companies give some of the rents they create to the workers, whether there is a union or not. Even in the absence of formal unions, the authors argue, labor has some power over wages. For one thing, the performance of workers depends on whether they believe they are being fairly treated. If an industry is very profitable, workers expect their fair share, and performance in a company will suffer from either individual or collective actions if that is not given. So companies may buy peace with their workers. In fact, unions may not be the main cause of wage premiums; often unions may simply move into situations where wages have already been set above the average.

To develop their explanation for wage differentials, Katz and Summers look at the characteristics of high-wage industries. These industries, they find, are capital intensive and R&D intensive, earn above-average rates of return, and have above-average financial market valuations. The characteristics of these industries suggest that they have rents which can be shared with the workers. Capital- and R&D-intensive industries will lose heavily if their assets are kept idle.

The authors next discuss the implications of their conclusions for policy, particularly trade policy. The basic argument is this: when noncompetitive wage differentials exist, the level of employment in a high-wage industry is too low and the level of capital investment may be inefficient also, and the existence of these differentials should be considered when policy decisions are being made. For example, when choosing between policies to protect imports or encourage exports, the higher wages paid in export industries should be factored into the decision. And the case for policies to encourage capital investment or R&D should recognize that workers capture some of the returns to these expenditures.

The authors use two case studies to illustrate the effect of policies to favor good-job industries: the European Airbus and the intervention in

the world trade in automobiles. In their study of the development of the A300 jet, Baldwin and Krugman found that the subsidies paid to this project greatly affected the allocation of aircraft production between the United States and Europe and also lowered the price of this type of jet aircraft considerably by reducing the monopoly power of the U.S. industry. According to the study, the gain to European consumers was worth $1.47 billion, but the cost to taxpayers was about $1.42 billion, so the net effect was roughly a wash. However, say Katz and Summers, this calculation ignored the favorable impact on the European economy of the increase in high-wage employment generated by the Airbus A300. They estimate that the program had in fact a clear net positive effect, given that the industry pays a positive wage differential.

In his study of the policy to encourage domestic production of automobiles Dixit did take the wage and employment effects into account. And by doing so, he showed that the United States clearly benefited from the policy of encouraging the production of autos in U.S. facilities.

Some writers today believe the way to promote good jobs in the American economy is to restrict foreign trade. Katz and Summers say their analysis does not provide a case for protectionism. In the long run, they point out, the United States will be able to reduce its trade deficit, and indeed will have to run a surplus to pay interest on its foreign debts. The key question for trade policy is whether we are better off with a high or a low level of *both* exports and imports. To address this question, the authors present data showing that U.S. export industries generate high-wage jobs and that all the industries facing very severe foreign competition are low-wage industries except for automobiles. After being adjusted for skill differences, wages in export-intensive industries are above average and import-intensive industries are below average. Therefore, *policies that promote the expansion of international trade will increase the number of good jobs and reduce the number of bad jobs.*

Katz and Summers conclude by summarizing the essential elements of their argument. Some industries are good-job and some are bad-job industries, even after allowing for differences in the skills and education of the workers employed in them. This means that when policy decisions are being made, the fact that one policy encourages employment in high-wage industries is a valid point in its favor.

Robert Topel has studied theories of wage determination in his own

work, and, though he thought Katz and Summers presented their paper in a balanced way, he was very critical of its conclusions. He believes that industry differences can be explained by differences in the quality of workers and in working conditions. The strong correlation between the industry wage differentials and the effects of personal characteristics (a result stressed by Charles Schultze in the meeting) suggests, he said, that high-wage industries are paying for high-quality workers. Using the observed differences available from the CPS is bound to be inadequate: consider how much individual heterogeneity there is in any roomful of people with the same level of education or observed attributes. Topel also believes that some high-wage industries, such as coal mining, are paying more because of poor working conditions.

Robert Hall and other participants in the meeting did find the evidence persuasive that industry differentials remain after correcting for worker quality and other factors. But Hall questioned the implications to be drawn from these differentials. He noted that large unexplained differences in wages exist along several dimensions. Since wages are so much lower in some states than others, for example, would Katz and Summers suggest subsidizing employment along geographic lines? The authors responded that regional wage differentials are different because people do not have to wait in line in order to change regions. However, Hall was not convinced that the industry wage differences demonstrate disequilibrium in the labor market that needs to be corrected. He did not see evidence that high-wage industries have the kind of excess supply of labor the Katz-Summers results would suggest.

Some participants expressed concern about the potential response of workers to a policy of subsidizing high-wage industries. If the high wages are the result of workers acting coercively to raise wages, with or without a union, then a subsidy will encourage other workers to act in the same way. Katz and Summers acknowledged this problem and do not propose a general policy of subsidizing high-wage industries.

Griliches on Trends in Patenting

Zvi Griliches' paper starts with the fact that the number of patents granted in the United States and in most other major industrial countries has declined. Patents granted to U.S. corporations peaked in about 1970,

and the rate of patenting has not really recovered since then. This slowdown in the rate of patenting has been accompanied by a worldwide slowdown in productivity growth, suggesting that the pace of technological change may have declined.

The data may also be interpreted in another way. Since Japan is the one major country that has not suffered a decline in its rate of patenting, the United States and perhaps the European countries may have lost their competitive edge in technology development. The decline in U.S. patenting may be symptomatic of a loss of competitiveness, not of a slowing of the opportunities for advance.

Griliches argues that it is a mistake to draw conclusions from the decline in the rate of patenting without looking hard at what the numbers mean. Patent data are subject to serious misinterpretation, he maintains. For one thing, the number of patent *grants* does not always follow the number of patent *applications*. A change in the number of patents granted could reflect a change in the standards used by the Patent Office in deciding whether to approve a patent, or a change in the resources that the Patent Office has available to consider and process patents. For another thing, the number of applications could change in response not only to the underlying flow of inventions but also to changes in the value that companies themselves place on patents. Companies may decide that it is not worth patenting certain types of inventions, or even that patenting is to be avoided because of the disclosure an application implies.

Griliches looks first at the total number of patents applied for and the number granted in the United States, including those applied for by foreigners. By extrapolation, he shows the rate at which patents would have been granted if the U.S. Patent Office had maintained a constant (65 percent) approval rate and had a constant period of delay between approvals and applications. He finds that the rate of patent grants would have been basically constant over the 1970s if these conditions had been true. Therefore, some of the changes that took place in the rate of patenting were the result of changes in the approval process. In the late 1960s, says Griliches, half the applications took more than three years before a patent was approved. Beginning in 1971 the Patent Office made an effort to reduce the patent backlog and this brought it down. Then the patent office ran into budgetary problems in the 1980s, and the backlog of applications began to grow again.

Griliches approaches this same issue more formally by estimating an equation that explains the rate at which patents are granted, based on the number of patent examiners in the Patent Office and the number of patent applications. He finds that the number of patents approved depends crucially on the number of examiners. In fact, the main effect of the number of applications on the number of patents approved is indirect: when there are more applications, the Patent Office hires more examiners, thereby increasing the number of patents approved.

Griliches illustrates his statistical findings in figure 3 of the paper, which shows that the short-run fluctuations in the total number of patents granted in the United States depend much more on the number of patent examiners than on the number of patent applications. And Griliches notes, in particular, that *the decline in the rate of patents granted in the 1970s was a "bureaucratic mirage" more than an indication of overall technological decline.*

Griliches turns next to the important issue of domestic versus foreign applications for U.S. patents. He argues at the outset that in assessing the pace of technological change, one should use the *total* number of patents (if any patent data are relevant) rather than the number of domestic patents. The total flow of new patents in the United States is a better indicator of the technological opportunities becoming available to the U.S. economy. The level of domestic patenting may be more relevant to the competitiveness of the U.S. economy or to the rate of return to U.S. research and development.

In figure 4 Griliches plots the number of domestic patent applications and two other economic indicators—GNP and the level of R&D spending. The figure shows that the number of patent applications relative to GNP reached a peak as early as the 1920s and has not risen to the same level since. The number of patents relative to R&D spending has been declining rather consistently. In the period after World War II, domestic patent applications remained fairly constant, whereas GNP and R&D spending both increased—with R&D spending rising faster than GNP.

This decline in patenting relative to GNP or the decline relative to the amount of R&D spending was noted by Jacob Schmookler many years ago. He suggested three reasons for it: problems in the Patent Office (of the kind just considered); changes in the mix of industries in the economy away from those that rely heavily on patenting and toward those that rely on secrecy or other methods of protection; and changes in the

judicial climate that made patents harder to enforce and hence less valuable. Griliches adds another reason to these three: the rise in the cost of applying for a patent (in the form of the number of skilled labor hours required to prepare an application) may have discouraged independent inventors and even some corporations from applying for patents.

One of these four reasons for the slowing in domestic patenting can be examined empirically; namely, the effect of changes in the industry mix. The traditionally high-patent industries have been drugs and chemicals, where patent protection has been most valuable. The electronics industry has been less prone to patent because of the fears of disclosure and because of the rapidly changing technology. If innovation had shifted from industries like chemicals and toward electronics, the rate of patenting, on average, would have been lowered, even with no change in the underlying pace of innovation. Griliches finds that this plausible idea does not seem to work. He adjusts for changes in the mix of industries by using data on the amount of R&D by industry, and finds that even if he uses a mix-adjusted patent series, it differs very little from the regular series. The propensities to patent in the different industries are not very different, and the mix of industry R&D has not changed much. The decline in patenting by U.S. industries has been pretty much across the board.

Griliches turns next to the relation between patents and R&D spending. In the modern economy, patents primarily grow out of organized R&D, and the overall trend in R&D spending has been upward. However, a slowing of the growth in R&D spending in the 1970s probably contributed to the decline in patenting in that period. Thus part of the decline in domestic U.S. patenting in recent years can be "explained" by weakness in R&D spending. This explanation begs the question, however, of why R&D spending slowed down. If spending slowed because companies judged that the opportunities for technological advance were weak, the fundamental cause of the weakness in patenting would still be the lack of opportunities.

Griliches disagrees, however, and recalls one of Schmookler's main ideas about patenting. Schmookler said that innovation is driven by changes in demand. As people increase their purchases of a product, this stimulates the search for new technology in that area. Griliches argues that the weakness in R&D spending in the 1970s is attributable to the climate of recession and economic uncertainty triggered by the energy crisis.

To look both at the short-run fluctuations in patents and R&D and at the more fundamental question of the long-run decline in the ratio of patents to R&D spending, Griliches reports estimates of an equation that relates the number of patents applied for by domestic companies to the R&D they perform and to other variables. The equation confirms the link between patents and company R&D, so that indeed the slowing of R&D growth in the 1970s had a negative effect on patenting. In addition, Griliches finds that increases in the size of U.S. defense spending have a large and significant negative effect on patenting. A 10 percent increase in defense spending leads to a 5 percent decrease in patenting. This finding is consistent, says Griliches, with the fact either that defense spending pulls resources away from inventive activity or that it channels such activity into areas that do not lead to patents. Griliches also finds that basic research in universities contributes positively to domestic U.S. patenting and that the rising cost of R&D has been a negative influence on patenting. These results, he notes, are rather tentative because his explanatory variables are correlated among themselves. For example, the variable that picks up the effect of company-funded R&D on patenting moves very closely with time and the path of GNP. However, he reviews some earlier statistical work he and others have done that tests some of these relationships using data on individual companies or industries. Such microeconomic data have confirmed the link between the amount companies spend on R&D and the number of patents they apply for, though they have suggested a puzzlingly short lag between the R&D spending and the patent application.

To summarize the results of his investigation of the causes of the changes in U.S. patent applications, Griliches says the following. From 1970 through 1979 the rate of patenting declined about 2 percent a year, about a third of which was due to the growth in defense spending. Over the same period, company R&D and basic research grew 1–2 percent a year, which should have canceled out the negative effects of the defense spending. Thus the fact that U.S. patenting declined over the period remains unexplained. Furthermore, over the entire 1954–87 period patenting did not increase, though R&D and basic research spending did. This too remains an unexplained puzzle.

Griliches then considers the implications of the trends in patenting, in particular whether the lack of increase in domestic U.S. patenting could partly explain the slowing of productivity growth. Drawing inferences from patent data about the slowing of growth is difficult. Several

studies have found that most of the value of patents is concentrated in a few very valuable patents. The majority of patents are worth very little. Thus any decline in the flow of patents could have occurred with very little change in the flow of major patents or perhaps with a dramatic drop in the number of key patents. If anything, recent estimates made of the value of patents (see the Pakes and Simpson paper) have shown that the average value per patent has grown. Therefore, the productivity effects of the absence of growth in the numbers of patents have probably been ameliorated by the increase in the contribution of each patented invention.

Griliches argues next that the contribution of patented inventions to total productivity growth is not large to begin with. Only about a quarter of the increase in total factor productivity over time, he says, is the result of such inventions. So even if there had been a slowing of the pace of invention reflected in the patent data, the effect on productivity growth would not have been large.

Finally, Griliches turns to the implications of his findings for the incentives for R&D spending. Because most observers find that the payoff to innovative effort in any narrowly defined area of technology will at some point run into diminishing returns, one frequently hears concern that the United States is running out of opportunities for innovation. But the exhaustion of opportunities in one narrow area does not mean there will be diminishing returns to R&D overall. As one pool of ideas is fished out, another is opened up, and the return to innovative effort is maintained. Griliches has shown that the number of patents per dollar of R&D spending has declined, so that unless the value of inventive activity per patent has increased, there must have been a decline in the return to R&D. But the thrust of his findings in this study is that the relation between patents and inventive activity has changed. Thus Griliches does not see grounds for pessimism about the future trend in R&D spending. The returns per dollar of spending should remain high even if the number of patents per dollar of spending declines, as it has for more than fifty years.

Mike Scherer, in his comments on Griliches' paper, noted that patent applications became much longer, as measured by number of pages, during the first half of the century, which suggests that each patent is much more complicated than it used to be. He also reported that in 1974 the average company expenditure on R&D per patent granted was almost

$600,000 and the average patent contained almost ten separate claims of invention, figures certainly not inconsistent with the idea that patents are becoming more complex. Using data based on entrants to a competition for the most significant technical advance, Scherer said he found little evidence that the cost per *invention* had been rising. Finally, Scherer warned against looking for too close a connection between patents and productivity growth.

William Nordhaus took a position somewhat different from that of both Griliches and Scherer. He suggested that the slowdown in productivity growth may have been caused by a fundamental decline in the rate of invention and that the evidence in Griliches' paper did not justify the opposite conclusion. The slowing of the growth of patents since the 1930s, he observed, would support the hypothesis that opportunities are being depleted. And he disagreed with Griliches' statement that it is the worldwide flow of patents which matters. Rather, Nordhaus said, what matters is the number of U.S. patents that are available to U.S. producers. Finally, he argued that since the values of patents are known to be very skewed, variations in productivity growth over time could arise from the loss of a relatively few key inventions, and this loss might show up in a fairly small decline in the number of patents.

Pakes and Simpson on the Value of Patents

In the last paper of this issue, Ariel Pakes and Margaret Simpson note first that a patent provides specific protection for an inventor: it is the principal mechanism by which rights over intellectual property are provided in market economies. This protection has a value to the holder of the patent, but since patent rights are seldom bought and sold outright, it is hard to determine exactly what that protection is worth. In a series of papers, Ariel Pakes and his coauthors have developed a methodology for valuing patents, using information on the fraction of patents that are renewed at various points over the life of the patent.

In their paper here Pakes and Simpson provide an overview and assessment of this methodology and an application and extension of it, using new data sources. The authors stress at the outset of their analysis that there is an important difference between the value of an invention and the value of the patent that protects it. Companies have many ways

of getting a return from their inventions other than through patents (see the paper by Richard C. Levin and others in *BPEA 3:1987*).

The value of a patent is an important number, however, because it tells us specifically about the legal protection that society provides to inventors. If policymakers are looking for ways to affect the incentives for invention, the patent system is an important policy instrument. For example, the provisions of the patent system were changed a few years back to provide more protection for companies that develop new drugs.

Pakes and Simpson characterize the inventive process as follows. An inventor has an idea that he or she believes can be patented and that may earn some commercial return. If the patent is granted, it gives the right to the holder to use the patented idea over a specific period. At the end of the period the patent must be renewed if it is to remain in force, and a renewal fee must be paid to achieve this (almost all the main industrial countries require renewal fees). At the time of renewal the patent holder must weigh the cost of the renewal against the value of using the patented idea until the next patent renewal point, plus the option value of being able to renew the patent for a subsequent period. At each renewal point it may not be worthwhile making the renewal, and the patent may simply lapse.

In general, the patent holder will not know exactly what the returns to a patent will be. Pakes and Simpson assume that the estimates patent holders form about the probable values of their patents in future years depend on the age of the patents and the returns they have been earning in the current year.

Perhaps the key issue for the analysis is to assess how much information the data set contains: how can information on renewals be translated into information on the values of the patents themselves? The authors' first result confirms the intuition that when one type of patent is more valuable than others, then a larger fraction of the more valuable group will be renewed at each renewal point, regardless of the specific schedule of renewal fees or the rate of discount applied to future returns. Since data are available on the fractions renewed at each point, they can then be used to test for differences in the values of the patents in the different groups.

To see how to go beyond this result, imagine a controlled experiment that varied the initial fee for acquiring the patent and the subsequent fees for renewing the patent from high to low. Doing so would change the

proportions of the patents that were renewed as the fee schedules were changed. And, as the authors demonstrate, with enough variability in these fees, one could in principle trace out the full distribution of the values in a group of patents. One could just see which patents dropped out at each point as the fee schedule was raised.

Any actual data set will not correspond to the controlled experiment, but in practice countries impose a series of fees over time, so that some patents drop out at each renewal point. This separates the valuable patents, as these are kept in force for many years, from the less valuable patents, as these drop out early. Roughly speaking, there is a correspondence between the value of a patent and the length of time it is kept in force.

The method is limited because having fairly small fixed fees imposed at certain renewal points is not the same as being able to vary the level of fees at will, in the ideal way described in the controlled experiment. Specifically, a problem would arise if the renewal fees were trivially small, for then almost all the patents would be renewed. Knowing which patents were renewed and which were not would convey very little information about the distribution of patent values. Patents that in fact differed in value would all get lumped together.

A different problem would arise if the renewal fees were always very large, for then most of the patents would not be renewed. One could therefore not distinguish the somewhat valuable patents from the patents with very little value. The wheat and the chaff would all drop out together.

In practice, most renewal fee schedules rise to between $500 and $1,000 (up to $2,000 for Germany). This is sufficient to induce most patents to drop out at some point before their statutory limit. About 95 percent of the patents in France and Germany are eventually dropped, at different points over their potential lifetimes. So most of the "students" get different grades on the renewal exam, and we can tell them apart. But at the same time these fees are trivial relative to the value of the most important patents. The value of patents is not evenly distributed around the average value; much of the value of all patents is contained in a small minority of very valuable patents for which the renewal fees are very small. So a group of very bright students ace the exam, and we cannot tell them apart. The renewal fees allow us to learn a lot about most of the patents but not enough about the very valuable patents.

Still, it is possible to learn something about even the very valuable patents from renewal data, because the ultimate value of a patent is not known to the holder at an early stage of the patent's life. The decision to renew depends upon how likely the holder thinks it is that a given patent will turn out to be one of the very valuable patents.

The authors review some of what has been learned from the renewal literature to date. In the work by Pakes and by Schankerman and Pakes it was found that patents are applied for at a point in the inventive process when there is still much uncertainty about the value of the invention and hence of the patent. Companies apply for patents at a fairly early stage in the process.

These studies found that the average value of patent applications in the United Kingdom and France was about $7,000. In Germany, where the ratio of approvals to applications is low, the average value of patents granted was $16,200. Given the total flow of patents, these figures imply that *the value of the current returns to all patents is only 11 to 16 percent of the value of current R&D spending in each of the countries*. Patents clearly are not the main mechanism providing a return to R&D. But they are still an important incentive to R&D relative to other public policy options, such as the R&D tax credit.

An important additional finding has already been alluded to: the distribution around the average value is very skewed, with 1 percent of patents having values in excess of $85,000 in France and the United Kingdom and a little higher in Germany. *More than half of the value of all patents accrues to between 5 and 10 percent of the patents*.

Pakes and Simpson next ask if renewal data allow the construction of a measure of the flow of patented ideas that is better than simply using the number of patents per year. They attach weights to the patents in different groups depending on the average value of the patents in the group, with the values estimated from the renewal data. That work, which covered patents from the 1950s, 1960s, and early 1970s, found that the patents which ended up being dropped before age five had almost no value. These represented a third of the French patents and about 5 percent of the German patents. In the French data even the patents discarded before age eleven, two-thirds of the patents, had little value, as did the German patents discarded before age eight, 27 percent of the total.

After these cutoff ages of eight and eleven, the weights increased by about 35 percent a year up to the final year. The patents that are renewed

throughout their possible lives require a weight of close to two and a half times the weight given to the patents that were renewed up to, but not including, the final year. These are the patents that the authors would like to know more about.

Griliches' paper showed us that patent data may not accurately reveal trends in the flow of new inventions. Pakes and Simpson find that some of the year-to-year fluctuation in the number of patents is eliminated if value weights are used to make an adjustment for quality. Years with many patents are often years with many low-value patents. Even more important is the result that Pakes and Schankerman originally reported for France and Germany, and that is confirmed here by Pakes and Simpson for Norway and Finland: *the average value of a patent was increasing in those countries with decreasing numbers of patents from the mid-1960s until the end of the sample period in the mid-1970s*. Since the number of patent applications was falling in almost all of the countries at that time, apparently *falling quantity was being offset by rising quality*. Griliches also suggested this result in his discussion of trends in patenting, but it runs counter to Nordhaus's view.

Up to this point Pakes and Simpson have been mostly reviewing the methodology of patent renewal models and what has been learned from them. They turn now to some new analysis using data from Norway and Finland. These are small countries, but they provide a unique opportunity because the authors were able to separate out the patents in them by industry and by the country of the patent applicants. Despite their size, the countries provide a sample that tells us how U.S., Japanese, and other major country patents differ from one another, by industry. These data were located by Pakes and Simpson by sending out a survey to over 100 patent offices asking about the data that are collected by each of them and their availability for research. This is the first time that these data have been used for renewal analysis, and the paper provides information on how other researchers can use them.

The analysis looks first at whether countries differ in their propensity to renew the patents they hold in Norway and Finland, without regard to differences by industry. The answer is yes there are differences, with the Japanese being more likely to renew than the United States or continental Europe, and Britain being less likely to renew. In both Norway and Finland, the home countries' own patents were the least likely to be renewed.

When the patents are disaggregated by industry as well as by country,

the country effects become largely insignificant. Thus the Japanese renew more because their patents are in industries that typically renew more. Even so, the home country patents are still less likely to be renewed, even after allowing for industry differences.

The pattern revealed by the industry breakdown is as follows. The industries whose patents are more likely to be renewed and, therefore, have more valuable patents are chemical patents, especially drugs, followed by mechanical, heavy industry, and finally low-tech industry patents. The high value per patent in the chemical industry is familiar, but note that food and kindred products, and lumber, wood, and paper, are classified with the chemical patents in the above breakdown, as part of the group that has the patents with the highest value.

In commenting on the paper, Edwin Mansfield stressed that the value of a patent is not the same as the value of the invention being patented. He has found that in pharmaceuticals, oil, and machinery about 80 percent of inventions are patented, whereas in primary metals and autos the percentage is only 60 percent. Thus patent counts may be misleading measures of inventive output, said Mansfield, and he wondered how much the Pakes and Simpson results tell us how about the current policy debate over intellectual property protection. Richard Levin stressed, however, that patents are a key policy instrument.

Kenneth Judd said the work done by Pakes and Simpson and other coauthors on valuing patents with renewal data was very interesting, but he expressed some concern about the validity of the assumptions underlying the analysis. Pakes and Simpson assume that the value of a patent is determined exogenously. But, said Judd, patents have to be viewed in terms of the overall strategy of the companies in an industry. If one company has a technological lead in a field, it may decide not to patent its inventions, but may let it be known that if other companies try to enter the field, then it will certainly file for patents that will exclude these other companies. Thus the fact that a company has the option to file a patent has a value to it, even though there is no observed patent. Differences in strategic behavior may be a reason for differences in the propensity to patent of different industries. Judd also commented on the fact that the cost of renewing patents includes some administrative costs that may be as large as the renewal fee. The Pakes and Simpson results may be sensitive to the assumed size of the fees.

Mike Scherer expressed some concern about the ability of renewal

data to value patents in industries where the technology is moving very rapidly. Some patents are very valuable, but only for a short period of time. These patents may be dropped after only a few years, making them look like patents that were dropped because the underlying inventions turned out to have little value.

MARTIN NEIL BAILY