Banking in the 21st Century

Based on a speech presented by President Santomero at the New Jersey Bankers Association Annual Convention, Aventura, Florida, March 27, 2004

BY ANTHONY M. SANTOMERO

Although U.S. consumers are catching up with their counterparts in the rest of the world in their use of electronic payment methods, paper checks remain a popular way to make payments here. How is the Federal Reserve supporting an orderly evolution of the payment system as consumers shift away from paper? One way is the Fed's advocacy of legislation such as Check 21, which is scheduled for implementation in October of this year. In this quarter's message, President Santomero describes some of the implications of this new legislation for both the Federal Reserve System and depository institutions.

The Federal Reserve System has three main responsibilities: conducting monetary policy, supervising and regulating banks, and maintaining an efficient and effective payments system. This third area — the payments system — has seen a great deal of change in recent years. Perhaps the greatest change has been the shift from paper forms of payment, that is, cash and checks, to electronic methods such as credit cards and debit cards.

We know that people's payments choices evolve slowly. It takes time for them to become comfortable with new payment methods. In the meantime, it is important that familiar, well-established payment methods be there for them. The Fed is committed to working with financial institutions to improve the reliability and efficiency of the *current* generation of payment vehicles and to foster innovation and support the *next* generation of payments vehicles. Both commitments are important.

However, I want to focus on another way the Fed is supporting an orderly evolution of the payments system: through its advocacy of a legal and regulatory framework that enables greater innovation in the marketplace — innovation by banks and other payment service providers — in response to evolving technologies and customer preferences.

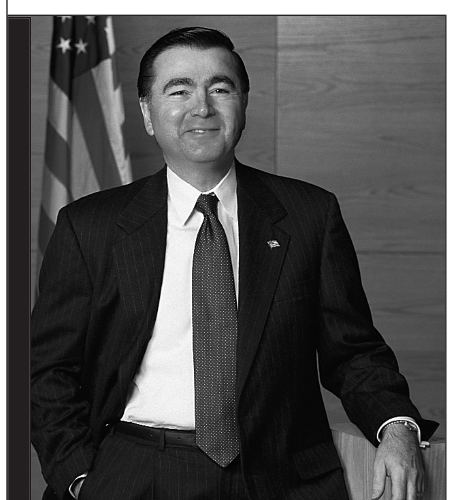
The Check Clearing for the 21st Century Act, also known as Check 21, is an example of such legislation.[1] The Federal Reserve vigorously supported Check 21 from the outset. We saw it as a means of enabling the marketplace to achieve greater efficiency and reliability in payments, proceeding at its own pace and in its own way.

PAYMENTS AND THE CHECK

In recent years, we have experienced a surge in electronic retail payments: credit cards, debit cards, and smart cards, as well as direct deposit and direct debit through the ACH. The emerging shift to electronics has been documented in a Philadelphia Fed Research Department analysis using data from the Federal Reserve Board's Survey of Consumer Finances.[2]

[1] For a synopsis of this act, see the Philadelphia Fed's *Banking Legislation and Policy*, January-March 2003 (www.phil.frb.org/files/blp/blp103.pdf).

[2] See Loretta J. Mester, "Changes in the Use of Electronic Payments," Federal Reserve Bank of Philadelphia *Business Review*, Third Quarter 2003 (www.phil.frb.org/files/br/brq303lm2.pdf).



Anthony M. Santomero, President, Federal Reserve Bank of Philadelphia

According to the survey, less than 18 percent of households used debit cards in 1995. By 2001, nearly half of all households were using them. Meanwhile, the percentage of households using automatic bill payment, although still relatively small, nearly doubled. Not only did usage of all means of electronic payments increase, but the increases were registered across all demographic categories: all age groups and all levels of income and education.

Other statistics show that the conversion of check payments to electronic transactions continues to grow rapidly. In fact, this past year, these so-called electronic check payments more than doubled, and they now exceed 100 million transactions per month.

So consumers are using electronic payments. In effect, the U.S. is catching up to the rest of the world in its use of electronic payments. Undoubtedly, the trend will continue. Indeed, a competitive marketplace will drive banks to offer new electronic payment vehicles in response to consumer demand for greater convenience at lower cost.

Yet, the use of that expressly American payment vehicle — the paper check — remains widespread here. Research financed by the Fed indicates that check use in the U.S. peaked in the mid-1990s and has been steadily declining since then. But Americans still write about 40 billion checks a year. That represents about half of the nation's retail non-cash transactions. Checks are likely to represent a significant share of payments for a long time. People see them as a very convenient, reliable, and familiar payment instrument. Bankers see them as a substantial source of revenue. So while checks will continue to decline, they will not completely disappear any time soon.

Over the years, banks have become quite efficient at processing paper checks. But as check volumes decline, the pressure will be on to find new processing efficiencies, and processing electronic information is more efficient than processing paper. Recognizing this, the Fed has been committed for some time to enabling greater use of electronics in check processing. Check 21 is an important step in that direction.

CHECK 21 AND EFFICIENCY

The goal of Check 21 is to foster innovation in the payments system and to enhance efficiency. It does this by facilitating check truncation and electronification via imaging without making it mandatory.[3] Check 21 accomplishes this simply by authorizing the use of a new negotiable instrument: the substitute check.

Notice that Check 21 does not require collecting banks to truncate or image checks, nor does it require paying banks to accept electronic images. Check 21 requires only that paying banks accept substitute checks as well as originals. Whether they accept the substitute checks in paper form or electronic form is strictly their decision.

The expectation is that Check 21 will, in fact, increase electronic presentments and foster the electronification of checks at the earliest possible stage of processing. The speed with which this evolution occurs is hard to predict. Like the use of the check itself, it is likely to be a gradual process.

Over time, we should see substantial efficiency gains as a result of Check 21. An article in the *American Banker* states that once the full effects of Check 21 are realized, it is estimated that the banking industry could potentially reduce its check processing costs by over $2 billion a year. That estimate includes a reduction of $250 million in transportation costs alone. In addition, banks will benefit from improved availability of funds and greater efficiency in processing return items.

Check 21 does not require collecting banks to truncate or image checks, nor does it require paying banks to accept electronic images. Check 21 requires only that paying banks accept substitute checks as well as originals.

The substitute check must contain payment information identical to that on the original check and must be in a specific machine-readable format. Properly created, the substitute check is the legal equivalent of the original check. Under Check 21, a collecting bank can truncate and image the original checks it receives for deposit, process the checks electronically, and then print and deliver substitute checks at a location near the paying bank for presentment.

CHECK 21 AND RELIABILITY

Thus far, I have emphasized the efficiency gains we expect from Check 21. Indeed, industry buzz has long expounded on the efficiency gains from the electronification of checks. But there is another benefit: mitigating

[3] Check truncation refers to removing the original paper check from the check collection process and replacing it with electronic information related to the original check.

risk. Check 21 will help alleviate the danger of checks being lost or delayed during transport. As clearing time shrinks, credit risk is reduced as well.

From the Fed's perspective, reducing the risks associated with the physical transportation of checks is an important benefit of Check 21. Our experience in the aftermath of 9/11 focused our attention on the value of electronification from the standpoint of reliability. The interruption of air travel — and check transportation — in the days after the 9/11 attacks pushed the Fed's check float to over $47 billion, more than 100 times the normal level.

Clearly, if image exchange had been more prevalent among banks, the impact would have been much smaller. Indeed, it was the opportunity to reduce check processing's dependency on the transportation system that motivated the Fed to approach Congress in late 2001 with the proposal that would become Check 21.

RESPONDING TO CHECK 21

After much discussion with the industry and consumer groups, Congress passed Check 21 in late 2003 and set implementation for October 28, 2004. It will take some time for the full effect of Check 21 to be realized.

Many banks are already using check imaging, both to streamline internal operations and to enhance services to their customers. Check 21 offers banks the option of using those images to collect from any paying bank by presenting a substitute check. Whether exercising this option makes economic sense, of course, depends on the relative cost of presenting a substitute check rather than the original, either directly or through a third-party provider.

Banks that use imaging must also make some important decisions about aggregating and archiving. As banks begin to receive electronic files from more institutions, it may make sense to outsource these activities to a third-party provider.

Some banks are not using imaging yet because the internal efficiencies and customer service benefits do not justify the cost. With Check 21, the expanded opportunity to transmit images for presentments may make imaging cost-effective.

But even if imaging does not make sense for some banks, Check 21 will require even non-imaging banks to accept presentments of substitute checks. That means bank customers will likely be getting back some substitute checks with their statements. Banks must have ready a plan for how they will familiarize customers with these new instruments and address their concerns.

> From the Fed's perspective, reducing the risks associated with the physical transportation of checks is an important benefit of Check 21.

As I indicated earlier, when a check is truncated at its bank of first deposit and a substitute check is created, the collection process is enhanced by expediting presentment, improving availability, and eliminating transportation costs. But consider an extension of this process, one in which presentment of the substitute check is replaced by an image. In this case, further benefits are extended to the paying bank. Accepting images for deposit eliminates back-office capture of the check as well as the inconvenience of transporting paper checks.

Let's extend this scenario even further. Suppose the bank of first deposit receives images from its own ATMs, from its own branch offices, or even from its corporate customers. Image capture earlier in the process will further benefit the bank by eliminating check transportation and the need for proofing and encoding and processing with check sorters. All this is made much easier with the passage of the new legislation.

CHECK 21 AND THE FED AS SERVICE PROVIDER

The Federal Reserve will also be responding to market needs during this period of substantial change in the payment patterns of consumers and businesses alike. To make this point, I'll outline how the Fed, as a provider of check services, is preparing to support the banking industry's best use of options under Check 21.

The Fed is investing in technologies that enhance Check 21. We want to provide all customers, regardless of size or location, the opportunity to embrace and take advantage of the many benefits of Check 21 — without significant investment on their part.

The Reserve Banks have recently upgraded and standardized their check platforms so that their bank customers will have access to the same check processing and adjustment services at all of its locations. Philadelphia installed the common platform in the fourth quarter of last year.

Now, we at the Fed are looking to October and preparing the services we will make available once Check 21 is implemented. The Reserve Banks will soon be rolling out a variety of new and improved products, services, and solutions designed to support banks' best use of their new options under Check 21. In addition, we will offer image deposit services and improved deposit deadlines and availability. For our payor services, we will have lower prices for items we collect via images. Also, we will convert paper items to images where we can expedite collection.

Our national archiving service, FedImage, is already available. We are expanding our capability to produce substitute checks, and we intend to increase our services available on the web.

Over time, Check 21 will provide financial institutions opportunities to broaden deposit options and extend deposit cut-off hours. This will have an effect on the Federal Reserve, too. As the Fed increases its processing efficiencies, it will pass the gains on to its bank customers in the form of accelerated availability and enhanced deposit deadlines. In short, the Fed and its bank customers are both involved in and affected by the changes imposed by the recent legislation, and the Fed is prepared to respond to Check 21's implications for itself and for the banking industry it serves.

In addition, as part of its public service role, the Federal Reserve is working to develop communication tools that will help banks educate their customers about Check 21.

So, we at the Fed will continue to develop our products and expand our electronic capacity in response to the market's evolution and our customers' needs. At the same time, we will take steps to foster an environment for improved payments system efficiencies and vibrant private-sector innovations. Check 21 is an important step toward this goal, but it is by no means the final step. The impact of this change and those that follow will ultimately transform the U.S. payments system and enable a radical restructuring of its service capabilities.

CONCLUSION

We at the Federal Reserve are pleased to see Check 21 nearing its implementation date. While people increasingly rely on electronic forms of payment, checks remain an important and trusted payments vehicle. Recognizing this, the Fed proposed Check 21 as a means of enabling the industry to make check processing more efficient and reliable. Now, banks must consider how to use the new options offered by Check 21 to create shareholder value and improve service quality.

As we move forward, the infrastructure and convention of check processing will evolve, generating new check products and services and new ways to deliver them. We at the Fed look forward to working with banks in achieving a common goal: realizing all of the efficiencies this new legislation offers. BR

We at the Fed will continue to develop our products and expand our electronic capacity in response to the market's evolution and our customers' needs.

What Test Scores Can and Cannot Tell Us About the Quality of Our Schools

BY THEODORE M. CRONE

How to best judge the quality of our schools is a thorny issue. Now, the No Child Left Behind Act, which was signed into law in January 2002, mandates standardized testing in math and reading for students in grades three through eight. The test scores will then be used both to gauge the students' level of proficiency in these subjects and to evaluate the schools' performance. But emphasizing test scores as a measurement of the quality of schools raises several questions. In this article, Ted Crone looks at some of these questions and warns us to be cautious in how we use test scores.

On January 8, 2002, President Bush signed into law the No Child Left Behind Act (NCLB), the latest reauthorization of the Elementary and Secondary Education Act. When fully implemented, the new law will require that students in grades three through eight take statewide standardized tests every year in math and reading. The scores on these tests will be used to determine whether students have achieved the required level of proficiency in these subjects for their grade level. Schools will be evaluated and rewarded or penalized on the basis of the test results. Since states are obligated under the law to release annual report cards on the schools, the general public is also likely to view these test scores as the primary measure of school quality.

This increased emphasis on standardized test scores as a measure of school quality and a tool for accountability raises the issue of what test scores can and cannot tell us about the quality of our schools. Should we be looking at average test scores or changes in test scores as the measure of quality? How much of a difference in either of these measures is significant? Finally, how can we distinguish the school's contribution to these test scores from the effects of the students' innate abilities, their family, social, and economic backgrounds, and the abilities and backgrounds of their peers in the classroom?

LARGE-SCALE TESTING IS NOT NEW TO THE U.S. EDUCATION SYSTEM

The beginning of large-scale external testing in the U.S., that is, tests developed outside the schools in which they are used, goes back to the mid-19th century.[1] Initially, such testing was limited. But the use of standardized tests increased significantly in the two decades after the development and publication of the Stanford Achievement Test in 1923. Between World War II and the 1960s, standardized tests were primarily used to evaluate students and curricula; they were not commonly used to hold schools accountable for student performance. Except for tests such as the SAT, which is used for college admissions, there were few direct consequences for the students or the schools associated with the scores on standardized tests.

In the 1960s two new programs at the national level expanded the role for large-scale testing. Title I of the Elementary and Secondary Education Act, enacted in 1965, provided federal funds to schools with a large percentage of low-income students. The act required the periodic testing of students in the program to assess its effectiveness. Also, in the late 1960s, the Education Commission of the States sponsored the first set of tests

Ted Crone is a vice president and head of the regional and urban economics section in the Research Department of the Philadelphia Fed.

[1] For a brief history of large-scale testing in the U.S., see the report from the U.S. Congress, Office of Technology Assessment. See also the article by Laura Hamilton and Daniel Koretz, and the two articles by Koretz.

known as the National Assessment of Educational Progress (NAEP). These tests have been administered periodically since 1969 to a random sample of 9-, 13-, and 17-year-olds in reading, math, and science to measure progress over time. A parallel set of NAEP tests was developed in the 1980s to be given in specific grades rather than to students based on their age. Besides being given to a national sample, this set of tests is given every two years to a sample of fourth and eighth graders in participating states and provides a basis of comparison among the states.

At the state level, large-scale standardized testing took on a new role in the 1970s. Many states adopted minimum-competency testing as a requirement for promotion or graduation or as a benchmark for assigning students to remedial programs. Prior to 1975, only two states had mandated any kind of minimum competency test; by 1980, however, 29 states had mandated such tests.[2] Minimum competency tests are essentially a tool for holding students accountable; as such, they require that each student take the test and that a cutoff score be established to determine who meets the minimum competency level. Results from these types of tests are likely always to be considered the best measure of academic competency for primary and secondary students.

The role of large-scale standardized tests expanded again in the 1980s and 1990s. Besides using standardized tests to hold *students* accountable, states began to use them to hold *schools* accountable, rewarding or penalizing schools based on test scores (the so-called high-stakes testing).[3] The first wave of reform began in the early 1980s and was given momentum by the publication of *A Nation at Risk*, a critical report on the state of American education by the National Commission on Excellence in Education. By the end of the decade many of the standardized tests introduced in this first wave of reform came under criticism for being too narrowly focused and not testing higher-level skills. A second wave of reform in the 1990s introduced standardized tests that were not as dependent on the multiple-choice format and that emphasized a broader range of skills. These were sometimes referred to as "tests worth teaching to." Assessment programs also began to rely on other measures *in addition to* test scores to evaluate students' achievement levels (for example, portfolios of students' work, presentations, and longer term projects).

Despite the reform efforts in the 1980s and 1990s and some improvement in national scores, achievement levels of U.S. students remained unacceptably low at the beginning of this century, giving rise to the testing requirements of NCLB. (See *Achievement Levels of U.S. Students.*) NCLB mandates yearly testing in math and reading in grades three through eight no later than the 2005-06 school year.[4] States are allowed to develop and administer their own tests, but they must specify what constitutes the acceptable level of proficiency for each grade. A sample of fourth and eighth graders from each state must also participate in the state-level NAEP tests every other year to provide a basis of comparison with the state's own tests.

NCLB requires that all students in each school reach the state-designated proficiency level on the state's own tests by the end of the 2013-14 school year. Prior to the

[2] See U.S. Congress, Office of Technology Assessment, p. 59.

[3] Test scores are also the most frequently used output measure in studies that estimate the effects of various school inputs; see the 1997 article by Eric Hanushek.

[4] NCLB continues the previous requirement that students be tested at least once in these two subjects in grades 10-12. By the 2007-08 school year, states will be required to test students in science at least once in the grade spans 3-5, 6-9, and 10-12.

2013-14 school year, schools that have not reached 100 percent proficiency must make adequate yearly progress toward that goal.[5] Adequate progress must be made for all students and for major subgroups of students (by race, ethnicity, income, and disability). The penalties for not achieving adequate progress become progressively severe.

For students in any school that fails to make adequate progress for two consecutive years, the district must provide them with a choice of public schools they can attend, and the state may be required to spend up to 5 percent of its federal funds under Title I to pay for that option. Subsequent years of inadequate progress result in further penalties. After five consecutive years of inadequate progress, districts are required to set up an alternative governance structure for the school. This could include reconstituting it as a charter school, turning over management to a private company, or having the state run the school. Thus, NCLB has significantly raised the stakes for schools based on student performance on standardized tests.

TO WHAT EXTENT DO STANDARDIZED TEST SCORES MEASURE SCHOOL QUALITY OR PERFORMANCE?

Individual Student Scores. Scores on standardized tests are primarily a measure of student achievement or competence in the subject being tested. They provide a better basis of comparison between students in different classrooms or different schools than scores on teacher-generated tests or course grades. Standardized test scores are not a perfect measure of achievement or competence, however. A written test cannot capture the full range of a student's abilities, and every test involves a certain amount of measurement error.[6] The reliability of a test is measured by the standard error of measurement or the degree to which the scores would spread out around the average score if the same student took the test many times. The measurement error on standardized tests can stem from a number of random factors, such as the student's health on the day of the test, the form of the test the student receives, or how well the student slept the night before. A mark of a well-designed test is that the measurement error is small relative to the range of scores on the test. For example, scores for the SAT I test used for college admissions range from 200 to 800, and the standard error of measurement is 30 points. In practice, this means that a test-taker could be 68 percent sure that her score on the test is within 30 points either way of her "true score" or average score if she took the test many times. She could be 95 percent sure that her score on the test is within 60 points either way of her true score.

> The measurement error on standardized tests can stem from a number of random factors, such as the student's health on the day of the test, the form of the test the student receives, or how well the student slept the night before.

The existence of measurement error raises a serious issue for minimum competency tests. There will be some misclassification *in both directions* when cutoff scores are used to determine which students meet the minimum level of proficiency. Each time the test is given, some students who are above the required achievement level are likely to score below the minimum and vice versa. For this reason most states that require a minimum proficiency score allow the students to take the test several times.[7] Unless students are allowed to take minimum competency tests more than once, the temptation will always be to lower the cutoff score to account for the measurement error in a single test score. The incentive to lower proficiency levels on tests is compounded by the fact that the national legislation provides no national standard for proficiency. Each state is allowed to set its own proficiency levels.

Average School Scores. Although tests are primarily a measure of individual student achievement, average scores or the percent of students scoring above a certain level are increasingly being used as measures of school quality and accountability. Usually a school will have some students with high scores and some with low scores, so the average score for the school will be somewhere in between, and the range of average scores across schools is much narrower than the range of individual scores for all students. For example, on the math and reading tests administered to fifth, eighth, and 11th graders as part

[5] There is some exception to the 100 percent goal for the learning disabled.

[6] See the article by Vi-Nhuan Le and Stephen Klein.

[7] See Chapter 3 of the report from the Center on Education Policy. Of the 19 states that had adopted a high-school exit exam in 2003, all but one allowed students to take the test two or more times. The one exception was Washington State, and the state's minimum competency requirement had not yet gone into effect.

of the Pennsylvania System of School Assessment (PSSA) in 2002, the range for school scores was only 50 percent to 60 percent as wide as the range for individual student scores in the state.[8] Researchers have consistently found that most of the variation in test scores is accounted for by the variation in individual students' scores within schools rather than by the variation between schools.[9] Thomas Kane and Douglas Staiger (2002b) report that the variation in fourth-grade math and reading scores in a typical North Carolina school is about 90 percent as large as the variation among all the state's fourth graders. The large variation in scores *within* schools is an argument for the NCLB requirement that not only schools as a whole but also major subgroups within each school make adequate yearly progress toward proficiency.

Since average scores for schools will be reported in states' annual reports, it is important to understand how reliable these average scores are and how well they measure the quality of the school. Average scores for a class are a more reliable measure of the "true" class average than is any individual's score of his true average. Many of the random factors that affect individual students' scores (for example, a student's health or the form of the test) tend to cancel out when scores are averaged across an entire class. However, some random factors, such as a distraction in the classroom, poor lighting, or imprecise instructions from the teacher, can affect average scores for the class as a whole. In their study of math and reading scores for fourth graders in North Carolina from 1992 to 1999, Kane and Staiger estimated that these types of random factors accounted for a relatively small percentage of the variation in average school scores — only 3.6 percent of the total variation among mid-size schools (Table, column 2, row 1).[10]

How much of the remaining variation in average test scores is due to differences in the instructional quality of the schools? Certainly, not all of it. Some of the variation in average scores across schools is due to differences in the quality of the students. The quality of students differs not only across schools and school districts but also across cohorts or age groups within the same school. In any given year, the students in a particular grade may be brighter than the students in other years *even though* they come from the same neighborhood. In their North Carolina sample, Kane and Staiger estimated that almost 11 percent of the variation in the combined reading and math scores for the fourth grade among mid-size schools is due to drawing a different sample of students from the neighborhood each year (Table, column 3, row 1).[11]

Thus, for the typical school random factors and cohort effects (an abnormal number of good or poor students) account for about 15 percent of the variation in school scores. But these factors influence the average scores of smaller schools more than those of larger schools. As a result a greater percentage of smaller schools tend to be at the top and bottom of the distribution of average scores in a given year. Kane and Staiger estimated that the combined effect of random factors and different cohorts accounts for less than 10 percent of the variation in average fourth-grade reading and math scores among the largest schools and almost 20 percent of the variation in scores among the smallest schools.[12] NCLB recognizes the problems with the high variability in scores for small samples by not requiring that average scores be reported for subgroups when the number of children is small.

Most of the variation in the average *level* of test scores from school to school is persistent; that is, it is not due to factors that change on a yearly

Most of the variation in the average *level* of test scores from school to school is persistent; that is, it is not due to factors that change on a yearly basis.

[8] For example, the individual math scores for the 11th grade ranged from 700 (at the first percentile) to 1893 (at the 99th percentile); the average school scores ranged from 770 (at the first percentile) to 1460 (at the 99th percentile). See the report from the Pennsylvania Department of Education.

[9] In general, the variation among schools' average scores accounts for only 10 percent to 20 percent of the total variation in test scores; the rest is due to variation among individual students within schools. The Coleman report in the mid 1960s found that nationally about 16 percent or less of the variation in reading and math scores on achievement tests for sixth, ninth, and 12th graders could be attributed to variation across schools. David Figlio (February 13, 2002) reported that only 14 percent to 15 percent of the variation in math and reading scores in two Florida school districts could be attributed to the variation between schools.

[10] See Kane and Staiger, 2002b. Mid-size schools are those in the middle quintile by size; on average they have 56 fourth-grade students.

[11] This does not include differences in the student population across neighborhoods serviced by different schools.

[12] The largest quintile of schools in North Carolina has an average of 104 students in the fourth grade, and the smallest quintile has an average of 28 students in the fourth grade. The greater variability in test scores for smaller schools than larger schools due to these transitory effects has also been documented for Chile. See the paper by Kenneth Chay, Patrick McEwan, and Miguel Urquiola.

basis. Kane and Staiger's analysis of the variation in fourth-grade math and reading scores in North Carolina suggests that about 85 percent of the variation in the *level* of test scores across mid-size schools is persistent (Table, column 1, row 1). Evidence from the PSSA also shows that the relative differences in scores across schools are persistent. The correlation of 11th-grade scores for public high schools for consecutive years between 1998 and 2002 is approximately 0.88 for math and 0.80 for reading.[13]

[13] Although schools' relative PSSA test scores are fairly stable across years, in every year between 1999 and 2002, there are examples of the average 11th grade math or reading score for a school moving from the 25th to the 75th percentile in the state or vice versa.

Who's Responsible for These Persistent Differences in Test Scores? Are these persistent differences a measure of the quality of the school or a measure of the abilities and backgrounds of the students? Economic studies of the educational process have identified three major influences on student achievement besides the quality of the school: the student's innate ability and family characteristics and the characteristics of the student's classroom peers.[14]

Teachers are well aware of the wide range of student abilities from the learning disabled to the gifted. But it is difficult to get a pure measure of the innate ability of students. Initial test scores are not a pure measure of innate ability; by the time students enter the school system their achievement levels have been influenced by a number of environmental and social factors. Moreover, as students progress through the school system, their achievement levels are the result of their cumulative educational experience.

Family characteristics, such as the parents' education and income, can also affect the level of student achievement and test scores. Education takes place not only in the classroom but also at home; in general, students whose parents are more highly educated have a better educational environment in the home. For example,

[14] See the article by Byron Brown and Daniel Saks, and the 1979 and 1986 articles by Hanushek.

TABLE

Sources of Variation in Fourth-Grade Test Scores for Mid-Size Schools in North Carolina

(Average number of fourth graders = 56)

	% of Variance Due to Persistent Characteristics of the School	% of Variance Due to Purely Random Factors	% of Variance Due to Differences in Cohorts Within the School
Combined Reading and Math Scores for Fourth Grade	85.4%	3.6%	10.9%
Change in Combined Reading and Math Scores for Fourth Grade from One Year to the Next (Variation is 40 percent as great as variation in scores across schools)	29.1%	16.4%	54.5%
Change in Combined Reading and Math Scores from Third Grade to Fourth Grade (Variation is 23 percent as great as variation in scores across schools)	51.6%	35.5%	12.9%

Source: Author's calculations from Kane and Staiger, 2002b, Table 2. Mid-size schools are those in the middle quintile based on size. Time frame: 1992-99.

students who are exposed to a richer vocabulary in their family conversations are more likely to perform better in language arts than students who are exposed to a limited vocabulary. Income matters as well. Higher income families can better afford certain aids to education, such as computers or private tutors. In addition, students from higher income families are more likely to have more educational experiences like foreign travel.

Finally, a number of studies have found that the achievement levels and other characteristics of fellow students in the classroom can have an effect on a student's own achievement and test scores — the so-called peer-group effect.[15] Peers can provide motivation for a student. They can contribute to learning through direct interaction, or they can affect the learning process in the classroom. A disruptive student clearly hinders the learning process for his peers, but a bright student can aid the process by asking questions that help other students as well.[16] A good set of peers in the classroom can increase the quality of the school by enhancing the learning environment, but in the U.S. public school system, classroom peers are largely determined by the families who choose to live in the neighborhood, not by the school.

The cumulative effect of students' innate abilities, family backgrounds, and peers will be reflected in tests scores. Schools, however, have little or no influence over these factors, raising the issue of how test scores can be used to judge the school's contribution to learning. One suggestion is to use *changes* in test scores rather than the *level* of scores to measure school quality and performance.[17]

ARE CHANGES IN TEST SCORES A BETTER MEASURE OF SCHOOL QUALITY?

Changes in Scores for a Given Grade. There are several ways to measure changes in test scores in a system of school accountability.[18] The first is to compare this year's score for a given grade with last year's score for that grade, for example, the change in the average test score for fourth grade. Of course, we are not comparing the same students in this exercise, and so it is difficult to know how much any increase in the average score represents an improvement in student achievement. The school's contribution to this increase in the score is equally difficult to assess. Kane and Staiger estimate that more than 50 percent of the variation in the annual change in fourth-grade reading and math scores across mid-size schools in North Carolina is due to the fact that a different cohort of students is being tested each year, and each cohort has a different average level of ability (Table, column 3, row 2).

> Random factors such as a large number of students with a cold on the day of the test can also contribute to the change in scores for a given grade from one year to the next.

Besides these cohort effects, random factors such as a large number of students with a cold on the day of the test can also contribute to the change in scores for a given grade from one year to the next. According to Kane and Staiger, the combination of cohort effects and these kinds of random factors accounts for more than 70 percent of the variation in the annual change in fourth-grade scores in North Carolina.[19]

Changes in Scores for a Given Cohort of Students. A partial solution to the problem of comparing two different cohorts of students is to compare this year's average fourth-grade score with last year's third-grade score. But this is only a partial solution for two reasons: The composition of the class may have changed as some students enter or leave the class, and even if there has been no change in the composition of the class, different cohorts of students advance at different rates. If a cohort of particularly able students has moved from third to fourth grade this year, a larger than average increase in scores may not be due to the school at all. The change in scores from one grade to the next tends to be considerably less variable than the change in scores for a given grade. But even in this case, only about half the variation in the change in scores can be attributed to differences

---

[15] See the articles by Anita Summers and Barbara Wolfe; Vernon Henderson, Peter Mieszkowski, and Yvon Sauvageau; Ron Zimmer and Eugenia Toma; Caroline Hoxby; and Erick Hanushek, John Kain, Jacob Markman, and Steven Rivkin. Peer-group effects are difficult to isolate and hard to separate from school effects, since students generally attend school for some years with most of their classroom peers. Joshua Angrist and Kevin Lang found only very weak evidence of peer-group effects in their study.

[16] See the article by Edward Lazear and the one by Hanushek et al.

[17] See Hanushek's 1986 article, and the article by Hanushek and Lori Taylor.

[18] See the article by Laura Hamilton and Daniel Koretz.

[19] See Kane and Staiger, 2002b. In a study of test scores in Florida, David Figlio and Marianne Page found that the correlation between *changes* in average test scores in consecutive years for a given grade at a school was negative, supporting the notion that changes in test scores are a noisy measure of school quality.

in the quality of the schools (Table, column 1, row 3). Kane and Staiger estimate that in mid-size schools in North Carolina about 13 percent of the variation in the average change in test scores from third to fourth grade is due to the cohort that is advancing that year, and more than 35 percent is due to purely random factors (Table, columns 2-3, row 3).

Changes in Individuals' Test Scores. A more refined measure of the value added by a school is the improvement in individual student scores over time rather than the improvement in class scores from one year to the next.[20] Students' test scores are highly correlated from one year to the next, so it may take a longer period to capture meaningful changes in a student's scores compared to the average change.[21] But data on individual students' test scores are difficult to maintain over time, especially for students who are very mobile. Furthermore, tracking individual students does not solve all the issues of identifying the school's contribution to any improvement in scores. Family background and innate ability influence not only the level of scores at a point in time but the rate of change as well. A student whose father or mother has a graduate degree in engineering is likely to get more help on his algebra homework and, therefore, advance more quickly than the student whose parents did not graduate from high school. Kane and Staiger found that students in North Carolina whose parents had a higher level of education had greater gains in test scores from the end of third grade to the end of fourth grade. Whether we compare schools based on the *level* of test scores or some measure of *change* in scores, the school's contribution has to be determined in light of the innate abilities and backgrounds of the students.

Teaching to the Test. Another word of caution has to be raised about changes in scores on high-stakes tests whose results have serious consequences for the school. No matter how we measure changes in test scores, there is a tendency in the early years after a new high-stakes test is introduced for scores to rise rapidly. Daniel Koretz provides a striking example of inflation in high-stakes test scores.[22] He and his colleagues tracked student performance in two school districts on several tests; some were high-stakes tests and some were not. The figure illustrates what happened to third-grade math scores in one district that changed its high-stakes test between 1986 and 1987. In the final year in which the old test was given, the median grade equivalent was 4.3 for the third graders in the district. In the first year of the new test, the grade equivalent dropped to 3.7, but by the fourth year of administering the new test, the median grade had climbed back to 4.3. In the fourth year of the new test, Koretz and his colleagues administered the old test to a random sample of third graders. Their median score on the old test was 3.7. Scores on the new district-wide test had increased substantially in the four years, and scores on the old test had dropped.

The initial rapid rise in high-stakes test scores is often attributed to the practice of "teaching to the test." There is evidence that teachers do

---

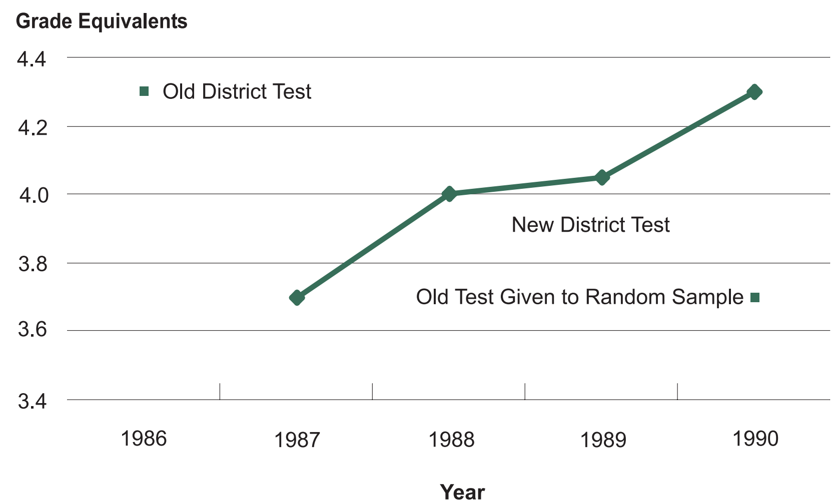
[20] Figlio and Page found that the ranking of Florida schools based on the improvement in individual scores is very different from the ranking based on the average level of scores in a given year. But the available data only allowed them to calculate the change in individual reading scores from the fourth to the fifth grade.

[21] In Kane and Staiger's study of North Carolina schools, the correlation between individual students' standardized third- and fourth-grade scores was 0.80.

[22] See the article by Daniel Koretz, Robert Linn, Stephen Dunbar, and Lorrie Shepard, and both articles by Koretz.

FIGURE

Median Grade Equivalent (Old Test and New Test) Third-Grade Mathematics



Source: Daniel M. Koretz, *Journal of Human Resources*, 2002. Used with permission of the author.

spend more time on the subjects tested in their grade than on other subjects.[23] In any given subject, teachers can emphasize the material they know will be covered on a high-stakes test. These are not necessarily negative consequences of high-stakes testing. *If* high-stakes tests adequately cover the essential material to be learned in each grade, these practices can enhance the teaching in the classroom. Teaching to a well-designed set of tests can improve both test scores and student achievement. But we cannot assume that every improvement in test scores is an improvement in overall academic achievement. Some classroom practices improve test scores on high-stakes tests but have little or no effect on achievement levels. For example, teachers learn over time how to administer tests with less confusion, and they prepare students for the format of the new high-stakes test.

One check on whether higher test scores are measuring true gains in achievement or simply reflect score inflation is to compare the improvement in scores on high-stakes tests with improvement in other test scores. Researchers have compared gains in several state-mandated tests with gains in the NAEP tests taken in the state. The results are mixed. Gains in test scores from the Kentucky Instructional Results Information System (KIRIS) were not matched by gains in the state's NAEP scores.[24] In the first two years of the program, fourth-grade reading scores increased dramatically on the KIRIS tests but did not increase at all on the NAEP tests. In the first four years of the program, fourth- and eighth-grade math scores increased three-and-a-half to four times more on the KIRIS tests than on the NAEP tests. Perhaps the most publicized high-stakes testing program in the country has been the Texas Assessment of Academic Skills (TAAS). Stephen Klein and his associates compared gains in TAAS scores from 1994 to 1998 with gains in the Texas NAEP scores.[25] *Both* sets of tests showed gains in reading and math in fourth and eighth grade, but the gains on the TAAS tests were much larger than those on the NAEP tests. Moreover, other educational outcomes such as graduation rates or plans to attend college have not improved with the gains in the Texas test scores.[26] Unlike the situation in Kentucky and Texas, increases in scores on the North Carolina state tests were about the same as increases in the NAEP scores in the state. This may be because the North Carolina tests are more similar to the NAEP tests.[27] The possibility of serious grade inflation on high-stakes test scores reinforces the need for a comparison test such as the NAEP, against which we can measure any improvement in high-stakes test scores.

One check on whether higher test scores are measuring true gains in achievement or simply reflect score inflation is to compare the improvement in scores on high-stakes tests with improvement in other test scores.

SUMMARY: SOME CONSIDERATIONS ON THE USE OF TEST SCORES TO EVALUATE SCHOOL QUALITY

Test scores are primarily a measure of the achievement levels of individual students, but they are increasingly being used to measure the quality of schools. This new role for testing is a response to the performance of U.S. students relative to students from other industrialized countries and to the large percentage of U.S. students who do not meet proficiency levels on standardized tests. The new testing programs are designed to hold schools as well as students accountable. Test scores and changes in test scores are one of the few quantitative measures of school quality we have, but special precautions need to be taken when test scores are used to evaluate schools rather than students.

Perhaps the most popularly accepted notion in judging the quality of schools is that *all* students should achieve a minimum level of competency based on some standardized test in order to graduate or be promoted. But there is some measurement error in the score on every test, and some students who are above the minimum in achievement will not receive the minimum score on a single administration of the test. Therefore, if all students are required to score above the minimum on a single administration of the test, states will be tempted to lower the cutoff score for proficiency to account for the measurement error. Students should have more than one chance to achieve the minimum score

[23] See the article by Brian Stecher.

[24] See the paper by Daniel Koretz and Sheila Barron.

[25] See the article by Stephen Klein, Laura Hamilton, Daniel McCafferty, and Brian Stecher. Robert Linn, Eva Baker, and Damian Betebenner also point out that the percent of students meeting proficiency levels in the TAAS tests increased much faster than the percent of students meeting proficiency levels on the Texas NAEP tests.

[26] See the article by Martin Carnoy, Susanna Loeb, and Tiffany Smith.

[27] See Thomas Kane and Douglas Staiger, 2002a.

on these tests, and this should be true of tests that have serious consequences for the school as well as those that have serious consequences for the students.

Average school scores are less susceptible to measurement error than individual student scores. But the average score may not measure the school's contribution to the students' achievement for several reasons. Each cohort of students in a school will differ in their abilities, and the family characteristics and innate abilities of students will differ from school to school. Moreover, peer effects can magnify these differences. Therefore, if we want to use average scores to judge the quality of schools, we must look at scores over several years and compare scores for schools that have students from similar backgrounds.

Theoretically, changes in test scores should be a better measure of the school's contribution to student achievement than average scores. But there is a lot of random variation in the changes in scores for a given grade or for a given class from one year to the next. Longer-term trends in test scores can eliminate some of this random variation in the changes in scores. But not all of the long-term improvement in class scores or individual scores can be attributed to the school. Family characteristics and peer effects influence how quickly students advance in their education. So every easily accessible measure of student achievement has some drawback as a measure of school quality.

Despite the shortcomings of standardized test scores as a measure of school performance there is no generally recognized substitute; test scores simply have to be used with caution. Improvements in high-stakes test scores should be checked against improvements in other tests such as the state-level NAEP tests. Other measures of student achievement, such as course grades and performance on longer-term projects, can be incorporated into the evaluation of school quality. Finally, other criteria, such as graduation rates and the percent of students attending college are important in evaluating how well our schools perform. BR

Appendix: Achievement Levels of U.S. Students

Since the 1960s, a number of countries have administered math and science tests so that student achievement can be compared across countries. U.S. students have tended to score in the middle of the pack or lower in these international comparisons.[a] In the latest Third International Mathematics and Science Study (TIMSS) conducted in 1999, eighth-grade students in the U.S. ranked 19th in math among the 38 countries participating. In science, they ranked 18th out of 38 (Table A1). A number of explanations have been offered for the poor ranking of the U.S. in the TIMSS tests relative to nations like Japan, Korea, the Netherlands, and Australia, but there are no simple explanations for the differences in performance across countries.[b] Nonetheless, the rankings suggest considerable room for improvement in the American education system.

The trend and dispersion in student achievement within the U.S. are illustrated by the scores on the two types of tests given as part of the National Assessment on Educational Progress (NAEP) — the national trend tests and the state tests. The scores from the long-term trend NAEP tests offer the best assessment of student achievement over time, since the tests have changed very little since they were first administered. The average math and science scores on these tests show a pattern of deterioration in the 1970s, improvement in the 1980s, and a leveling off in the 1990s. The math scores have shown the most consistent improvement (Figures A1 and A2). Reading scores have shown little sustained improvement since the tests were first administered (Figure A3). For all age groups (9, 13, and 17) the latest reading scores are not significantly higher than they were in 1980.[c]

The state-level NAEP tests, which were first administered in the early 1990s, differ from the tests that capture the national trend because they are adjusted over time to reflect changing curricula and they are given in specified grades, not at given age levels. The National

[a] See Eric Hanushek's 1998 article.
[b] See the article by Deborah Nelson.
[c] See the report from the U.S. Department of Education.

TABLE A1

1999 TIMSS SCORES

8th Grade Math		8th Grade Science	
Singapore	604	569	Chinese Taipei
Korea	587	568	Singapore
Chinese Taipei	585	552	Hungary
Hong Kong	582	550	Japan
Japan	579	549	Korea
Belgium	558	545	Netherlands
Netherlands	540	540	Australia
Slovak Republic	534	539	Czech Republic
Hungary	532	538	England
Canada	531	535	Belgium
Slovenia	530	535	Finland
Russian Federation	526	535	Slovak Republic
Australia	525	533	Canada
Czech Republic	520	533	Slovenia
Finland	520	530	Hong Kong
Malaysia	519	529	Russian Federation
Bulgaria	511	518	Bulgaria
Latvia	505	515	United States
United States	502	510	New Zealand
England	496	503	Latvia
New Zealand	491	493	Italy
Lithuania	482	492	Malaysia
Italy	479	488	Lithuania
Cyprus	476	482	Thailand
Romania	472	472	Romania
Moldova	469	468	Israel
Thailand	467	460	Cyprus
Israel	466	459	Moldova
Tunisia	448	458	Macedonia
Macedonia	447	450	Jordan
Turkey	429	448	Iran
Jordan	428	435	Indonesia
Iran	422	433	Turkey
Indonesia	403	430	Tunisia
Chile	392	420	Chile
Philippines	345	345	Philippines
Morocco	337	323	Morocco
South Africa	275	243	South Africa

Source: U.S. Department of Education. National Center for Education Statistics. *Pursuing Excellence: Comparisons of International Eighth Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999.*

TABLE A2

Percent of Students Scoring Below Basic Level (NAEP)

	Math		Reading	
	4th Grade	8th Grade	4th Grade	8th Grade
Nation (public schools)	24	33	38	28
Alabama	35	47	48	35
Alaska	25	30	42	33
Arizona	30	39	46	34
Arkansas	29	42	40	30
California	33	44	50	39
Colorado	23	26	31	22
Connecticut	18	27	26	23
Delaware	19	32	29	23
District of Columbia	64	71	69	53
Florida	24	38	37	32
Georgia	28	41	41	31
Hawaii	32	44	47	39
Idaho	20	27	36	24
Illinois	27	34	39	23
Indiana	18	26	34	23
Iowa	17	24	30	21
Kansas	15	24	34	23
Kentucky	28	35	36	22
Louisiana	33	43	51	36
Maine	17	25	30	21
Maryland	27	33	38	29
Massachusetts	16	24	27	19
Michigan	23	32	36	25
Minnesota	16	18	31	22
Mississippi	38	53	51	35
Missouri	21	29	32	21
Montana	19	21	31	18
Nebraska	20	26	34	23
Nevada	31	41	48	37
New Hampshire	13	21	25	19
New Jersey	20	28	30	21
New Mexico	37	48	53	38
New York	21	30	33	25
North Carolina	15	28	34	28
North Dakota	17	19	31	19
Ohio	19	26	31	22
Oklahoma	26	35	40	26
Oregon	21	30	37	25
Pennsylvania	22	31	35	24
Rhode Island	28	37	38	29
South Carolina	21	32	41	31
South Dakota	18	22	31	18
Tennessee	30	41	43	31
Texas	18	31	41	29
Utah	21	28	34	24
Vermont	15	23	27	19
Virginia	17	28	31	21
Washington	19	28	33	24
West Virginia	25	37	35	28
Wisconsin	21	25	32	23
Wyoming	13	23	31	21

Source: http://nces.ed.gov/nationsreportcard/
The three states in the Third Federal Reserve District are shaded.

Assessment Governing Board, which oversees the test, has adopted three achievement levels for reporting the results — basic, proficient, and advanced.[d] The basic level "denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade" (www.nagb.org/about/achievement.html).

The No Child Left Behind Act required all states to participate in these tests for fourth- and eighth-grade students by the 2002-03 school year. The results were not encouraging (Table A2). Nationwide, 24 percent of fourth-grade public-school students and 33 percent of eighth graders scored below the basic level in math. Even in the best performing states, 13 percent of fourth graders and 18 percent of eighth graders scored below the basic level. In the three worst performing states, more than one-third of the fourth graders scored below the basic level and in 10 states more than 40 percent of the eighth graders scored below basic.[e] On the reading tests 38 percent of fourth graders and 28 percent of eighth graders nationwide scored below the basic level. Even in the best performing states, 25 percent of fourth graders and 18 percent of eighth graders scored below basic. In the 13 worst performing states, more than 40 percent of fourth graders scored below basic in reading, and in seven states, more than one-third of the eighth graders scored below basic in reading. These results suggest that the need for improvement in student achievement is not limited to a few states or school districts.

[d] These are not related to the proficiency levels to be determined by each state according to the No Child Left Behind Act.
[e] These numbers exclude the District of Columbia where more than 50 percent of the fourth- and eighth-grade students scored below basic on the math and reading tests.

FIGURE A1

Average Math Scores (NAEP)

Source: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. *NAEP 1999, Trends in Academic Progress: Three Decades of Student Performance.*

FIGURE A2

Average Science Scores (NAEP)

Source: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. *NAEP 1999, Trends in Academic Progress: Three Decades of Student Performance.*

FIGURE A3

Average Reading Scores (NAEP)

Source: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. *NAEP 1999, Trends in Academic Progress: Three Decades of Student Performance.*

REFERENCES

Angrist, Joshua D., and Kevin Lang. "How Important Are Classroom Peer Effects? Evidence from Boston's METCO Program," NBER Working Paper 9263 (October 2002).

Baker, Eva L., and Robert L. Linn. "Validity Issues for Accountability Systems," CSE Technical Report 585, Center for the Study of Evaluation, National Center for Research on Education (December 2002).

Brown, Byron W., and Daniel H. Saks. "The Production and Distribution of Cognitive Skills Within Schools," *Journal of Political Economy*, 83 (1975), pp. 571-93.

Carnoy, Martin, Susanna Loeb, and Tiffany L. Smith. "Do Higher State Test Scores in Texas Make for Better High School Outcomes?" Research Report Series RR-047, Consortium for Policy Research and Education, November 2001.

Center on Education Policy. *State High School Exit Exams Put to the Test.* Washington, DC, August 2003.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiloa. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools," NBER Working Paper 10118, November 2003.

Coleman, James S., et al. *Equality of Educational Opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare, 1966.

Figlio, David N. "Aggregation and Accountability," presented at Will No Child Truly Be Left Behind? The Challenge of Making This Law Work, a conference sponsored by the Thomas B. Fordham Foundation, February 13, 2002.

Figlio, David N., and Marianne E. Page. "Can School Choice and School Accountability Successfully Coexist?" in Caroline M. Hoxby (ed.): *The Economics of School Choice.* Chicago: University of Chicago Press, 2003.

Hamilton, Laura S., and Daniel M. Koretz. "Tests and Their Use in Test-Based Accountability Systems," in Laura Hamilton, Brian M. Stecher, and Stephen P. Klein (eds.): *Making Sense of Test-Based Accountability in Education.* Rand Corporation, 2002, pp.13-49.

Haney, Walt. "The Myth of the Texas Miracle in Education," *Education Policy Archives*, 8, 41 (August 2000).

Hanushek, Eric A. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources*, 14 (1979), pp. 351-88.

Hanushek, Eric A. "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24 (1986), pp. 1141-77.

Hanushek, Eric A. "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, 19 (1997), pp. 141-64.

Hanushek, Eric A. "Conclusions and Controversies about the Effectiveness of School Resources," *Economic Policy Review*, Federal Reserve Bank of New York, March 1998, pp. 11-27.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. "Does Peer Ability Affect Student Achievement?" NBER Working Paper 8502 (October 2001).

Hanushek, Eric A., and Lori L. Taylor. "Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement," *Journal of Human Resources*, 25 (1990), pp. 179-201.

Henderson, Vernon, Peter Mieszkowski, and Yvon Sauvageau. "Peer Group Effects and Educational Production Functions," *Journal of Public Economics*, 10 (1978), pp. 97-106.

Hoxby, Caroline. "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working Paper 7867 (August 2000).

Jaeger, Richard M. "The Final Hurdle: Minimum Competency Achievement Testing," in Gilbert R. Austin and Herbert Garber (eds.): *The Rise and Fall of National Test Scores.* NY: Academic Press, 1982, pp. 223-46.

Kane, Thomas J., and Douglas O. Staiger. "Improving School Accountability Measures," NBER Working Paper 8156 (March 2001).

Kane, Thomas J., and Douglas O. Staiger. "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*, 16 (2002a), pp. 91-114.

Kane, Thomas J., and Douglas O. Staiger. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems," in Dianne Ravitch (ed.): *Brookings Papers on Education Policy*, 2002b, Washington, DC: Brookings Institution, pp. 235-83.

REFERENCES

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffery, and Brian Stecher. "What Do Test Scores in Texas Tell Us?" *Education Policy Analysis Archives*, 8, 49 (October 26, 2000).

Koretz, Daniel M. "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity," *Journal of Human Resources*, 38 (2002), pp.752-77.

Koretz, Daniel M. "Using Student Assessments for Educational Accountability," in Eric A. Hanushek and Dale W. Jorgenson (eds.): *Improving America's Schools.* Washington, DC: National Academy Press, 1996, pp. 171-95.

Koretz, Daniel M., and Sheila I. Barron. "The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)," Santa Monica: Rand Corporation, 1998.

Koretz, Daniel M., Robert L. Linn, Stephen B. Dunbar, and Lorrie A. Shepard. "The Effects of High-Stakes Testing on Achievement: Preliminary Findings About Generalization Across Tests," presented at *Effects of High-Stakes Educational Testing on Instruction and Achievement*, symposium presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April 5, 1991.

Lazear, Edward P. "Educational Production," *Quarterly Journal of Economics*, 116 (2001), pp. 777-803.

Le, Vi-Nhuan, and Stephen P. Klein. "Technical Criteria for Evaluating Tests," in Laura S. Hamilton, Brian M. Stecher, and Stephen P. Klein (eds.): *Making Sense of Test-Based Accountability in Education.* Rand Corporation, 2002, pp. 51-77.

Linn, Robert L., Eva L. Baker, and Damian W. Betebenner. "Accountability Systems: Implications of Requirements of the No Child Left Behind Act of 2001," *Educational Researcher* 31, 6 (2002), pp. 3-16.

Meyer, Robert H. "Value-Added Indicators of School Performance," in Eric A. Hanushek and Dale W. Jorgenson (eds.): *Improving America's Schools.* Washington, DC: National Academy Press, 1996, pp. 197-223.

National Commission on Excellence in Education. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: Government Printing Office, 1983.

Nelson, Deborah I. "What Explains Differences in International Performance? TIMSS Researchers Continue to Look for Answers," *Policy Briefs*, Consortium for Policy Research in Education, September 2003.

Pennsylvania Department of Education. *The Pennsylvania System of School Assessment: Handbook for Report Interpretation 2002 PSSA Mathematics and Reading Assessment*. Harrisburg, PA: November 2002.

Rothstein, Jesse M. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition Among Jurisdictions," Working Paper 3, Princeton University, Education Research Section, October 2003.

Stecher, Brian M. "Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practice," in Laura S. Hamilton, Brian M. Stecher, and Stephen P. Klein (eds.): *Making Sense of Test-Based Accountability in Education*. Rand Corporation, 2002, pp. 79-100.

Summers, Anita A., and Barbara l. Wolfe. "Do Schools Make a Difference?" *American Economic Review,* 67 (1977), pp. 639-52.

U.S. Congress, Office of Technology Assessment, *Testing in American Schools: Asking the Right Questions.* (OTA-SET-519) Washington, DC: U.S. Government Printing Office, February 1992.

U.S. Department of Education. National Center for Education Statistics. *Pursuing Excellence: Comparisons of International Eighth Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999.* NCES 2001-028, by Patrick Gonzales, Christopher Calsyn, Leslie Jocelyn, Kitty Mak, David Kastberg, Sousan Arafeh, Trevor Williams, and Winnie Tsen. Washington, DC: U.S. Government Printing Office, 2000.

U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics. *NAEP 1999, Trends in Academic Progress: Three Decades of Student Performance.* NCES 2000-469, by J.R. Campbell, C.M. Hornbo, and J. Mazzeo. Washington, DC: 2000.

Zimmer, Ron W., and Eugenia F. Toma. "Peer Effects in Private and Public Schools Across Countries," *Journal of Policy Analysis and Management*, 19 (2000), pp. 75-92.

The Software Patent Experiment*

BY ROBERT HUNT AND JAMES BESSEN

Over the past two decades, the scope of technologies that can be patented has been expanded to include many items previously thought unsuitable for patenting, for example, computer software. Today, the U.S. Patent and Trademark Office grants 20,000 or more software patents a year. Conventional wisdom holds that extending patent protection to computer programs will stimulate research and development and, thus, increase the rate of innovation. In this article, Bob Hunt and Jim Bessen investigate whether this has, in fact, happened. They describe the spectacular growth in software patenting, who obtains patents, and the relationship between a sharp focus on software patenting and firms' investment in R&D.

When it comes to patents, the U.S. has undergone an almost accidental process of legal innovation over the last two decades. Standards have been eased: We now issue patents for inventions that, in the past, would not have qualified for protection. In addition, the scope of technologies that can be patented has been increased to include, among other things, gene sequences, computer programs, and methods of doing business.[1]

Bob Hunt is a senior economist in the Research Department of the Philadelphia Fed.

This article investigates the effects of extending the patent system to a field of technology — computer software — known for rapid innovation well before software patents became commonplace. According to our estimates, the United States Patent and Trademark Office (USPTO) now grants at least 20,000 software patents a year, and the numbers are growing rapidly. The European Commission is debating a proposal to formally recognize the patentability of computer programs in member countries. These changes have been controversial, but they are typically justified by the argument that making patents easier to obtain will increase the incentive to invest in research and development (R&D) and, therefore, the rate of innovation.[2] In policy circles, it is fair to say this is the conventional wisdom.

There is sound empirical evidence that, for at least some industries, the availability of patents is an important factor that explains the willingness of firms to invest in R&D. For example, a number of surveys establish the important role that patents play in the U.S. chemical and pharmaceutical industries.[3] But these surveys also show

---

* An earlier version of this article was published in: *Patents, Innovation, and Economic Performance*, Proceedings of the Conference of the Organization for Economic Cooperation and Development on "IPR, Innovation, and Economic Performance," Paris, August 2003, OECD, Paris (forthcoming).

[1] For an examination of the economic effects of changing patent standards, see Bob Hunt's 1999 *Business Review* article. For a brief history of intellectual property rights for computer programs, see Hunt's 2001 *Business Review* article.

[2] Many studies examine the relationship between growth in R&D and growth in productivity or economic output. See, for example, the working paper by Dominique Guellec and Bruno van Pottelsberghe de la Potterie and the review article by Zvi Griliches.

[3] See, for example, the article by Richard Levin, Alvin Klevorick, Richard Nelson, and Sidney Winter and the working paper by Wesley Cohen, Richard Nelson, and John Walsh. Using data compiled for that working paper, Ashish Arora, Marco Ceccagnoli, and Wesley Cohen present evidence that firms that rate patents as both important and effective tend to do more R&D.

Jim Bessen is a lecturer at Boston University School of Law.

that in many other industries, patents are not regarded as either very important or effective in protecting one's innovations. Other general reviews of the effects of the patent system reach ambiguous conclusions: Patents help in many circumstances but not in others, and in some instances, the effects may be deleterious.[4]

The research described in this article suggests there is some reason for concern about the economic effects of software patents.[5] We found that software patents are not closely related to the creation of computer programs — the vast majority of software patents are obtained by firms outside the software industry.[6] We also found that firms that focus on software patents, in the sense that a higher share of their new patents is software patents, have tended to focus less on research than other firms. Interpreting these facts is difficult, but they do suggest that the relationship between the increased availability of software patents and the incentive to invest in R&D is more complicated than is often assumed in the policy debate. In short, we did not find evidence in favor of the conventional wisdom.

CHANGES TO OUR PATENT SYSTEM

The American patent system has changed in a number of important ways over the last quarter of a century. Some of these changes include the relaxation of standards used to determine whether an invention qualifies for patent protection and the elimination of the so-called subject matter exception that precluded the patenting of computer programs.

What Is a Patent? For more than 200 years, the U.S. government has used patents to reward inventors for their discoveries. The reward is a grant of the legal right to exclude others from making, using, or selling the patented invention for a limited period of time.[7] If the patent is infringed, the patent owner may sue the infringer to recover lost profits. If the infringement was willful, the court may award additional damages.

> The American patent system has changed in a number of important ways over the last quarter of a century.

In certain ways, a patent is a custom design. The inventor's right to exclude is limited to those *claims* applied for and granted by the patent office. Those claims are based, at least in part, on the description of the invention contained in the application to the patent office.

Not Every Invention Can Be Patented. U.S. patent law permits a patent to be granted only for inventions that are useful, new, and nonobvious. The first two requirements are fairly intuitive and sensible. One view of patents is that they are a bargain with inventors: The government grants a temporary monopoly on an invention, but only if it is both useful and represents an advance over our existing knowledge, which patent lawyers call the prior art. In exchange, the inventor must disclose the nature of the invention, which is described in the patent document itself. The third requirement, *nonobviousness*, is less clear. It rules out the patenting of an invention that would have been obvious to a practitioner in the relevant field at the time the invention was made. In other words, a patentable invention must be more than a trivial extension of the prior art.

Our patent law and many judicial decisions provide instructions on how the nonobviousness requirement should be applied. During the 1980s, a number of judicial decisions revised these instructions in significant ways. In practice, the modified test for nonobviousness is easier to satisfy than the one applied prior to the early 1980s. As a result many more inventions now qualify for patent protection.[8] Other judicial decisions made it easier for a patent holder to obtain a preliminary injunction — a court order prohibiting a potentially infringing activity even before the question of infringement is decided by the court. Today, the threat of a preliminary injunction often carries significant weight in negotiations between patent holders and alleged infringers.

---

[4] For recent reviews, see the reports by the Federal Trade Commission, the article by Stephen Merrill et al., and the article by Nancy Gallini.

[5] This article is based on Jim Bessen and Bob Hunt's 2004 working paper.

[6] We identify software firms as those companies included in Standard Industry Classification (SIC) 7372 (software publishers) as coded in Standard and Poor's *Compustat* database in 1999. For some purposes, we use a broader definition of software firms, that is, those classified in SIC 737 (computer programming, data processing, and other computer-related services).

[7] Today, a U.S. patent expires 20 years after the date of application. In the past, the patent term ran for 17 years from the date of grant.

[8] The changes in the 1980s were instituted by the Federal Circuit, a specialized appeals court for patent and certain other cases, created in 1982. For more information about these decisions and their effect on subsequent litigation, see the article by Adam Jaffe. The economic effects of reduced patentability standards are examined in Hunt's forthcoming article.

Subject Matter Exceptions. As a general principle, the American patent system is not designed to treat different kinds of inventions differently. For example, when Congress passed the 1952 Patent Act, the committee report endorsing the bill stated that the new law was meant to apply to "everything under the sun made by man."[9] These words are often mentioned in judicial decisions where a federal court is confronted with the problem of interpreting Congress's intent in drafting that law.

One exception to this rule was computer software. In the 1972 decision *Gottschalk v. Benson*, the Supreme Court ruled that the computer program in question was a mathematical algorithm and, therefore, unpatentable subject matter. But it did not take very long before new decisions began to blur this seemingly bright distinction between computer programs and other inventions. For example, in the 1981 decision in *Diamond v. Diehr*, the Supreme Court ruled that an invention incorporating a computer program could be patented as long as the new and nonobvious aspects of the invention did not consist entirely of the software. Even this distinction gradually eroded.[10]

Any real difference in the treatment of software and other inventions was eliminated after a 1994 appeals court decision (*in re Alappat*) upheld the patentability of a computer program that smooths digital data before displaying it as a waveform on a computer monitor. Shortly after that decision, the Patent and Trademark Office issued a comprehensive revision to its examination guidelines for computer-related inventions. Thereafter, the number of software patents granted increased dramatically (Figure 1).

TAKING A CLOSER LOOK AT SOFTWARE PATENTS

Despite considerable interest in the effects of granting patents on computer programs, there is no official list of software patents. The USPTO maintains a detailed system for classifying patented inventions by technology field — a sort of Dewey decimal system for patents. But there is no explicit classification for software inventions. Instead, researchers must devise their own ways of identifying software patents.[11] The data used in this article are based on a simple keyword search of the USPTO's database of patents issued after 1975. We looked for patents that used the words "software" or "computer program" in the description of the invention.[12]

According to this definition, about 1,000 software patents a year were granted in the early 1980s, increasing to about 5,000 a year in 1990. The rate had doubled again by 1996. Nearly 25,000 software patents were granted in 2002. This was a period of very rapid growth in patenting — the number of patents of any kind granted in 2001 was 1.7 times larger than in 1981 — but the growth in software patents was much larger still. As a result, the share of all patents counted as software patents increased from about 2 percent in the early 1980s to nearly 15 percent by 2002 (Figure 2).

[9] Senate Report No. 1979 82d Congress, 2nd Session (1952), p. 5.

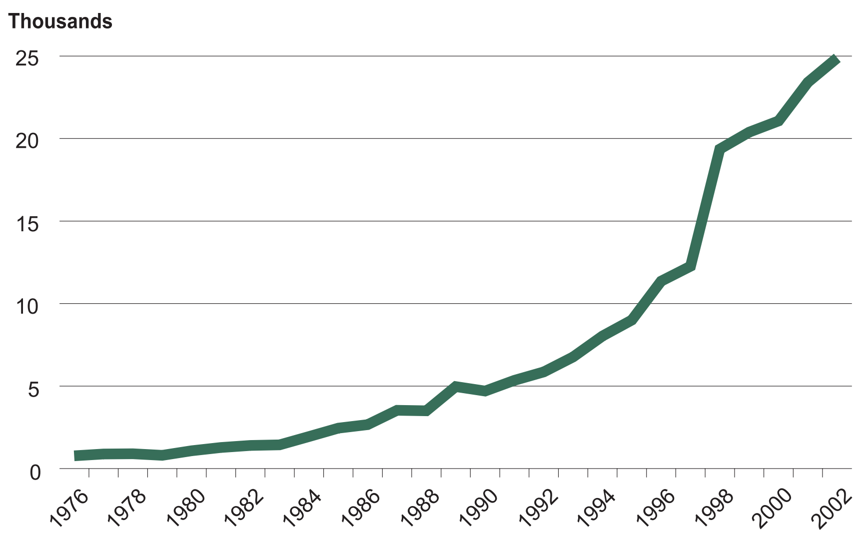
[10] For additional information on the changing treatment of software in patent law, see Hunt's 2001 article in the *Business Review*.

[11] For details on the different approaches, see the articles by John Allison and Mark Lemley, and by John Allison and Emerson Tiller, and the one by Stuart Graham and David Mowery.

[12] The exact search query is found in the Data Appendix. For a comparison of this definition and the resulting patent counts with others in the literature, see our working paper.

FIGURE 1

Software Patents Granted in the U.S. (1976-2002)



Source: U.S. Patent and Trademark Office and authors' calculations. Plots software patents by grant date.

Software Patents Are an American Phenomenon. We can learn something about inventors and the owners of patents by examining information contained in the patent document itself. This information reveals that software patents are a (relatively) home-grown phenomenon. During the 1990s, 70 percent of software patents were obtained by inventors living in the U.S.; that is significantly higher than the share of domestic inventors for all other patents (53 percent). Similarly, 70 percent of all software patents owned by companies went to firms headquartered in the U.S.; 51 percent of all other patents owned by companies went to American firms.[13]

Established Firms Obtain Most Software Patents. The typical owner of a software patent is a relatively large, well-established firm. During the 1990s, companies obtained a larger share of software patents than other patents (88 percent vs. 80 percent). To put it another way, individuals were relatively less likely to obtain their own software patent than a patent on another kind of invention.

We can also compare the financial characteristics of firms that obtain software patents and other kinds of patents. We obtained detailed financial data on several thousand U.S. firms from Standard and Poor's *Compustat* database, and using some existing databases and our own research, we matched patents to those firms.[14] We then used this information to compare the median firm ranked in terms of (1) the number of software patents obtained and (2) the number of other patents obtained.[15] The median firm ranked in terms of software patents is much larger than the median firm ranked by other patents. If size is measured in terms of market value, the median software patentee is twice as large as the median patentee of other inventions ($24 million vs. $12 million). Measured in terms of sales, it is 50 percent larger ($13 million vs. $9 million). Measured in terms of spending on research and development, it is 68 percent larger ($956 million vs. $376 million).

Most Software Patents Don't Come from the Software Industry. We were surprised to find that the vast majority of software patents are not obtained by firms associated with computer software. In the second half of the 1990s, firms in the software industry received 1 percent of all patents granted to firms included in the *Compustat* file and at most 7 percent of all software patents (Table).[16] Manufacturers accounted for three out of four software patents. The top five firms in terms of software patents obtained in 1995 were IBM, Motorola, Hitachi, AT&T, and Hewlett-Packard. Nine of the top 10 firms ranked by software patents received in 1995 were on the list of the top 20 firms ranked by patents of any kind.

[13] The USPTO data indicate the owner of a patent at the time it was issued. The owner may be the individual(s) who made the invention or an organization (assignee), such as a firm or a government agency.

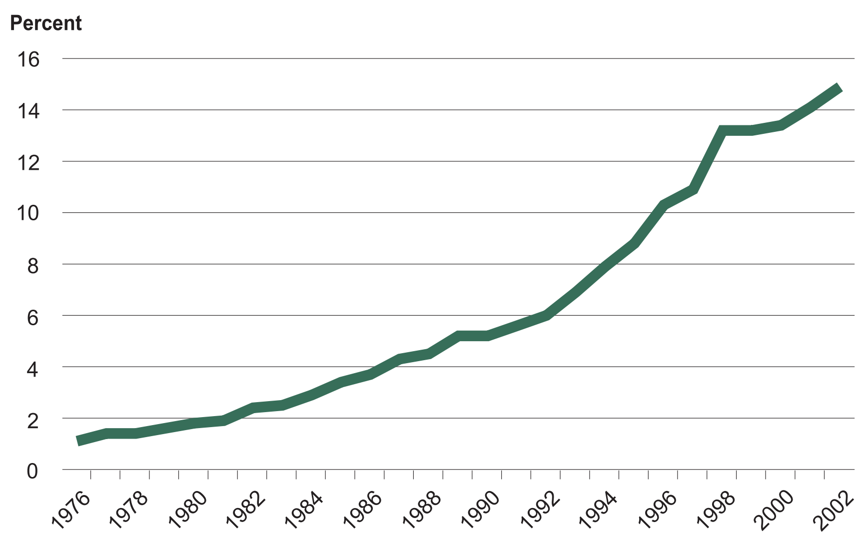
[14] *Compustat* includes information on virtually all firms that file 10-K and 10-Q reports with the U.S. Securities and Exchange Commission. Our matching of patents to firms is based primarily on information contained in the NBER Patent Citations Data File and data generously provided to us by Tony Breitzman of CHI Research. For details on the matching process, see the Data Appendix.

[15] The median identifies the firm where 50 percent of all firms have more patents than it does and 50 percent of all firms have fewer patents than it does.

[16] These statistics are for successful patents applied for during 1994-97. For this calculation, the software industry is defined as firms included in the SIC 737, but excluding IBM, which alone accounted for 6 percent of software patents granted. We treat IBM separately to rule out the possibility that these patterns are the result of a single, large company's activity.

FIGURE 2

Software Patent Share in the U.S. (1976-2002)



Source: U.S. Patent and Trademark Office and authors' calculations.
Plots the percentage of patents granted in a year that are software patents.

TABLE

The Distribution of Software Patents (1994-97)

<table>
<tr><th></th><th>(1)
Share of all software patents</th><th>(2)
Share of all programmers</th><th>(3)
Share of all programmers & engineers</th><th>(4)
Share of all patents</th><th>(5)
All patents/ R&D</th><th>(6)
Software patent propensity</th></tr>
<tr><td>Manufacturing</td><td>75%</td><td>11%</td><td>32%</td><td>88%</td><td>3.8</td><td></td></tr>
<tr><td>Chemicals (SIC 28)</td><td>5%</td><td>1%</td><td>2%</td><td>15%</td><td>2.5</td><td>1.5</td></tr>
<tr><td>Machinery (SIC 35)</td><td>24%</td><td>3%</td><td>7%</td><td>17%</td><td>4.2</td><td>4.4</td></tr>
<tr><td>Electronics (SIC 36)</td><td>28%</td><td>2%</td><td>7%</td><td>27%</td><td>6.8</td><td>9.6</td></tr>
<tr><td>Instruments (SIC 38)</td><td>9%</td><td>1%</td><td>4%</td><td>11%</td><td>7.1</td><td>8.7</td></tr>
<tr><td>Other manufacturing</td><td>9%</td><td>5%</td><td>13%</td><td>18%</td><td>2.3</td><td>1.9</td></tr>
<tr><td>Nonmanufacturing</td><td>25%</td><td>89%</td><td>68%</td><td>12%</td><td>3.0</td><td></td></tr>
<tr><td>Software publishers (SIC 7372)</td><td>5%</td><td rowspan="2">33%</td><td rowspan="2">18%</td><td>1%</td><td>1.0</td><td rowspan="2">1.0</td></tr>
<tr><td>Other software*</td><td>2%[#]</td><td>1%[#]</td><td>2.8[#]</td></tr>
<tr><td>Other nonmanufacturing</td><td>4%</td><td>55%</td><td>49%</td><td>4%</td><td>3.4</td><td>3.8</td></tr>
<tr><td>Addendum: IBM</td><td>6%</td><td>—</td><td>—</td><td>2%</td><td>5.0</td><td></td></tr>
</table>

Notes: This table is based on patents issued by the U.S. Patent and Trademark Office applied for during the years 1994-97 and matched to a firm in the *Compustat* data set. Data on computer programmers are from the Bureau of Labor Statistics Occupational Employment Survey (various years) and the numbers include system analysts. The fifth column reports patents granted per $10 million of R&D in 1996 dollars. The last column reports the relative patent propensity (for software patents) estimated in the statistical analysis contained in Bessen and Hunt's 2004 working paper and described on page 27. The numbers in column 6 are presented relative to the estimated software patent propensity of firms contained in the business services sector (SIC 73). For example, the estimated software patent propensity for the chemical industry is 1.5 times that for SIC 73.

*Firms in SIC 73, excluding those firms in SIC 7372.

[#]Excludes IBM.

Firms in just three manufacturing industries (machinery, electronics, and instruments) alone accounted for 66 percent of software patents granted to firms — a number that significantly exceeds their impressive 54 percent share of patents of any kind. These numbers are even more remarkable when we examine the distribution of computer programmers across these industries.[17] These are presumably the workers responsible for creating most new computer programs. Manufacturers of machinery, electronics, and instruments employed only 6 percent of all computer programmers and yet they obtained 66 percent of software patents. Firms outside the manufacturing sector employed 90 percent of computer programmers, but together they accounted for only 25 percent of software patents. It would appear that the distribution of software patents across industries reflects something other than the creation of software.

[17] Our data on computer programmers come from various editions of the *Occupational Employment Survey*, published by the Bureau of Labor Statistics. We thank Joseph Bush of the BLS for his assistance.

WHY ARE THERE SO MANY SOFTWARE PATENTS?

The previous section shows that firms obtain many software patents today, but they either could not, or did not wish to, obtain them in the past. Of course, the software sector of the economy has also grown rapidly over time. But these explanations tell us very little about why firms obtain software patents, and they potentially exaggerate the effects of legal changes by ignoring economic and other factors that may have contributed to the explosion in software patenting. Let's look at the differences in the software-patenting behavior of firms across industries and over time, and let's look for any relationships between a firm's software patenting behavior and its R&D investments.

Estimating the Propensity to Patent. Industries vary significantly in their propensity to patent — that is, the average number of patents obtained from a given amount of resources spent on developing new products and processes. For example, during the mid-1990s, firms in the machinery, electronics, and instrument industries received between four and seven patents (of any kind) for every $10 million in R&D they spent (see column 5 of the Table). That compares with only about one patent per $10 million in R&D for firms in the software industry. Based on this simple calculation, all else equal, if we observed a $10 million increase in R&D in each of these industries, we would expect to see four to seven more patents by manufacturers of machinery, electronics, and instruments and one additional patent by software companies.

A more sophisticated analysis shows that firms apply for more software patents when they are both more research-oriented and more capital-intensive and when the industry workforce consists of more programmers and engineers.[18] We did not find a difference in the propensity to patent software between old and young firms except in the software industry. There, new firms have a significantly lower propensity to patent software than older firms in the same industry.[19]

Manufacturers, in general, have a much higher propensity to patent software than do firms in the software industry (see column 6 of the Table). After we account for R&D and other factors, firms in the machinery, electronics, and instruments industries obtain software patents at a rate four to 10 times higher than firms in the software sector. In addition, the propensity to patent software is significantly higher for firms in industries in which their peers obtain more patents (of all kinds) per employee. In short, the pattern of software patenting across U.S. firms seems to be more closely related to industry-wide variations in the utilization of patents in general than to the resources devoted to creating software.

Manufacturers, in general, have a much higher propensity to patent software than do firms in the software industry.

The Rise in the Propensity to Patent Software Over Time. The average annual increase in the number of successful applications for software patents between 1987 and 1996 was 16 percent. Our analysis shows that changes in firms' R&D and capital investments, employment of programmers, and other factors explain about one-third of the rate of increase in software patents. The remaining two-thirds (about 11 percentage points) represent an increase in the propensity to patent software over time. Compared with the rate for 1987, and holding all other factors constant, firms were successfully applying for 50 percent more software patents in 1991 and more than 150 percent more by 1996.

It is likely that a good part of this increase in the propensity to patent is the result of changes in the legal treatment of software patents.[20] Such changes might work in two ways. The cost of obtaining software patents relative to the cost of obtaining any other patent may have fallen. Alternatively, the economic benefit conferred by a typical software patent, again relative to the benefit conferred by any other kind of patent, may have increased. Or both of these may be true. In other words, our analysis suggests that the relative profitability conferred by obtaining software patents increased over time.

DO MORE SOFTWARE PATENTS MEAN MORE R&D?

Ordinarily, when a firm obtains additional patents, the profits

[18] This section is based on a regression analysis where we controlled for firm and industry characteristics and allowed for the possibility that patent propensity has changed over time. Details may be found in our 2004 working paper. For an excellent example of this methodology applied to the American semiconductor industry, see the article by Bronwyn Hall and Rosemarie Ziedonis

[19] A new firm in our analysis represents the first five years that a company is reported in *Compustat*.

[20] We cannot attribute all of this residual increase to legal changes because the pattern can also be explained by productivity growth, that is, increases in the number of inventions per programmer. Our review of the available studies suggests that any reasonable estimate of productivity growth in software would explain less than half of the residual increase in the propensity to patent software over time.

earned on its inventions should rise. This should encourage the firm to engage in more R&D. Similarly, when a firm engages in more R&D, it should invent more, and that should make it easier to get additional patents. This is the traditional incentive theory of patents: By granting firms more and stronger property rights — that is, the right to capture more profits from their R&D investments — the government can stimulate innovation.

Complements or Substitutes? When economists think about a problem like this, they often inquire whether the variables in question (for example, R&D and software patents) are *complements or substitutes*. In the standard textbook exposition, two goods are complements if a *fall* in the price of one good induces an *increase* in the consumption of the other. An example of two goods that might act as complements might be coffee and cream. On the other hand, two goods are substitutes when a *fall* in the price of one good causes a *decrease* in the consumption of another good. An example of two goods that might act as substitutes would be public transit and automobiles.

This intuition about complements and substitutes also applies to a firm's choice of inputs. For example, if the cost of information technology (IT) declines, it is entirely possible that a firm will purchase more of the technology *and* hire more computer programmers who are skilled in using that technology. If that did happen, we would say that IT equipment and computer programmers are complementary inputs.[21] If, on the other hand, we observed a decline in the number of computer programmers, we might conclude that IT equipment has substituted for computer programmers, who have become more expensive relative to the cost of IT equipment.

To economists, then, the conjecture that making software patents easier to obtain will increase investments in R&D is a claim that these patents and R&D are complementary inputs in the production of profitable innovations. All else equal, the legal changes described earlier increased the return from obtaining software patents relative to other patents. We have already seen that one response was a very large increase in the number of such patents (the quantity demanded rose as the cost of software patents fell), even after we've controlled for other factors. But what has the effect been on demand for R&D?

The Relationship Between R&D and Software Patents Has Changed. To answer this question, we can examine the relationship between changes in a firm's research intensity (typically measured by the ratio of R&D to sales) and changes in the firm's focus on *software patents* — its new software patents divided by all its new patents — over time.[22] An increasing focus on software patents should reflect a decline in the firm's cost of obtaining software patents relative to other patents. A positive correlation between changes in research intensity and changes in focus on software patents would suggest that software patents and R&D are indeed complementary inputs.[23]

Our research shows that, all else equal, during the late 1980s, firms that increased their focus on software patents tended to increase their R&D intensity, but the relationship was weak. In other words, more likely than not, software patents and R&D were complementary inputs during the 1980s. For the 1990s, we found a much stronger relationship, but it was negative: All else equal, increases in share of software patents were associated with decreases in research intensity. This suggests that in the 1990s, software patents substituted for R&D.[24]

This effect is concentrated in the machinery (including computers) and electronics (including semiconductors) industries and the software industry broadly defined[25] — in other words, among the industries that account for

By granting firms more and stronger property rights — that is, the right to capture more profits from their R&D investments — the government can stimulate innovation.

[21] Note that the total value of IT equipment purchased might rise or fall depending on how much more IT equipment is purchased in response to the drop in price. A firm's demand for a good is said to be elastic when a decline in price, expressed as a percent change, induces a larger increase, again expressed as a percent change, in the quantity demanded.

[22] In our statistical analysis, we examined the relationship between changes in firms' R&D intensity and changes in their focus on software patents over five-year intervals. We controlled for changes in input prices (including information technology), size of the firm, new vs. established firms, employment of computer programmers, and idiosyncratic factors specific to the firm or industry.

[23] A positive correlation means that increases in focus on software patents are associated with increases in R&D intensity and decreases in focus on software patents are associated with decreases in R&D intensity.

[24] We also found evidence that this negative correlation had become even more negative by the late 1990s, but we cannot be certain this is true.

[25] Here we count all firms in SIC 73 (Business Services) as software, and this includes IBM. If we exclude IBM from SIC 73, we do not find a systematic relationship between increases in focus on software patents and changes in R&D intensity among the remaining firms in SIC 73.

about two-thirds of all software patents. Outside of those industries, there was no systematic relationship, during the 1990s, between an increase in focus on software patents and changes in firms' R&D intensity.

Overall, the effect is economically significant. Taking the analysis literally, if the number of software patents grew only as rapidly as all other patents after 1991, the average R&D intensity of U.S. firms would be about 7 percent higher than was actually recorded in 1997. This represents about $9 billion in additional private R&D spending for the entire U.S. economy. It also represents about five years of the annual average increase in the research intensity of American firms since 1953.

But it is important to emphasize that the analysis does not identify the exact relationship that explains why an increased focus on software patents is associated with a decline in research intensity. In the language of statistics, this approach identifies a correlation but not causation. Still, we can compare the patterns identified in the data with a number of hypotheses about the effects of software patents.

RECONCILING THEORIES WITH THE DATA

While we can't provide a full explanation of what happened, we can compare the facts identified so far with a variety of hypotheses or theoretical arguments that appear in the debate over changes to the U.S. patent system.

The Incentive Theory. The first of these is the traditional incentive theory, which argues that by making available stronger property rights at lower cost, firms will have an increased incentive to engage in R&D. This conventional wisdom is often cited in arguments that favor extending patent protection to computer programs in Europe. Is our evidence consistent with this theory?

The answer seems to be no. We observe that the vast majority of software patents are obtained by firms outside the software industry and with little investment in the inputs (computer programmers) required to develop software inventions. The distribution of software patents seems to

The increases in the total number of software patents and in share of software patents are consistent with firms' responding to a decline in the relative cost of obtaining these patents or, alternatively, an increase in their cost effectiveness.

follow more closely the general pattern of industry-wide propensities to patent than anything peculiar to software in itself. In general, industries known for prodigious patenting account for the vast majority of software patents obtained. Firms located in such industries have a higher propensity to patent software.

The increases in the total number of software patents and in the share of software patents are consistent with firms' responding to a decline in the relative cost of obtaining these patents or, alternatively, an increase in the their cost effectiveness. But the negative correlation between increases in firms' focus on software patents and their R&D intensity in the 1990s suggests that firms may be substituting for R&D with software patents.

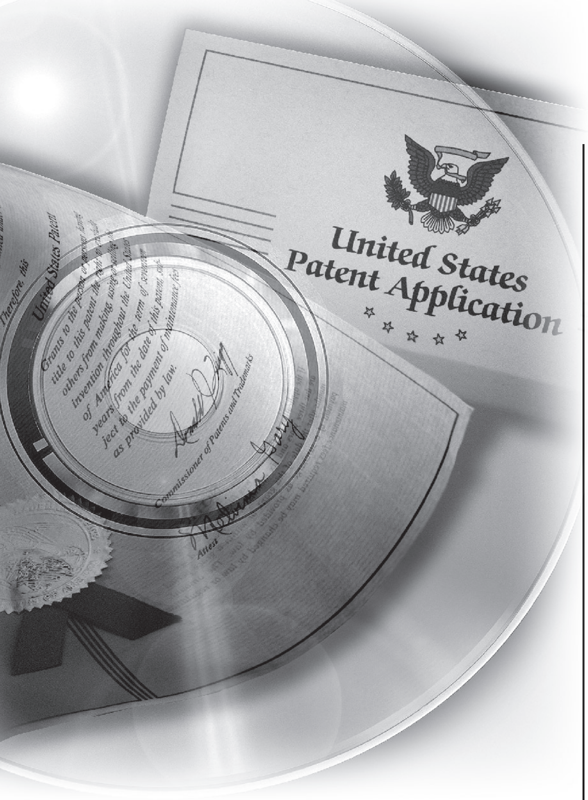
A Productivity Shock. Another hypothesis is that the U.S. economy has experienced a large productivity shock that favored inventions implemented via computer programs. Such a shock would be consistent with a large increase in software patenting and the long-run trend toward higher research intensity among American firms. But it is inconsistent with the negative correlation between increases in share of software patents and R&D intensity. What's more, the observed increase in the propensity to patent software seems too large to be explained entirely by advances in the productivity of computer programming.

One variation on the productivity-shock hypothesis points to the potential for outsourcing of software development as the market for prepackaged software expanded. In other words, firms might have chosen to purchase software rather than to develop it internally. Such outsourcing could explain a decline in research intensity, but firms outsourcing their software development would also likely reduce their focus on software patenting. Conversely, software developers that benefit from outsourcing might be expected to increase both their R&D investments and software patenting. Neither of these patterns is consistent with the data.

It has also been suggested that the use of software in the R&D process significantly reduces the cost of doing R&D, and this might explain the observed negative correlation between focus on software patents and R&D intensity. But previous studies have shown that firms respond elastically to changes in the cost of doing R&D.[26] In other words, the quantity

---

[26] See, for example, the article by Philip Berger and the article by Bronwyn Hall and John van Reenan. In addition, our analysis takes into account changes in the cost of software and the share of computer programmers in the industry workforce.



of R&D that firms engage in increases by at least as much, in percentage terms, as the decrease in its cost. Thus, even if software reduces the cost of research and development, R&D intensity should not fall and might even increase.

Patent Thickets. In contrast to the incentive theory already described, suppose instead that firms in an industry assemble large patent portfolios in order to extract royalties from competitors and to defend themselves from similar behavior by their rivals. Economists have come to describe such an environment as a patent thicket.[27] In theory at least, extensive competition in patents, rather than inventions, may occur if firms rely on similar technologies and the cost of assembling large portfolios is not very high. In such an environment, firms may compete to tax each others' inventions — for example, by demanding royalties — and, in the process, reduce their competitors' incentive to engage in R&D.[28]

The outcome of patent litigation and licensing agreements often depends on the size of the firm's patent portfolio. This creates an incentive for firms to build larger patent portfolios, especially when their rivals focus on patents as a competitive strategy. Economists sometimes describe this type of environment as a *prisoner's dilemma*.[29] All firms would be better off if they did not act in this way, but each firm would be worse off if it did not respond to a surge in patenting by their rivals. Under these circumstances, firms may find themselves competing in court, rather than in the marketplace.

The changing legal treatment of software patents might explain a systematic change in the behavior of some firms. During the early 1980s, patents were relatively costly to obtain, and this might have discouraged substitution away from R&D and toward strategic patenting. By the mid-1990s, software patents became a relatively inexpensive way to expand patent portfolios. This may have increased the attractiveness of a strategy that emphasizes patent rights over a strategy based on R&D.

The patent thicket explanation is consistent with the observed rise in propensity to patent and the negative relationship between changes in share of software patents and research intensity in certain industries. It might also explain why software patents are more common in industries with high propensities to patent (machinery, electronics, and instruments) rather than in industries that focus primarily on developing software. Also, it is consistent with the observation that the propensity to patent is higher in industries in which firms obtain more patents per employee.

CONCLUSION

Nearly 50 years ago, scholar Fritz Machlup presented the results of his study on the efficacy of the patent system to the U.S. Congress. He concluded: "If one does not know whether a system as a whole (in contrast to certain features of it) is good or bad, the safest policy conclusion is to muddle through....If we did not have a patent system, it would be irresponsible...to recommend instituting one. But since we have had a patent system for a long time, it would be irresponsible, on the basis of our present knowledge, to recommend abolishing it."

What would Machlup say about a significant expansion of the patent system and a significant change in patentability standards, instituted in the absence of much evidence about the likely effects? Yet this is precisely what has happened in the U.S. over the last quarter of a century.

These changes are often justified on the basis of conventional wisdom: Granting more and stronger property rights will necessarily stimulate innovation. Our evidence suggests this assumption may be incorrect in the case of software patents. If, instead, the legal changes create patent thickets, the result might well be less innovation.

---

[27] For evidence of this phenomenon in the electronics and semiconductor industries, see the article by Peter Grindley and David Teece and the one by Bronwyn Hall and Rosemarie Ziedonis.

[28] For a theoretical model of this intuition, see the 2003 working paper by James Bessen.

[29] The term is derived from the example of two suspects arrested and interrogated separately. If they both remain silent, the prosecutor has little evidence, and each will receive a small penalty. If one suspect rats and the other doesn't, the rat will reduce his own punishment, but the silent one will be punished severely. Knowing each other's incentives, both suspects rat on each other.

Data Appendix

Identifying Software Patents

We count as a software patent any utility patent (excluding reissues) granted after 1975 that satisfies the following conditions:

1. The terms *software* or *computer* and *program* appear in the specification;
2. The terms *antigen*, *antigenic*, and *chromatography* do not appear in the specification; and
3. The terms *chip*, *semiconductor*, *bus*, or *circuit* or *circuitry* do not appear in the title.

Using this algorithm, we identified 130,650 software patents granted in the years 1976 to 1999. For a comparison of this definition, and the resulting patent counts, with others used in the literature, see our working paper.

Matching Patents to Firms

Our statistical analysis relies on the matching of patents to companies in Standard and Poor's *Compustat* database. The majority of our matches are obtained from the NBER Patent Citations Data File (for details on that resource, see the working paper by Bronwyn Hall, Adam Jaffe, and Manuel Trajtenberg). We supplemented those matches using information graciously provided to us by CHI Research (for information about these data, see www.chiresearch.com/information/customdata/patdata.php3.

Both of these sources link a numeric assignee number issued by the USPTO with an alphanumeric CUSIP code that can be used to identify firms contained in *Compustat*. In addition, we also matched the patents of the 25 largest publicly traded software firms and 100 other large R&D performers not already matched in the data provided by the other sources.

The resulting data set includes patents matched to 4,792 distinct subsidiaries and 2,043 parent firms. Over the period 1980-99, our sample accounts for 68 percent of successful U.S. patent applications by domestic nongovernment organizations (mostly corporations) and 73 percent of software patents granted to these organizations. The matched firms accounted for 91 percent of R&D spending reported by U.S. firms in *Compustat*. These coverage ratios are quite stable over the two decades.

Still, only 37 percent of R&D performers in the *Compustat* data set were matched to their patents. This suggests the possibility of *selection bias* in some of our results, because firms successfully matched to their patents may somehow be systematically different from firms not matched. In particular, our coverage of the smallest and newest firms in *Compustat* is not likely to be as good as our coverage of larger and older firms. We conducted a number of statistical tests for selection bias and found this possibility had little or no effect on the results reported here. For details, see our working paper.

REFERENCES

Allison, John R., and Mark A. Lemley. "Who's Patenting What? An Empirical Exploration of Patent Prosecution," *Vanderbilt Law Review*, 58, 2000, pp. 2099-2148.

Allison, John R., and Emerson H. Tiller. "An Empirical Analysis of Internet Business Method Patents," mimeo, University of Texas at Austin, 2001.

Ashish, Arora, Marco Ceccagnoli, and Wesley M. Cohen. "R&D and the Patent Premium," NBER Working Paper 9431 (January 2003).

Berger, Philip. "Explicit and Implicit Effects of the R&D Tax Credit," *Journal of Accounting Research*, 31, 1993, pp. 131-71.

Bessen, James. "Patent Thickets: Strategic Patenting of Complex Technologies," ROI Working Paper, 2003.

Bessen, James, and Robert Hunt. "An Empirical Look at Software Patents," Federal Reserve Bank of Philadelphia Working Paper 03-17/R, March 2004.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," NBER Working Paper 7552 (2000).

Federal Trade Commission. *To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy.* Washington, DC: Federal Trade Commission, 2003.

Gallini, Nancy T. "The Economics of Patents: Lessons from Recent U.S. Patent Reform," *Journal of Economic Perspectives*, 16, 2002, pp. 131-54.

Graham, Stuart J.H., and David C. Mowery. "Intellectual Property Protection in the U.S. Software Industry," in Wesley M. Cohen and Stephen A. Merrill, eds., *Patents in the Knowledge-Based Economy,* Washington, D.C.: National Academies Press, 2003, pp. 219-58.

Griliches, Zvi. "Productivity, R&D, and the Data Constraint," *American Economic Review,* 84, 1994, pp. 1-23.

Grindley, Peter C., and Teece, David J. "Managing Intellectual Capital: Licensing and Cross-Licensing in Semiconductors and Electronics," *California Management Review,* 39, 1997, pp. 8-41.

Guellec, Dominique, and Bruno van Pottelsberghe de la Potterie. "R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries," OECD STI Working Paper 2001/3, 2001

Hall, Bronwyn H., and John van Reenen. "How Effective Are Fiscal Incentives for R&D? A Review of the Evidence," NBER Working Paper 7098, 1999.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498, 2001.

Hall, Bronwyn H., and Rosemary Ziedonis. "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995," *RAND Journal of Economics*, 32, 2001, pp. 101-28.

Hunt, Robert M. "Patent Reform: A Mixed Blessing for the U.S. Economy?" Federal Reserve Bank of Philadelphia *Business Review* (November/December 1999), pp.15-29.

Hunt, Robert M. "You Can Patent That? Are Patents on Computer Programs and Business Methods Good for the New Economy?" Federal Reserve Bank of Philadelphia *Business Review,* First Quarter, 2001, pp. 5-15.

Hunt, Robert M. "Patentability, Industry Structure, and Innovation," *Journal of Industrial Economics* (forthcoming).

Jaffe, Adam B. "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process," *Research Policy*, 29, 2000, pp. 531-57.

Kortum, Samuel, and Josh Lerner. "What Is Behind the Recent Surge in Patenting?" *Research Policy*, 28, 1999, pp. 1-22.

Levin, Richard C., Alvin Klevorick, Richard Nelson, and Sidney Winter. "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3, 1987, pp. 783-831.

Machlup, Fritz. *An Economic Review of the Patent System*. Study No. 15, U.S. Senate Committee on the Judiciary, Subcommittee on Patents, Trademarks and Copyrights. Washington: U.S. Government Printing Office, 1958.

Merrill, Stephen A., Richard C. Levin, and Mark B. Myers, eds., *A Patent System for the 21st Century.* Washington: National Academies Press, 2004, www.nap.edu

National Science Foundation. "Research and Development in Industry: 2000," Arlington, VA: National Science Foundation, Division of Science Resources Statistics, 2003.

USPTO. "Examination Guidelines for Computer-Related Inventions," *Federal Register*, 61, 1996. pp. 7478-92.

Cases

Diamond v. Diehr 450 U.S. 175 (1981).

Gottschalk v. Benson, 409 U.S. 63 (1972).

In re Alappat, 33 F.3d 1526 (1994).

CEOs, Clerks, Computers, and the Rise of Competition in the Late 20th Century

BY LEONARD NAKAMURA

A new era of heightened creative destruction that began in the late 1970s also ushered in a new era of heightened competition. Such intensified competition has made leaders of large industrial enterprises vulnerable to a level of uncertainty previously reserved for managers of small and new firms. Consequently, managerial careers now less often have benign endings. In addition, CEOs have become less reliable servants of the corporation. The story was much different during the previous 100 years. From the 1870s to the 1970s, the large industrial corporation was highly stable. Its stability was derived, in part, from investments in a corporate structure that centered on a sales and administrative staff. In this article, Leonard Nakamura argues that the electronics revolution of the 1970s sharply reduced the value of this corporate asset and made corporations more susceptible to competition.

The relationship between the modern corporation and its leadership has undergone a sea change over the past 30 years. The large industrial corporation appeared to be a serene republic, protected like 14th century Venice by a beneficent nature and as likely to exist for centuries. The leader of this corporation — the chief executive — was a sort of philosopher king, dealing with long-term strategic issues and delegating most of the direct oversight of the corporation's daily operations to the chief operating officer.

In his introduction to A.P. Sloan's memoir, *My Years with General Motors*, Peter Drucker described the General Motors' president's relationship to the founding of Chrysler Motors: "Chrysler started the automobile company that bears his name in large part because Sloan pointed out to him the opportunity created by the decline of the Ford Motor Company in the mid-twenties, but also because Sloan clearly saw that with Ford rapidly going downhill, GM, in its own interest, needed a strong competitor." This passage forcefully calls to mind a different era, a time when corporate managers were so secure in their corporations' solidity that they could deliberately encourage the growth of a formidable competitor. This is a far cry from GM's aggressive and successful legal pursuit of European purchasing manager Jose Ignacio Lopez when he left General Motors to work for Volkswagen in 1993. (GM forced Volkswagen to fire Lopez and recompense GM for the secrets Lopez took with him.)

There have been sweeping changes in the corporate environment since the late 1970s, when the new era of heightened creative destruction emerged.[1] In this new era, heightened competition has made leaders of large industrial enterprises vulnerable to a level of uncertainty previously reserved for managers of small and new firms. Chief executives of Fortune 500 corporations were once seen as professional, farsighted, and deeply interested in the longevity of their firms; now they find their ability to control their destinies substantially diminished.

Leonard Nakamura is an economic advisor and economist in the Research Department of the Philadelphia Fed.

---

[1] Economist Joseph Schumpeter coined the phrase creative destruction. Schumpeter theorized that creativity was the prime engine in a modern economy and profits were the fuel. Allowing creative workers to temporarily capture monopoly profits — for example, by granting patents — promotes "creative destruction": new goods and livelihoods replace old ones. See my *Business Review* article from 2000.

This heightened competition — firms' ability to enter and conquer markets via new products and processes — need not imply small firms or profits. It may well involve a series of firms that rapidly prosper and attain substantial profits and size, such as Microsoft or Merck. But attaining and maintaining profits has become more hazardous work. Managers and shareholders find their firms increasingly at risk, both from outside competitors and from insiders who threaten to take their talents and become competitors.

The upshot has been that managerial careers now less often have benign endings. Forced resignations of corporate CEOs have become more common, and more CEOs are hired from outside the firm, rather than being promoted from among the ranks of its senior executives (see the study by Huson, Parrino, and Starks). Corporate leaders have found themselves in a harsher economic environment, their jobs riskier, their boards more demanding.

In their turn, CEOs may have become less reliable agents for their shareholders and boards of directors. It should not be surprising if CEOs demand higher wages to compensate for their greater risk. Nor is it a great stretch to imagine that boards must deepen their oversight over CEOs in order to maximize shareholder value in this new environment. Indeed, although this is a conjecture that has not been researched, the recent spectacular examples of corporate fraud may partly be a byproduct of the decline in solidarity between shareholders and their corporate employees.

The story was much different in the previous century. The large industrial corporation was highly stable from the 1870s to the 1970s. According to Alfred Chandler, Harvard Business School's august historian of business, this stability was based both on tangible investments in plant and equipment and on investments in a corporate structure whose core purpose was information processing: the sales and administrative staff. I argue that the electronics revolution in the 1970s sharply reduced the value of this corporate asset and made corporations far more vulnerable to competition.

CREATORS AND CLERKS: OCCUPATIONAL CHANGE FROM 1900 to 1980

Chandler wrote two influential tomes on the rise and stability of the great industrial corporation: *The Visible Hand* and *Scale and Scope*. Chandler showed that many of the corporations that were the first modern large producers in their industries in the late 19th and early 20th centuries were still leading their industries in the 1970s. He attributed this longevity to three complementary investments:

Investment 1: Production. The first investment was in a large, scale-economy production facility. Typically, the scale of this plant was substantially larger than its competitors', enabling substantially cheaper production and larger profits. In several cases, a few plants accounted for a substantial fraction of the industry's total capacity. Of course, this physical investment was not easily expropriated by employees of the firm. One cannot easily imagine a disgruntled group of managers resigning en masse and taking with them an oil-refining complex or a section of railroad, complete with trains!

This physical investment was buttressed by additional investments in information, or rather in corporate employees whose collective task was to process information.

Investment 2: Sales. One task was collecting information about the corporation's customers. A large and efficient production facility, after all, is valuable only if its immense output can be sold. This required a disciplined, intelligent, and well-trained sales force. The sales force was the eyes, ears, and voice of the corporation for its customers, in a period when orders, invoices, and payments were processed by hand or typed.

Investment 3. Coordination. The corporation also needed to create a management and clerical team that coordinated sales and production decisions and accounted for every order, invoice, and item in inventory. Within the extended bureaucratic hierarchy of the great corporation, accurate financial and operational data were crucial to permitting individual units to act autonomously while remaining accountable to the firm as a whole. It also enabled management to identify problems promptly, allocate resources overall, and plan for future ventures. This coordinating mechanism was the nervous system of the great industrial corporation.

Within the extended bureaucratic hierarchy of the great corporation, accurate financial and operational data were crucial to permitting individual units to act autonomously while remaining accountable to the firm as a whole.

Chandler showed that these three complementary investments formed a barrier to entry that few potential entrants could surmount. One way of articulating Chandler's argument is that the sales force and coordinating management protected the production facility investment against technological innovation. For example, suppose a new technological innovation made a market leader's production facility obsolete. Before an entrant could take full advantage of the innovation, it would have to build up a sales force and management bureaucracy, an expensive and time-consuming affair. In the intervening time, the incumbent could usually duplicate the innovation and keep most of the business for itself.[2]

One example Chandler gives is the Standard Oil Company, whose Cleveland refineries in 1870 were the largest in the world. Standard Oil's large volume enabled it to garner special deals from railroads and to develop an alliance of oil refiners that controlled much of the kerosene output of the U.S. To break the power of the Standard Oil alliance, crude-oil producers set up the rival Tidewater Oil Company with new technology: Tidewater built a huge pipeline from northwest Pennsylvania to Bayonne, New Jersey, where the company eventually built a massive refinery. But the Standard Oil alliance was able to maintain overall dominance of the kerosene market because of its greater organizational capabilities. It constructed its own pipeline, even though doing so required an investment 10 times its previous capitalization, and built new, even larger refineries because it maintained its domination over kerosene sales. Indeed, Tidewater was dependent on Standard Oil to market Tidewater's kerosene in Europe. As Chandler laconically puts it, "Not surprisingly, Tidewater soon came under the financial control of Standard Oil."[3]

Although challengers to the market leaders did appear, the number of players in many industries where these barriers to entry existed remained small, and there was little turnover among the leaders. The corporations that led industries in 1890 or 1915 often still led them in 1975. As these corporations grew — as Standard Oil grew from kerosene to gasoline, as Ford grew from Model Ts to Tauruses, and as DuPont grew from gunpowder to synthetics — so did their corporate and sales staffs.[4]

In addition, corporate managers developed a new professional attitude. In discussing the development of the railroads, Chandler noted that "because of the special skills and training required and the existence of a managerial hierarchy, the railroad managers came to look on their work as much more of a lifetime career than did the plantation overseer or the textile mill agent."[5]

As corporations became arenas for professional advancement, the career concerns of management bent the interests of the corporation toward corporate stability. Indeed, in the 1960s and 1970s, as Gordon Donaldson has written, "The essence of the corporate mission … at many … companies of that day was the concept of the individual corporation

As corporations became arenas for professional advancement, the career concerns of management bent the interests of the corporation toward corporate stability.

as an independent and self-sustaining economic and financial entity within which all primary constituent interests, including shareholders, could fulfill their economic objectives. Growth, diversification, and a higher degree of independence from the public capital markets were essential ingredients of long-term economic self-sufficiency."[6] In that era in which diversification of shareholders was still limited, reducing corporate risk was often seen as valuable to the shareholders as well as to employers.

Donaldson also noted another aspect that characterized the corporation then: "Jobholders at all levels have traditionally looked to the individual corporation as the source of their lifetime economic welfare. It was an expectation to be encouraged, since an unconditional commitment from

---

[2] The existence of entry barriers did not imply, of course, complete freedom from failure for the incumbent. A large corporation could fail to take advantage of its market position and ignore rather than adopt and improve upon rivals' innovations. Indeed, Ford Motor Company failed to follow General Motors' market segmentation strategy for over two decades and might well have disappeared. But the very fact that Ford could turn itself around after World War II is testimony to the great stability of large corporations during most of this period.

[3] See Chandler, *Scale and Scope*, p. 95.

[4] Alfred Chandler's vision of the rise of the American corporation remains the central academic interpretation. While Chandler's work has been criticized, it has largely stood the test of time, and most of the critiques to date have modified rather than overthrown its basic themes. See, for example, Louis Galambos's review article on "The U.S. Corporate Economy in the Twentieth Century," and its bibliographic note, in the Cambridge *Economic History of the United States*.

Some economic historians have emphasized the role of government in helping maintain and develop the market power of the great industrial corporations. In this view, government regulation of public utilities and transportation, as exemplified by the regulated monopoly of the Bell System, provided crucial support for the large industrial corporation. But government regulation itself is likely to stifle productivity rather than enhance it. We have to turn to Chandler to understand why the large industrial corporations grew and were able to sell aggressively in competitive world markets.

[5] See Chandler, *The Visible Hand*, p. 87.

[6] See Donaldson, *Corporate Restructuring*, p. 23.

the workforce served the best interest of corporate leadership." [7]

THE RISE OF THE CHANDLERIAN WORKERS: CLERKS, MANAGERS, AND CREATORS

Until about 1980, the continuing rise of the industrial state required an increasing proportion of information processing workers because, until then, office information remained expensive to automate. From 1900 to 1980, as mass production spread over the American economy, the sales and clerical workforce rose from being a relatively minor component of American employment to one of the largest: from 7.5 percent of total employment to 27.3 percent, about 2.5 percentage points a decade (Figure 1). By comparison, agricultural workers (including farmers and farm managers) represented 37.5 percent of the workforce in 1900, while blue collar workers (craft workers, operatives, and laborers) represented 35.8 percent of the workforce in that year. By 1980, agricultural workers had nearly disappeared, falling to 2.9 percent of the workforce, and blue collar workers, while remaining comparatively unchanged, still fell to 31.2 percent of the workforce.

Corporate managers also increased substantially as a proportion of the workforce, rising from 5.8 percent of total employment to 10.3 percent over the same period (Figure 1). Management during this period was self-confident, foresighted, and autonomous to an extent almost unheard of today. Managing a firm was by no means easy: Leaders such as A.P. Sloan of General Motors and Thomas Watson of IBM ran their firms during the uncertainties and difficulties of the Great Depression and World War II. But despite these difficulties, the large industrial corporation assumed a magisterial, almost immortal stature. Managers climbed the corporate bureaucracy secure in the knowledge that whether they as individuals won or lost in the corporate game, the positions at the top would remain to be handed down by the incumbents.

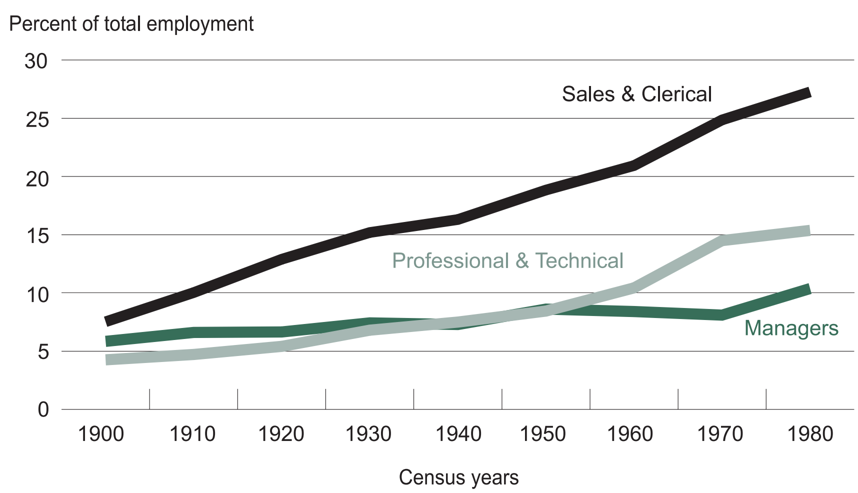
Creativity on a Tight Leash. Of course, corporations had to develop new products, with all the risks that doing so entailed. But their competition tended to come from their peers — other incumbent large corporations — rather than from smaller firms. Competition among incumbents meant that each firm had to be solicitous of its existing stable of products. New product introduction was generally orderly, so as not to excessively cannibalize currently profitable products.

Who created these new products? The occupations most directly concerned with creativity include engineers, scientists (including computer systems analysts and scientists), and writers, artists, entertainers, and athletes. These workers, whom I call creative workers, are most closely associated with research and development (R&D), software, design, the arts, and the media. Between 1900 and 1980, creative workers as a proportion of the workforce base rose faster than either the sales force or management segments, quintupling from 0.7 percent to 3.8 percent.

But creative workers were not the masters of this world; managers were. If new products came into conflict with existing products, the new products gave way, and the existing products remained. During most of this period, creative employees generally found it difficult to strike out on their own because the manufacturing plants, distribution systems, and their corporate bureaucracies were crucial to bringing new products rapidly to market.

FIGURE 1

From 1900 to 1980, sales and clerical workers rose from 7.5% to 27.3% of all workers (2.5 percentage points per decade)



Source: U.S. Census, Employment by Occupation, 1900-1980.

[7] Donaldson, p. 25.

A clear example of the corporation's control over creativity is the annual model changeover instituted by General Motors as a means of maximizing profitability. General Motors orchestrated the rate of change of models in terms of styling and technological progress, introducing technological innovations in its top-of-the-line cars and bringing them, over time, to its entire product line. The phrase "planned obsolescence," which was widely used in the 1950s, described a sales effort to encourage drivers to trade in their cars for the latest model and bespoke the corporation's control over the rate of technological change.

In this ancien régime of the stable corporation, creators were neither as important to corporate profits nor as much of a threat as they would soon be. The great industrial corporations performed most of the private research and development in the U.S. Creativity was leashed, and the managers were in control.

In the 1970s, however, distribution channels began changing more rapidly, as information processing became increasingly automated. Advances in electronics began to outmode many of the roles of information workers. An indication of the changing usefulness of electronics is that from 1977 to 1985, business investment in computers rose abruptly in economic importance. This was associated with the deployment of PCs, minicomputers, and video terminals, which increasingly permitted the processing, transfer, and storage of information with little or no human intervention. Computer purchases rose from 0.3 percent of GDP to 0.8 percent (Figure 2). This was the starting point of the computer's becoming — through its remarkable technological progress — ubiquitous. The proportion of the U.S. economy's resources used to produce computers has remained roughly at this level ever since 1985.

THE RISE OF COMPUTERS AND THE DECLINE OF THE TYPING POOL: 1977 TO 2002

In the wake of this eruption of expenditure on computer hardware, the growth of sales and clerical employment slowed dramatically. Some occupations, such as typists, began to decline absolutely. If the trend in the growth of the number of clerks established from 1900 to 1980 had continued to the end of the century, these workers would have been 32.5 percent of the workforce. Instead, sales and clerical workers were 25 percent of the workforce, less than the 27.3 percent in 1980 (Figure 3).[8]

In short, the electronics revolution of the 1970s made large chunks of the existing corporate bureaucracy obsolete. The investment that corporations had made in these information systems was sharply devalued. The automation of information meant that new entrants could far more easily enter markets, particularly if they had technology that could surpass that of the existing market leader.

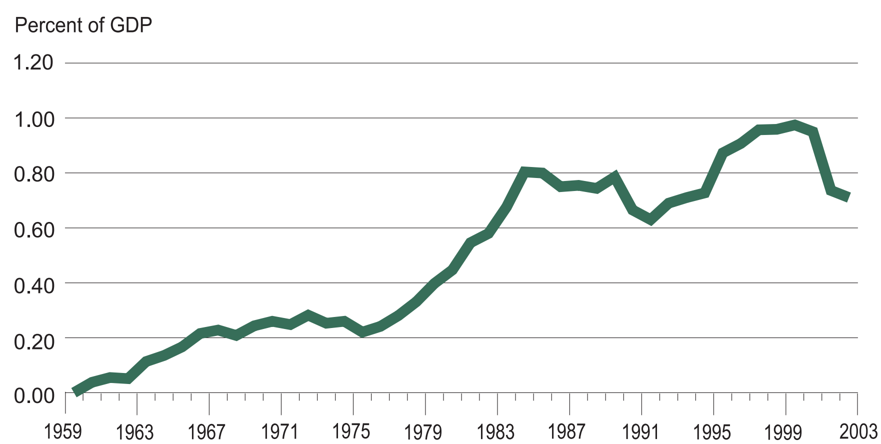
For example, the size of stores and the number of different items on their shelves increased substantially because it became easier for store managers to track sales and inventory, to change prices, and to order new stock.[9] Indeed, tracking sales and ordering

[8] These data are based on the U.S. Bureau of Labor Statistics' 1990 system for categorizing occupations. In the 2000 census, a new system was used, which raised the proportion of sales and clerical workers by about 1 percentage point.

[9] To give one example, variety at the average supermarket accelerated in 1980s and 1990s (see my 1999 article). Items per store in grocery supermarkets surveyed in *Progressive Grocer* magazine rose 2.7 percent annually from 1960 to 1970 and 1.8 percent annually from 1970 to 1980. By contrast, from 1980 to 1990, items per store rose 5.8 percent annually, and from 1990 to 1994, 4.4 percent annually. With the stores providing more space, the rate of innovation accelerated. Annual new product introductions in grocery categories rose from 1,365 in 1970 to 2,689 in 1980 to 13,244 in 1990.

FIGURE 2

Investment in Computers Surged from 1977 to 1985*



*Gross private investment in computers, nominal

Source: U.S. Bureau of Economic Analysis fresh stock could increasingly be done at the national level. That, in turn, implied less need for a manufacturer to send sales workers to individual stores: A sales pitch at the retailer's corporate headquarters could stand in for a hundred visits to store managers. And that might give a new manufacturer access to consumers in numbers near those of an established industrial giant.

Microsoft routed IBM from the market for personal computer software despite the fact that IBM had as many as 400,000 employees in the mid-1980s, while Microsoft had fewer than 2,000 as late as the end of 1987. Microsoft did not suffer the fate of the Tidewater Oil Company: Its lack of an extensive corporate bureaucracy and sales force did not prevent it from selling to millions of consumers. Of course, the Internet has further expanded small firms' ability to rapidly seize markets with new products.

Consequently, many large industrial corporations whose sales and clerical workforce had become outmoded by the electronics revolution have found themselves besieged by sharply increased competition. The workforce that had been a barrier to entry could actually impair the corporation's ability to resist entry by a superior product, since laying off or retiring the now-redundant workers is typically an expensive and disruptive process. Naomi Lamoreaux, Daniel Raff, and Peter Temin have also written about the decline of the Chandlerian corporation in this period, emphasizing the inability of these corporations to change rapidly in response to these new conditions.

As these natural barriers to entry were falling, artificial barriers such as tariffs and government regulation were also being reduced. Globalization has increased foreign competitors' ability to enter our markets and, at the same time, has increased the value of new products by widening the potential market for them. Deregulation has reduced or removed government protection of monopoly power. Telecommunications, trucking, airlines, banking, electric utilities, and pharmaceutical companies have been subject to changes in regulation that have increased competition.

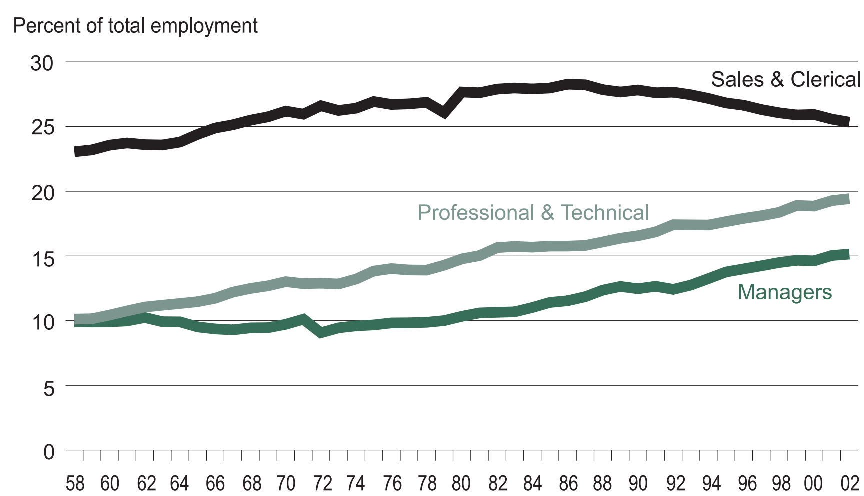
In principle, globalization could have increased the power and stability of large industrial corporations, as the global reach of their marketing expanded. In practice, the great industrial corporations did not fare well. GM, Ford, and Chrysler found their market successfully invaded by Japanese and German competitors. Xerox, IBM, Kodak, Caterpillar, U.S. Steel, AT&T — icons of American industry — all lost their once-solid grips on their core markets.

McKinsey management consultant Tom Peters had been a follower of Chandler in 1982, arguing in his bestseller *In Search of Excellence* (written with Bob Waterman) that the large industrial corporation could be successful as long as its sales force enabled it to pay close attention to its customers. But the very corporations that Peters held up as models of excellence in 1982 stumbled badly as the decade progressed. As the great industrial corporations revealed feet of clay, Peters turned apostate in 1987, proclaiming in the introduction to his new bestseller *Thriving on Chaos* (p.3), "There are no excellent corporations."

Increasing Competition Means Increased Risk. As the large industrial corporation lost its ability to defend its markets, creativity became the new key to profitability. But innovations are very risky. A very few have extraordinarily large returns, while most others have little or no value. F.M. Scherer and Dietmar Harhoff have shown this for a variety of groups of patented and licensed products in Germany and the U.S. One new data set they constructed is based on corporate estimates of the value of patents originating in Germany and the United States. They studied a sample

FIGURE 3

Sales and clerical worker growth slowed in the late 1970s and began to decline in the late 1980s



Source: U.S. Bureau of Economic Analysis of patents filed in Germany in 1977, all of which had sufficient value to warrant paying 16,000 deutsche marks (roughly $13,000 in today's dollars) in annual renewal fees until the patents' expiration at full term in 1995. Scherer and Harhoff asked the corporations that had owned the patents what the total value of the patents had been during their lifetimes. Of the more than 600 patents originating in Germany, they found that the most valuable five patents made up 54 percent of the total value of the entire group of patents. And the top 60 patents made up 80 percent. With innovations so risky, the corporate pursuit of innovation as the main source of profit implies more risk for the corporation.

Further evidence of the riskiness of creativity comes from research on its impact on future earnings arising from R&D expenditures. S.P. Kothari, Ted Laguerre, and Andrew Leone showed that from 1972 to 1997, R&D investments generated future earnings that were three times as uncertain as investments in plant and equipment; investments in advertising were about as risky as investments in plant and equipment.[10]

As competition to produce innovations has accelerated, corporations have faced increased individual risk, as reflected in the stock market volatility of corporate share prices. Financial analysts distinguish the risks that are faced by individual firms separately from general economic or market conditions ("idiosyncratic" risks), such as the failure of individual products or brands, and risks, such as interest-rate changes, that are common to the entire stock market ("aggregate" risk) and to all firms within a given industry ("industry" risk). John Campbell and co-authors documented that idiosyncratic corporate risk rose substantially relative to market and industry risk. During the period they examined (1962 to 1997), overall market risk increased very little. But when the risk of individual shares is broken out into aggregate risk, industry risk, and idiosyncratic firm risk, it is evident that idiosyncratic risk rose substantially after 1980. Thus, it appears that the heightened competition between corporations translated directly into heightened risk for the individual corporation. This heightened risk has a direct consequence: Hedging idiosyncratic risk requires more diversification. Through the 1970s a portfolio of 20 stocks was considered big enough to diversify away most idiosyncratic risk. By the early 1990s, that was no longer true; instead, Campbell and co-authors found that it took 50 stocks to achieve the same benefits of diversification. On the other hand, this greater risk to the individual corporation appears to have gone hand in hand with investors' increasing familiarity with mutual funds that enabled stock market investors to inexpensively diversify their holdings.

Some of this increased risk may well be due to the riskiness of creativity. The authors also suggest that some of this risk may be due to changes in corporate governance. There has been a strong tendency over the past two decades to break up conglomerates and replace them with firms that specialize in a given industry, as described in Mitchell Berlin's article. This tendency, in turn, may be due to improved financial markets that can provide firms with good access to capital despite heightened risk. Shareholders, being more diversified, could afford to ignore the heightened risk of individual corporations.

As competition to produce innovations has accelerated, corporations have faced increased individual risk, as reflected in the stock market volatility of corporate share prices.

CORPORATE AMERICA ADJUSTS TO GREATER RISK

How did this heightened risk affect corporate management? Recent research suggests one consequence: In recent years, boards of directors and block shareholders more frequently either forced out the existing chief executive or replaced the chief executive with a new leader drawn from outside the corporation.

Heightened Risk for Managers. Mark Huson, Robert Parrino, and Laura Starks documented the rise of forced turnover and outsider succession from 1971 to 1994. In their data on leaders of major corporations, the proportion of involuntary turnovers among corporate chief executive officers rose from 10 percent (1971-76) to 23 percent (1989-94). Similarly, successions in which outsiders were appointed chief executive officer rose from 15 percent (1971-76) to 30 percent (1989-94). When an outsider becomes CEO, the corporate board is failing to use this ultimate promotion as a reward for current managers, thus expressing a lack of faith in existing management.

Did this change occur because shareholders have become more demanding of corporations?

[10] They measure uncertainty of future earnings by using the standard deviation of after-tax corporate profits in the five years after the investment.

There is no doubt that the rise of institutional management of pensions and mutual funds and other large pools of investment funds, including those of corporate raiders, has made the typical shareholder more mobile, less concerned with the stability of individual shares, and more concerned with market risk-adjusted rates of return. However, Huson and co-authors showed that this rise does not appear to have been driven by changes in corporate governance or the intensity of the takeover market. Instead, it appears that corporate risk — profit slowdowns and stock declines — was the driving factor.

Thus, it appears that corporate leaders more often were viewed as failing to maximize the corporation's value in the intense competition that developed over the course of the 1980s and 1990s. The intense competition apparently made corporate boards believe that the right CEO was a rare individual and not necessarily one who could be found within the ranks of the corporation itself.

In addition, directors and shareholders may believe that only an outsider can carry out the successful restructuring of a corporation. If, for example, a corporation needs to dispose of core parts of the business, an insider may be too loyal to the past vision of the corporation to take the necessary steps quickly enough for corporate survival.

Structural Change in the Corporation. Another reaction to heightened risk was that top managers became busier. The corporate hierarchy has tended to flatten out, with more managers reporting directly to the chief executive officer. Raghuram Rajan and Julie Wulf documented that corporations became flatter between 1986 and 1999. By 1999, the average chief executive officer had more positions reporting directly: on average, 7.2 positions, up from 4.4 in 1986. Thus, it appears that the CEO now has more day-to-day responsibilities. In keeping with this, fewer layers intervene between the CEO and division heads (the lowest managers with profit center responsibility), and the average firm has shed more than one layer. This circumstance does not reflect larger divisions (divisions shrank in size) or more employees in the corporation (the average remained roughly constant). Rajan and Wulf argue that this represents firms with more human capital than physical capital. One interpretation is that it reflects a switch from the Chandlerian corporation to a more creative and competitive environment. A free and rapid flow of ideas into action has become more important. A world of intense competition in new products is a world in which the pressure to bring a new product to market before one's rivals may require a leaner corporation.

> The reduction in the number of layers in the corporation and the rise in the number of individuals reporting to the chief executive officer imply more work day-to-day for the CEO than formerly.

More Expropriability? Another reason that corporations may have become flatter is that creativity is more expropriable than hardware. In the ancien régime of the Chandlerian corporation, which was centered on physical investment in a production facility, this type of expropriation was difficult to achieve.

Rajan and Luigi Zingales argue that pieces of the corporation may break off and compete against the original corporation: A middle manager can leave the firm and take subordinates along. But age is a counterweight to defections: Senior managers have less to gain by defection because their post-defection working life is short. If creative workers need their managers to cooperate with them in a startup to rival the original firm, flattening may help prevent departures by making the CEO and senior managers too important a part of each team to make departures tempting. While Rajan and Zingales have emphasized the organizational changes that accompany a shift away from physical capital to human capital, George Mailath and Andrew Postlewaite highlight a counteracting force that limits employees' ability to defect or to threaten defection in a firm where a large share of assets are intangible. They argue that it may be difficult for employees to coordinate their defection when the number of employees needed for a successful defection is large.

The reduction in the number of layers in the corporation and the rise in the number of individuals reporting to the chief executive officer imply more work day-to-day for the CEO than formerly. The corporate CEO has less time to focus on very long-term corporate issues. But this may not be such a great loss in a world that, because it is rapidly changing, may be less predictable.

A FRESH WIND FOR CAPITALISM?

I have argued that the rise of computers in the 1970s made life riskier for the large industrial corporation. The result has been more entry by smaller firms, more new product competition, and compelling vitality for the U.S. economy. The cost has been that life at the top within the large corporation has become tougher: riskier, faster, busier. Corporate hierarchies have flattened, and CEOs spend more time with their division heads and perhaps less time contemplating the long view.

Corporate executives are being treated as if their decisions mattered much more to corporate profitability and are being held accountable accordingly. The talent and effort required to successfully run a corporation may well have risen substantially. In such circumstances, it would be surprising if corporate salaries were not rising to compensate for the heightened demands and shortened careers.

All this implies greater conflict in the relationships among shareholders, boards of directors, and top corporate officers. Recent episodes of corporate wrongdoing may be a symptom of uneven progress toward new institutional structures. BR

REFERENCES

Berlin, Mitchell. "Jack of All Trades? Product Diversification in Nonfinancial Firms," Federal Reserve Bank of Philadelphia *Business Review*, May/June 1999.

Bils, Mark, and Peter Klenow. "The Acceleration in Variety Growth," *American Economic Review* 91, May 2001, pp. 274-80.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexaio Xu. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance* 56, 2001, pp. 1-43.

Chandler, Alfred. *The Visible Hand: The Managerial Revolution in American Business.* Cambridge, Mass.: Harvard University Press, 1977.

Chandler, Alfred. *Scale and Scope: The Dynamics of American Capitalism.* Cambridge, Mass.: Harvard University Press, 1990.

Donaldson, Gordon. *Corporate Restructuring: Managing the Change Process from Within.* Boston: Harvard Business School Press, 1994.

Galambos, Louis. "The U.S. Corporate Economy in the Twentieth Century," in Stanley L. Engerman and Robert E. Gallman, (eds.), *The Cambridge Economic History of the United States, Vol. III: The Twentieth Century.* Cambridge, UK: Cambridge University Press, 2000, pp. 927-67.

Huson, Mark R., Robert Parrino, and Laura T. Starks. "Internal Monitoring Mechanisms and CEO Turnover: A Long-Term Perspective," *Journal of Finance* 56, December 2001, pp. 2265-97.

Jackson, Tim. *Inside Intel*. New York: Dutton, 1997.

Kothari, S.P., Ted E. Laguerre, and Andrew J. Leone. "Capitalization Versus Expensing: Evidence on the Uncertainty of Future Earnings from Capital Expenditures Versus R&D Outlays," *Review of Accounting Studies*, 7, 2002, pp. 355-82.

Lamoureaux, Naomi R. *The Great Merger Movement in American Business, 1895-1904.* Cambridge, UK: Cambridge University Press, 1988.

Lamoreaux, Naomi R., Daniel M. G. Raff, and Peter Temin, "Beyond Markets and Hierarchies: Towards a New Synthesis of American Business History," *American Historical Review,* 108 (April 2003), pp. 404-33.

Mailath, George J., and Andrew Postlewaite, "Workers Versus Firms: Bargaining Over a Firm's Value," *Review of Economic Studies*, 57, July 1990, pp. 369-80.

Nakamura, Leonard. "The Measurement of Retail Output and the Retail Revolution," *Canadian Journal of Economics* 32, April 1999, pp. 408-25.

Nakamura, Leonard. "Economics and the New Economy: The Invisible Hand Meets Creative Destruction," Federal Reserve Bank of Philadelphia *Business Review,* July/August 2000.

Peters, Tom. *Thriving on Chaos*. New York: Knopf, 1987.

Peters, Tom. *Liberation Management*. New York: Knopf, 1992.

Rajan, Raghuram G., and Julie Wulf. "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies," mimeo, March 2003.

Rajan, Raghuram G., and Luigi Zingales. "The Firm as a Dedicated Hierarchy: A Theory of the Origins and Growth of Firms," *Quarterly Journal of Economics*, 116, August 2001, pp. 805-951.

Scherer, F.M., and Dietmar Harhoff. "Technology Policy for a World of Skew-Distributed Outcomes," *Research Policy*, 29, 2000, pp. 559-66.

Sloan, Alfred P., Jr. *My Years with General Motors*. New York: Doubleday, 1990.

Managing the Recovery in Uncertain Times:

A Summary of the 2003 Philadelphia Fed Policy Forum

BY LORETTA J. MESTER

Managing the Recovery in Uncertain Times was the topic of our third annual Philadelphia Fed Policy Forum held on November 14, 2003. This event, sponsored by the Bank's Research Department, brought together a group of highly respected academics, policymakers, and market economists, for discussion and debate about the effect of uncertainty on economic decision-making. Our hope is that the 2003 Policy Forum serves as a catalyst for both greater understanding and more research on policymaking in the face of uncertainty.

In its March 2003 press release, the FOMC acknowledged the difficulties that increased uncertainty was creating for assessing the risk to the economic outlook: "In light of the unusually large uncertainties clouding the geopolitical situation in the short run and their apparent effects on economic decision-making, the Committee does not believe it can usefully characterize the current balance of risks with respect to the prospects for its long-run goals of price stability and sustainable economic growth."

Loretta Mester is a senior vice president and director of research at the Federal Reserve Bank of Philadelphia.

Although policymakers and forecasters always have to operate in an uncertain environment, the degree of uncertainty that surrounded the economy during this cycle stands out. In fact, a search of the *Wall Street Journal*'s archives shows that the words "uncertainty" or "uncertainties" appeared in that newspaper 20 percent more times in the three years 2001-2003 than during a comparable period 10 years ago, 1991-1993. This might not be the best measure, but it is an indication that "uncertainty" was on people's minds.

Those uncertainties began with the tragic attacks of September 11. The initial economic uncertainties surrounding that event — for example, whether the banking system would continue to operate and whether and when the markets would reopen — were resolved quickly. But new uncertainties arose in 2002. These included uncertainties surrounding the outcome of the war with Afghanistan, the possibility of continued terrorist threats, and the effects of the corporate accounting and governance scandals that broke in the summer of 2002, all of which created concerns about the staying power of the recovery. Another large uncertainty loomed at the beginning of 2003, namely, whether and when U.S. military action in Iraq would start. The military campaign was followed by uncertainty about the success of rebuilding and peacekeeping activities in Iraq.

On the economic front, questions about the efficacy of monetary policy in a low-inflation or stable-price environment added additional uncertainty. Layered on top of all this was the fact that the economy appeared to be undergoing some structural changes on its own. There was uncertainty about whether the usual economic dynamics continued to be at work or whether these dynamics had shifted. Thus, the Policy Forum began with a discussion of some of the sectors — the consumer, housing, investment, and labor markets — that appeared to have behaved somewhat differently during the recovery that began in November 2001. We then turned to implications of uncertainty for optimal monetary policy and the effect on the economy of the uncertainty raised by the corporate governance and accounting scandals.

An underlying theme that emerged during the presentations was that uncertainty assumes many guises and each can have a different effect on the economy and decision-making.

Anthony M. Santomero, president of the Philadelphia Fed, began the day discussing how the three factors usually cited as causing the very slow turnaround in the labor market during this recovery — weak aggregate demand growth, strong productivity growth, and an increasing use of foreign labor — were all part of the same phenomenon, namely, the unfolding impact of the technological revolution on the economy. The boom and subsequent bust of business spending on information and communications technology (ICT) had generated the most recent business cycle. The very strong tech spending in the late 1990s represented a mix of both good and bad judgments. An implication is that it took the business sector three years, from 2000 through 2002, to digest this major investment spending, reallocating it across firms and fully exploiting its capabilities to boost productivity and cut costs within firms. Only after this were firms ready to contemplate new investment.

The late 1990s' acceleration in ICT investment coincided with a marked pickup in productivity growth, which points to higher growth in potential GDP. To re-employ those who became unemployed or underemployed during the recession and early recovery, real GDP growth would need to be higher than this now higher growth of potential. Moreover, the ICT revolution has raised the potential for possible mismatches in the near term between workers' skills and businesses' requirements, which can slow the rate at which unemployed workers are re-employed, relative to previous recoveries. In the longer term, it also means that industries will need to restructure, similar to what we've experienced in previous technological revolutions. Santomero points out that while these transformations benefit society by leading to higher income growth overall, the transition can be very difficult for workers whose job requirements and locations change. In his view, the flexibility of the U.S. economy implies that markets will induce the required adjustments, and these adjustments will occur in a global context.

Santomero notes that the ICT revolution is creating an increasingly integrated market for both goods and services, including labor services. Information can be disseminated, transactions effected, and far-flung activities coordinated at lower cost than ever before. This means that as a result of the technology revolution, the demand for labor in the U.S. will become more sensitive to the labor market and other economic conditions abroad. Still, while economic forces will play out in a broader global context, the forces are not fundamentally different from those we've experienced in the past. Thus, in Santomero's view, economic stabilization policies, both monetary and fiscal, can still be effective in mitigating the impact of business cycles in this globally integrated economy in which market forces have an increasingly international scope.

The Policy Forum turned next to issues of how this business cycle has played out in various sectors and the role of uncertainty in the economic dynamics.

CONSUMER BEHAVIOR[1]

I had the pleasure of moderating the first session, which addressed the behavior of the consumer sector. Typically in recessions, consumer spending declines, but during the 2001 recession, consumer spending continued to grow. As interest rates fell to historically low levels, there were several waves of mortgage refinancings, which put money into homeowners' pockets. Increased home equity buoyed consumers' wealth, even as the stock market booms of the 1990s ended and the market significantly corrected. How important were these factors to consumer spending during the cycle? How do consumers respond to increases in their wealth? The session's papers underscored the importance of looking beyond the simple conventional wisdom in addressing these questions.

Frank Stafford of the University of Michigan, and director of the Panel Study of Income Dynamics (PSID), began his presentation by discussing the role of housing equity in the economy and what he terms the "refi puzzle." Almost 40 percent of households' nonpension wealth is in housing equity. The cost of accessing home equity can be high. It's not unexpected that when mortgage rates fall, possibly as the result of a change in monetary policy, one of the responses is a boom in refinancing. What's less expected is that there are people who refinance even during periods when interest rates are high. Moreover, there are people who refinance and pay a higher rate. This suggests there may be different motivations for refinancing. Stafford's research, which is co-authored by Erik Hurst of the University of Chicago and which uses micro data from the PSID, suggests that there are two motivations for refinancing. The first is the standard theory of refinancing: Refinancing is a financial option that is exercised when it is "in the money" (i.e., when interest rates fall enough relative to the costs of refinancing to make it financially worthwhile). Stafford's research indicates that people who refinance for this reason are not, in general, spending the proceeds. Instead, the money shows up elsewhere in their portfolio.

But there is a second motive for refinancing: a consumption option. Income is variable but home equity is a source of funds that can be

[1] Many of the presentations reviewed here are available on our web site at www.phil.frb.org/econ/conf/ policyforum2003.html.

unlocked via refinancing. People who face an unexpected need for funds may refinance to gain access to their equity for spending. Thus, they may refinance and remove equity from their home even if they end up paying a higher rate (and Stafford finds that those whose refinancing results in a loan-to-value ratio over 0.8 do pay a premium). Note that refinancing for this reason can occur even in a stable interest rate environment. His empirical results indicate that rational households use home equity to buffer shocks to income — refinancing allows households to optimally smooth their consumption in the face of income shocks — and liquidity-constrained households spend most of the proceeds from refinancing.

The implication is that refinancing provides monetary policy another channel through which to affect the economy. Stafford's empirical results suggest that the liquidity-constrained households that experienced unemployment (an income shock) converted 60 cents out of every dollar they removed from equity via refinancing into current consumption (and this is likely a lower bound); on average, they removed about $16,000. He estimates that there was $18 billion in new spending by these liquidity-constrained households generated by refinancings as the Fed cut interest rates in 1993-94.

But Stafford also points out the downside. His data show that people who paid premium rates to refinance in the late 1990s often subsequently got into financial distress and pulled back spending. Thus, policymakers cannot expect to use the mortgage refinancing channel recurrently over short time periods. Nonetheless, Stafford hypothesizes that the 2001 recession generated a substantial negative stock market wealth effect that was, to a substantial extent, counteracted by the housing refinancing channel.

Sydney Ludvigson of New York University continued the discussion of the marginal propensity to consume out of wealth. As the stock market surged in the late 1990s, household net worth grew 47 percent from 1995 to 2000. A growing concern was the potential negative effect a large stock market correction might have on consumer spending and the aggregate economy. Another concern was that in the absence of a correction, consumers would be reacting to the surge in the stock market with a lag, so that eventually there would be an acceleration in consumption and an increase in inflationary pressures. Ludvigson concludes that both risks were not significant, based on her research with Martin Lettau of New York University. This research addresses four common statements about consumption and wealth and shows that the conventional wisdom is misleading.

Misleading statement (1): "An increase in the stock market raises consumption via the wealth effect." This is misleading because it ignores the distinction between transitory and permanent increases in wealth and treats all wealth changes equally. But only permanent changes in wealth affect consumption. Consumers do not react to unsustained transitory changes. That's not to say that transitory changes can't be long-lived. Ludvigson estimates that a transitory wealth shock continues to affect asset values for a little over four years.

Moreover, Ludgvison's research indicates that nearly all wealth fluctuations are transitory. (The transitory and permanent changes in wealth, consumption, and income are identified using cointegration techniques, exploiting the fact that these three variables follow a common trend over the long term. Deviations from this common trend are transitory changes.) She estimates that over

Tom Sargent of New York University

the postwar period, only 12 percent of the variation in household net worth (wealth) is attributable to permanent changes; the other 88 percent is transitory and mainly driven by volatility of the stock market.

Misleading statement (2): "A good rule of thumb is that a $100 dollar gain in wealth raises spending by about $4." This is true only if the change in wealth is permanent. Most changes in wealth are associated with no change in consumption, since most changes in wealth are transitory.

Conventional estimates put the marginal propensity to consume out of wealth at 4 percent, but Ludvigson's research suggests that this overstates the effect of an increase in wealth on consumption, since it ignores the distinction between permanent and transitory increases in wealth. She estimates that a $100 increase in wealth would typically imply only a 60 cent increase in consumption, since most of the change in wealth is transitory. Since most of the gain in the stock market in the late 1990s boom was a transitory change, it wasn't going to have a large effect on consumption. Hence, worries at the time about the effect of a stock market crash precipitating a rapid retraction of consumer spending and perhaps causing a recession had little foundation.

Consider the bull market from 1995 to 2000. Households reacted to the increase in their wealth but took account of the possibility of a stock market correction. The increase in consumption was $1.70 per $100 increase in stock market wealth, not the $4 it would have been if consumers thought the increased wealth were permanent. Similarly, the stock market retreat from 2000 to 2002 elicited a reduction in consumer spending, but only by about 10 cents per $100 of wealth lost. According to Ludvigson, historically, the economy has not had a hard landing stemming from the effects of a stock market correction on consumption.

Misleading statement (3): "Consumers ignore daily volatility but spend if the gains in wealth stay for a few quarters." No, consumers ignore transitory changes, even if they last for several years.

Misleading statement (4): "It takes many quarters for permanent changes in wealth to affect spending." This is not true for most of consumer spending, which responds to permanent changes in wealth within about one quarter. Spending on durable goods takes about two quarters to respond, but durables make up only 10 percent to 15 percent of consumer spending. This means that in the late 1990s, it was unlikely that there was some pent-up consumption due to the stock market boom that was waiting in the wings to put upward pressure on inflation.

The implication of Ludvigson's findings is that policymakers need to carefully evaluate the type of wealth increase in order to forecast its effect on consumption. Failure to distinguish between permanent and transitory increases in wealth could lead to overstating the sensitivity of consumption to changes in wealth and, therefore, policy mistakes.

It is interesting to consider the relationship between the Stafford and Ludvigson results. In the audience discussion following the formal presentations, Ludvigson mentioned additional results that show that in contrast to aggregate wealth, the non-stock-market component of wealth, which includes housing, has a very small transitory component; most of the changes are permanent. Thus, households would be more responsive to changes in wealth derived from housing than that derived from the stock market. Stafford's results suggest that different households respond differently to changes in wealth; so it would be interesting to extend Ludvigson's aggregate results using Stafford's micro data allowing for heterogeneity across households. Perhaps different households respond differently to the permanent and transitory components of wealth.

> Failure to distinguish between permanent and transitory increases in wealth could lead to overstating the sensitivity of consumption to changes in wealth and, therefore, policy mistakes.

A further question is, How do households infer whether the changes in wealth are permanent or transitory? Note that the different marginal propensities to consume out of permanent and transitory changes in wealth may reflect this inference problem. It could be that it takes time for consumers to identify whether a change is permanent, so that the impact of changes in wealth on consumption is only gradual. And the more volatile the component of wealth, the more gradual its impact, since the inference problem is more difficult. This might explain the higher marginal propensity to consume out of changes in non-stock market wealth, which is less volatile, and changes in stock market wealth, which is more volatile.

Another question arises: If consumers are really able to identify the transitory and permanent components, why aren't they selling when stock market prices are over trend and buying when stock market prices are below trend, thereby eliminating the transitory fluctuations? One unsatisfying answer is investor irrationality. But models of rational investor behavior that allow for time-varying risk aversion on the part of investors (in particular, where risk aversion varies over the business cycle and is higher in bad economic times and lower in good economic times) can also help explain persistent transitory variation in the prices of risky assets. Households might be willing to buy risky assets at temporarily higher prices in good times, since they've become less risk averse, even if they expect lower prices in the future. And they might be unwilling to buy assets with temporarily low prices in bad economic times, since they've become more risk averse.

THE EFFECT OF UNCERTAINTY

The Policy Forum's next session looked at the behavior of two other sectors — labor markets and business investment — in an uncertain environment and concluded with a discussion of how to forecast when there is an increased level of uncertainty.

The labor market has figured prominently in discussions of the recent recession and recovery. Indeed, employment growth had been the "missing link" of this recovery until recently. **Richard Rogerson** of Arizona State University discussed the possible effect increased uncertainty might be having on the labor market, emphasizing that different types of uncertainty may have different effects. He began by pointing out that from a policy perspective, it is important to determine what the underlying causes of observed changes in the labor market are, since they may require very different policy prescriptions. But determining these causes is difficult because policymakers must analyze data in real time. Often, we economists first look at the data and then try to formulate a model that helps us understand the data. The model generally describes the steady state, a very stable situation, and then adds some shocks that induce fluctuations around that steady state. Then, when studying some particular economic event ex post, the economist will try to assess whether it is best thought of as a fluctuation around the steady state or a change in the steady state. This helps direct the search for the driving forces of the event. If one looks at historical episodes, one can tell whether, say, employment growth temporarily declined and then returned to its previous steady state or whether it has stayed down, which would be interpreted as a change in the steady state. But Rogerson points out that economists and policymakers have to look at the data in real time, so it won't be clear at any point in time whether growth will revert to its previous rate or whether it will stay down. In a situation like this, Rogerson says, additional information must be brought to bear. For example, one could look at other indicators of the economy. A rebound in other variables but a continued decline in employment might indicate that there has been a permanent decline in employment, i.e., a change in the steady state.

From a policy perspective, it is important to determine what the underlying causes of observed changes in the labor market are, since they may require very different policy prescriptions.

The current recovery is an interesting case. The economy continued to lose a significant number of jobs after the recession trough in November 2001. Indeed, the current data suggest that nonfarm payrolls fell by 1.4 million jobs from November 2001 through August 2003, before companies began rehiring in September 2003. In Rogerson's view, that situation should lead one to consider whether some more fundamental, longer run changes are taking place in the economy. One possibility was that increased uncertainty was depressing the steady-state level of employment by discouraging job creation.

But as Rogerson points out, in thinking about the effect of uncertainty on job creation, it is important to be more precise about what is meant by uncertainty. Hiring is a costly and risky endeavor. Recruiting and training workers takes time and money. The payoff from hiring is uncertain, not only because the quality of the worker is uncertain but also because future demand for the firm's product is uncertain. Although this might seem to imply that increased uncertainty would make firms less willing to hire, Rogerson points out that this need not be the case — it depends on what's meant by increased uncertainty. If increased uncertainty means that the firm's average return is the same but the distribution of possible returns is more dispersed (i.e., there's a mean-preserving spread of returns), this would not be bad for job creation. That's because for a firm, returns are truncated on the downside — if the return goes below a certain point, the firm will close — but they are not truncated on the upside — the firm gets to keep all of the upside gains. This type of increased uncertainty would actually be good for job creation.

But in Rogerson's view, this is probably not the type of increased uncertainty that affected the economy during the beginning of the recovery. Instead, firms thought there was a greater chance that something bad would happen that would cause them to lose their investment. This is not a spreading out of possible returns, but a greater weight on bad outcomes. Rogerson investigated the implications of this type of uncertainty for job creation. He modeled increased uncertainty as corresponding to a shorter expected lifetime for the job once it is created and assumed that job creation entails certain costs, including equipment, recruiting, and training costs. His research implies that a permanent 10 percent increase in uncertainty, i.e., a 10 percent shortening of the expected life of the job, would lead to a 0.5 percent decrease in employment in the steady state — not a trivial effect, but not a large one either.

In Rogerson's view an increase in uncertainty is unlikely to be the main reason for labor market weakness during this recovery: It would take massive increases in uncertainty of the type modeled to result in the employment declines seen early in the recovery. However, he does acknowledge that if the increase in uncertainty were perceived to be temporary, it might have larger effects, as firms wouldn't act until uncertainty was resolved.

In thinking about alternative explanations for labor market weakness, Rogerson notes that although along some dimensions all business cycles are the same, along other dimensions, there is some heterogeneity across those cycles. Current conditions need not be suggesting there has been a fundamental change in the labor market or business cycle. Comparing the current recovery to past recoveries, he shows there was considerable variation in how long it took the labor market to recover. The early part of the 1991-92 recovery looked similar to the current cycle, but the market eventually recovered strongly, i.e., it did not remain at a permanently lower steady state. Rogerson points out that in the 1970-71 recession/recovery, it took almost a year from the trough for labor markets to begin showing some recovery. Thus, other recoveries in the past have had somewhat of a "jobless recovery" aspect to them.

He concludes that it is too soon to say that the labor market weakness that occurred in 1991-92 and in the current recovery represents a change in the nature of businesess cycles. But if the business cycle has changed this time around, Rogerson suggests it could be changes in production methods, workforce options, or composition of economic activity, or it might be a reflection of the nature of the shocks that caused the cycle. He believes that to understand recoveries, one must also investigate the recessions that preceded them. That is, to come to any understanding about the changing nature of the business cycle, one must look at the whole cycle.

Andrew Abel of the Wharton School, University of Pennsylvania, turned our attention to business investment under uncertainty. He began discussing two theories of investment: the standard Tobin's q theory and the newer real options theory of investment. Although many economists believe that the theories reach different conclusions regarding investment under uncertainty, Abel shows that when these theories are correctly applied, they yield similar answers.

According to q theory, the rate of firm investment, i.e., the rate at which firms want to increase the capital stock, should be related to q, the market value of new additional capital relative to its replacement cost. Given that changing one's capital stock involves adjustment costs, a firm will undertake investment until the marginal cost of investment is equal to the marginal value of another unit of capital, q. This implies that the firm will invest up to the point where the net present value of the next unit of capital is equal to zero. Since marginal q is not observable, average q, measured as the value of the firm divided by the replacement cost of its capital stock, is used.

According to the real options theory of investment, the decision to invest now is an irreversible decision; the firm has given up the option of waiting for more pertinent information on which to base its investment decision, and the elimination of this option is a cost of investing today.

What are the implications of the theories for investment under uncertainty? Similar to the point made by Rogerson, it depends on the type of uncertainty and the type of investment. Consider the real options theory: If investment is irreversible (i.e., the capital has no value to others, so it cannot be sold), an increase in uncertainty should tend to discourage current investment, since it raises the value of this option, i.e., the cost of investing today. But if investment is reversible, this need not be the case, since the firm can undo the investment decision.

The effect of increased uncertainty on investment may depend on a number of other aspects of the economy, e.g., how competitive the economy is and whether firms use increasing or decreasing returns to scale technologies. For example, the literature suggests that under increasing returns to scale and perfect competition, an increase in uncertainty, holding all else equal, would raise a firm's q and therefore induce higher investment (for reasons similar to those Rogerson discussed regarding a firm's hiring decision when the increased uncertainty is a mean-preserving-spread of future returns). But under decreasing returns to scale or when firms have monopoly power and investment is costly to reverse, increased demand uncertainty can mean lower investment. Fortunately, q will give the correct signal: If higher uncertainty results in lower q, investment will decline, and vice versa.

Abel concluded his talk with an application of the q theory to the current business cycle, which was driven by the investment cycle: q rose significantly during the boom of the 1990s and then tumbled, and investment followed it with a slight lag. Abel was optimistic about future investment given the recent turnaround in q.

According to **Laurence Meyer** of the Center for Strategic and International Studies, and of Macroeconomic Advisers, uncertainty and forecasting go hand in hand. These uncertainties include forecasting errors; which model is correct and the parameters of that model; unknowable shocks that can hit the economy; identification of structural change (discussed by Rogerson); and data revisions or lack of data on important fundamentals such as potential growth. What's a forecaster to do with so much uncertainty? Meyer suggests that the first rule is "be humble." Next, identify what you know more about and what you know less about. Finally, continually learn, since the economy is a dynamic place and economic research is advancing as well. One of the important things a forecaster can do is know when an old story is over and when a new one is beginning.

Meyer outlined three major questions facing forecasters in the late 1990s, when he served as a Federal Reserve Board Governor: Was there a productivity acceleration and what were its implications? Was there a bubble in the stock market (irrational exuberance)? And would the U.S. experience a spillover effect from global financial turmoil?

It wasn't obvious using the real-time data that we were in the midst of a productivity acceleration, but even after that determination, knowing its implications wasn't trivial. In Meyer's view one lesson we learned from the second half of the 1990s was that the productivity acceleration wasn't just a major supply shock; it was a powerful demand shock as well, driving an investment and consumption boom. It led to above-trend output growth and rapid employment growth. We also learned that the productivity acceleration was a powerful disinflationary event and that it could significantly increase the equilibrium real interest rate, which has implications for monetary policy.

In Meyer's view one lesson we learned from the second half of the 1990s was that the productivity acceleration wasn't just a major supply shock; it was a powerful demand shock as well.

These developments were not consistent with the simplest models in which productivity growth affects wages and prices symmetrically. In these models, there should be no relationship between NAIRU (the non-accelerating-inflation-rate of unemployment) and productivity growth. In Meyer's view the late 1990s showed that when there's an unanticipated acceleration in productivity, there is a significant effect on short-run NAIRU and on inflation dynamics. Thus, it was time to change the model — in particular, productivity growth may have asymmetric effects on wages and prices, with prices responding more immediately to changes in productivity growth than do wages. In this case, prices fall in response to an unexpected rise in productivity, and NAIRU falls in the short run. Meyer believes that this is a big part of the reason the decline in unemployment didn't lead to higher inflation in the second half of the 1990s. (Later in the day, Al Broaddus, the recently retired president of the Richmond Fed, also spoke about this acceleration in productivity growth.)

Another big question faced Meyer and his fellow policymakers and forecasters: Was the run-up in the stock market based on "irrational exuberance" and therefore unsustainable? If so, a stock market correction could be expected in the near future. In Meyer's view economic performance over the past few years can be explained as a post-bubble hangover, dominated by imbalances inherited from the bubble period, including the capital overhang and over-leveraged corporate balance sheets. These imbalances made it difficult to forecast during this period. To the extent that these imbalances are corrected, it would be a mistake to weight them too heavily in forecasting the future. However, he points out that the future always has some roots in the past — there are legacies that affect the economy going forward.

Today's economy is also affected by several uncertainties. One of the largest in Meyer's view is whether there's been another acceleration in structural productivity growth. Understanding the source of the acceleration would be important in assessing its impact on the economy, a theme echoed earlier in Rogerson's presentation. Higher productivity growth has been associated with weaker labor markets this time, in contrast to the positive effect it had in the late 1990s. It could be that increased productivity growth derived from different sources may have a different impact on the economy. The acceleration in productivity growth in the late 1990s was driven by capital deepening, which went hand in hand with the investment boom and was beneficial for the labor market. The post-recession acceleration in productivity growth has been driven by competitive pressures and cost cutting, which has been a negative factor for labor markets.

Meyer concluded by addressing how monetary policymakers should respond in times of heightened uncertainty. One school of thought is that when there's uncertainty surrounding important fundamentals, the policymaker should attenuate his response. But another view is that the policymaker should be continually updating his model and reacting as aggressively as he normally would, based on the updated model. In the late 1990s, these two schools would have produced observationally equivalent outcomes: (1) if you are uncertain about the natural rate of unemployment, as the unemployment rate falls, you might not want to react to it, versus (2) as the unemployment rate falls without any acceleration in inflation, you revise down your estimate of the natural rate and therefore don't need to respond. Meyer says it is difficult to tell which of these was the dominant way of looking at policymaking in the second half of the 1990s.

Former Federal Reserve Governor Larry Meyer

OPTIMAL MONETARY POLICY

Important advances in macroeconomic theory and macroeconomic modeling in recent years have given economists the necessary tools to address questions about optimal monetary policy. **Robert King** of Boston University gave a progress report on the development of macroeconomic models with strong theoretical underpinnings that are being used to study optimal monetary policy. These models are now sometimes referred to as the New Neoclassical Synthesis (NNS). The new macro models incorporate rational expectations; neoclassical foundations so that one can study consumption and investment dynamics; and various mechanisms so that money can affect the real economy in the short run; they also can be used to evaluate alternative monetary policy rules. These models are fully articulated, in the sense used by Tom Sargent (our next speaker), in that the objectives of firms and households, the structure of the markets in which they interact, the nature of the shocks that hit the economy, and the policy instruments and information available to monetary policymakers are fully specified. Thus, the models have strong micro foundations.

The models have yielded a strong and consistent prescription for monetary policy: The central bank should target a low and stable rate of inflation. Current research in optimal monetary policy design is looking at a variety of motivations for departing from this strict inflation target, including price shocks, aggregate demand shocks, and financial market disruptions. But the general conclusion is that optimally there would be little variation in the price level.

King reviewed the important ingredients in these models. Rational expectations modeling, developed by Tom Sargent and others, inevitably led to general equilibrium analysis and the development of important computational and econometric tools. Economic agents often need to make forecasts about future events in order to make rational decisions, e.g., firms must forecast revenues and costs in order to decide whether an investment is worth undertaking. Under rational expectations, firms and other economic agents adjust their expectations over time to minimize forecast errors — i.e., they are not consistently fooled. Thus, outcomes won't differ consistently from expectations over time. To form its expectations, a firm must think about the product and labor markets in which it interacts, and this leads naturally to the general equilibrium models known as real business cycle (RBC) models. RBC models emphasize the importance of shocks to the real side of the economy and, in particular, the role of technical progress as a source of fluctuations. These models provide a powerful methodology for studying the interactions between the various agents in a macro economy.

King discussed the next major development, the so-called New Keynesian macro model. This model focuses on firms' price-setting behavior and produces a role for monetary policy in economic stabilization. A key ingredient is the notion of price stickiness, i.e., firms have some market power and set prices, but they hold those nominal prices fixed for some period because adjusting prices is costly. This, in turn, gives the central bank an avenue for affecting real activity by affecting real markups and relative prices. Note that a firm's pricing decisions in the New Keynesian model depend on expected future inflation, and these pricing decisions determine current inflation. Thus, current inflation depends on expected inflation.

The New Neoclassical Synthesis (NNS) builds on the New Keynesian model by embedding the New Keynesian price stickiness and imperfect competition in a fully articulated stochastic dynamic general equilibrium model with strong micro foundations. A major benefit of these models is that they can be used to systematically study the effects of alternative monetary policy rules on real economic activity. The NNS models indicate that monetary policy shocks can have large and persistent effects on real economic activity and that the choice of rule matters. The models also underscore the importance of credibility in the monetary policymaker, since current inflation depends on expected future inflation. In King's view the NNS underscores that the management of expectations may be a key part of the central bank's job and that imperfect credibility in the monetary policymaker may be very important in understanding particular historical economic episodes.

King concluded his presentation by discussing the implications of the NNS models for optimal monetary policy. Within the context of the NNS model, a policy of strict inflation targeting at a zero inflation rate (i.e., a price level target) is optimal because it eliminates the relative price distortions caused by the interaction of inflation and sticky prices, since there is no inflation; it stabilizes the average markup and thereby holds fixed the market power distortion, which the central banks cannot eliminate; and it generates the same level of real output that would occur if prices were fully flexible and not sticky. Note that this "natural rate of output" fluctuates with the real shocks that hit the economy. King indicates that estimates of the cost of inflation in these models are dependent on the details about price stickiness, which is assumed to be exogenous; he suggests this is one place where further research is necessary.

More recent work with NNS models suggests that optimal monetary policy may deviate from strict adherence to a price level target but that the models constructed to date indicate that optimal policy would allow only a little change in the price level. King conjectures that simple targeting rules may be close to optimal for a wide class of models and that this remains an important subject of ongoing and future research.

Thomas Sargent of New York University continued the discussion of optimal policy under uncertainty. The Federal Open Market Committee (FOMC) comprises 19 members. As Sargent points out, it could be that they have different goals and objectives. They might have different information on the economy — after all, the committee was designed to have regional representation. Or they might have different models of the economy — i.e., they may put different probability distributions on the various possible sequences of future economic outcomes. Sargent focused his discussion on this latter case.

As Sargent explained, rational expectations, which was discussed earlier by King, has been a powerful force in macroeconomics, and its power derives from the fact that it eliminates the possibility of different agents' having different views of the world. Rational expectations doesn't allow expectations to be free parameters; it makes them outcomes, and it delivers cross-equation restrictions that are important for deriving optimal policy. But as Sargent points out, rational expectations also eliminates any discussion of model mis-specification or multiple models. Under standard rational expectations, all agents have the same model. They can have different information, but they have the same economic model. In Sargent's view, learning may be technically difficult to analyze, but in a world with a single model, it is philosophically trivial — it's just the application of Bayes' Law to update beliefs.

But experimental evidence suggests that agents do not act according to Bayes' Law applied to a single model. Rational people behave as if they have multiple models, i.e., multiple probability distributions over various outcomes, in their heads. This is a very profound kind of model uncertainty, and it is not at all clear how learning should take place in such a world. (For example, should you apply Bayes' Law model by model?) In Sargent's view this type of model uncertainty is an important factor in monetary policymaking.

With Tim Cogley, Sargent is currently researching models of learning with an application of why it took the Fed so long to stabilize the great inflation of the 1970s. One story for Fed decision-making in this time period is that economic research in 1960 indicated there was an exploitable Phillips curve (i.e., a systematic tradeoff between inflation and unemployment), and in the mid-1960s, the Fed started trying to exploit it. Inflation started rising. Economic research then suggested there is no exploitable Phillips curve and that an optimal policy would target low inflation. The data in the early 1970s provided confirmation, and subsequently, the Fed returned to targeting low inflation.

This story is a learning story, with the Fed acting as a Bayesian decision-maker. But Sargent's research suggests that it is difficult to get this story to fit the facts. If the Fed was using the incoming data to update its beliefs about which was the correct economic model, by the early to mid-1970s, it would have put almost all the weight on the non-exploitable Phillips curve model as being the correct model. A Bayesian decision-maker, then, would have begun targeting inflation in the early to mid-1970s. But the Fed delayed. Sargent's research suggests that the Fed may have been "model averaging" instead. It's true that in the early to mid-1970s, the evidence suggested that the non-exploitable Phillips curve model was almost certainly correct, but there was still a very small chance that the exploitable Phillips curve model was right. If it was, targeting inflation would have yielded extremely bad economic outcomes. Given that, inflation targeting was too risky. The Fed behaved as a min-maxer — it chose the policy that yielded the best outcome under the worst-case scenario of assuming the wrong model was correct. It was only when the exploitable Phillips curve model was proven incorrect that the Fed began to target inflation.

Sargent's research on policymakers' learning under model uncertainty is relevant for today's FOMC, and it is likely to remain relevant for the foreseeable future because, despite the advances outlined by King, model uncertainty will not be going away anytime soon.

> **Broaddus advises policymakers that in addition to carefully monitoring incoming data, they must also use modern economic analytical tools to be successful.**

Our next speaker, **Al Broaddus**, president of the Federal Reserve Bank of Richmond (since retired), related firsthand knowledge of how an FOMC member learns and uses economic research in formulating policy decisions. Broaddus indicates that the economists in the Research Department at the Richmond Fed, as well as academic visitors in the department, keep him abreast of ongoing research in monetary economics. Some people believe that the FOMC sets its target rate for the federal funds rate mainly by looking at data on current economic conditions and assessing what it indicates about the current state of the economy and the near-term outlook, then determining how promptly and strongly to respond to this information, based on instinct and common sense. Broaddus thinks this characterization is somewhat misleading, since it suggests a lesser role for economic analysis, including relatively recent developments in the academic literature. In his view, economic analysis has frequently played a central role in determining policy, especially over the longer run. Indeed, Broaddus thinks that one of the Fed's greatest achievements over the last three decades was its role in breaking the high inflation of the late 1970s and early 1980s and subsequently helping to bring the rate down to its current low level. Broaddus indicates that economic research showing there was no exploitable systemic tradeoff between inflation and unemployment, discussed earlier by Sargent, paved the way for this accomplishment.

Broaddus next discussed three examples of how economic analysis guided his own thinking as a policymaker and how he has used economic principles in arguing his positions at FOMC meetings. The first involved the inflation targeting debate at the January 31-February 1, 1995, FOMC meeting.[2] Broaddus spoke in favor of inflation targeting, a position he continues to hold today. The underlying economic principle that in-

[2] Transcripts of the meetings Broaddus discussed are available on the Board of Governors' web site at www.federalreserve.gov/fomc/transcripts.

Fed Presidents Al Broaddus and Tony Santomero

formed his view was rooted in the idea of rational expectations, namely, that by announcing an explicit long-run inflation objective, the FOMC would enhance the credibility of its commitment to low inflation. In a rational expectations world, this increased credibility would make it less likely that inflation would reaccelerate but, if it did, would make it less costly to bring inflation back down.

Broaddus points out another benefit of inflation targeting: it would allow the FOMC to act more aggressively to help stabilize the economy in the short run, since its actions to do so would be less likely to reduce its credibility, thereby setting off an inflation scare. Broaddus continues to believe that an inflation target would be beneficial. Indeed, the recent experience with disinflation and proximity to the zero bound on the fed funds rate underscores the need to avoid not only inflation that is too high but also inflation that is too low. Broaddus stated that to him a 1 percent to 2 percent inflation target range for the core PCE would be acceptable.

Broaddus's second example involves the Fed's intervention in foreign exchange markets on behalf of the Treasury with the aim of affecting the value of the U.S. dollar. An extended discussion occurred during the November 15, 1994, FOMC meeting. Broaddus is opposed to such intervention based on the underlying economic principle that intervention cannot have a sustained effect on the value of the dollar unless it is supported by basic monetary policy. He points out the problem that would arise if the policy needed to support the dollar conflicted with the appropriate policy based on domestic economic conditions. At the very least, it might raise doubts about whether Fed policy will support domestic or external objectives. In Broaddus's view, the Fed's intervening on behalf of the Treasury might put the Fed's credibility as an independent monetary policymaker at risk unless the Fed "sterilized" this intervention, i.e., neutralized its effect on the fed funds rate by carrying out offsetting open market operations. But if the Fed did so, the interventions would be unlikely to have a sustained impact on the value of the dollar.

Broaddus's third example involves the recognition that an increase in trend productivity growth has important implications for monetary policy. In 1996 and 1997, the FOMC began to recognize that the U.S. might be experiencing a sustained increase in trend productivity growth. Faster trend productivity growth would imply slower growth in the cost of labor per unit of output for a while, since it would take time for real wages to catch up. As firms passed the lower cost through to lower prices of their final goods and services, this would put downward pressure on inflation. Most reasoned that as long as rising productivity growth kept inflation low, the FOMC could refrain from raising its funds rate target.

At the May 20, 1997, FOMC meeting, Broaddus discussed another possible implication of higher trend productivity growth for monetary policy, namely, that the equilibrium real interest rate might be higher as a result of higher trend productivity growth. Broaddus explained his economic reasoning. Higher trend productivity growth should cause firms to expect higher future earnings and workers to expect higher future wages in a world where the Fed has credibility for keeping inflation stable and, therefore, expected inflation stable. If so, at the prevailing level of real interest rates, firms and workers would want to bring some of this expected future income forward and would do so by borrowing against it. To prevent excessive current demand, current real interest rates would need to rise. Broaddus reports that his argument did not elicit a response during the FOMC meeting. But in his view, somewhat more preemptive tightening might have prevented some of the excess investment during the late 1990s boom, which was followed by an investment decline and recession.

Broaddus advises policymakers that in addition to carefully monitoring incoming data, they must also use modern economic analytical tools to be successful. He believes his colleagues on the FOMC understand this and that economic analysis has improved policymaking over the past 20 years.

CORPORATE GOVERNANCE

Our last session focused on corporate governance, the system by which a corporation is managed and controlled. The recent corporate accounting and governance scandals have brought governance issues to the forefront. Although the effect these scandals have had on real economic activity is difficult to measure, the scandals are often listed as one of the factors that put a damper on the early stages of the recovery. Indeed, the FOMC's August 13, 2002, press release read: "The softening in the growth of aggregate demand that emerged this spring has been prolonged in large measure by weakness in financial markets and heightened uncertainty related to problems in corporate reporting and governance."

Andrew Metrick of the Wharton School, University of Pennsylvania, discussed some of his research on the design of corporate governance structures. How much power should shareholders, the owners of firms, yield to managers? Yielding too much power creates the potential for agency problems, since the managers' and shareholders' objectives may differ. Yielding too little means giving up the benefits of the managers' expertise and resultant superior decision-making. In setting up the governance structure, shareholders want to be able to get rid of managers who aren't doing their jobs, but they also want to give managers the power to make the decisions necessary to run the firm.

Before the wave of hostile takeovers in the 1980s, large firms were effectively immune from takeover. But as a result of the merger wave, in the mid-1980s, firms started adopting takeover defenses, such as poison pills and greenmail, and other provisions, to prevent takeover and reduce shareholder rights. Around the same time, many states passed laws to prevent outside firms from taking over firms in their states. The takeover wave subsided, but most of these provisions remained with little change in the 1990s.

Using the variation in those provisions across firms, Metrick and his coauthors, Paul Gompers and Joy Ishii, developed a "governance index" to proxy for the level of shareholder rights at a large sample of firms. The governance index is constructed using 24 different provisions a firm might have in place that either decrease or increase shareholders' rights (e.g., poison pills, golden parachutes, severance contracts not contingent on a change in control of the firm, whether a supermajority of shareholders is needed to approve a merger, etc.) A higher level of the governance index means a higher level of managerial power relative to shareholder power. The researchers studied how well firms with different levels of shareholder rights performed in the 1990s and found that firms with stronger shareholder rights according to their index earned significantly higher returns than firms with weak shareholder rights. Stronger shareholder rights are also associated with higher profits, higher sales growth, lower capital expenditures, and fewer acquisitions made. Other researchers have found that firms with stronger shareholder rights generally have lower CEO pay and stronger pay-for-performance and that firms with weaker shareholder rights tend to overinvest in booms and then have to cut more as the economy weakens.

Metrick points out that the governance index is a simple construct and not a perfect measure of corporate governance and that the research cannot address causality (i.e., does good corporate governance lead to good performance? Or does good performance beget good corporate governance?). Nonetheless, in his view the results suggest that governance does matter for firm performance and decision-making and may have large macroeconomic implications. Metrick recommends empowering shareholders by dismantling takeover defenses, making it easier for shareholders to elect directors, and clearing the path for shareholder proposals that would be binding on a firm's management.

Peter Hooper of Deutsche Bank Securities, Inc., continued the discussion by pointing out that up to this point, legislation and regulation have driven improvements in corporate governance, but in his view, investor preference will increasingly drive future improvements. As research by Metrick, Deutsche Bank, and others show, investors have good reason to take corporate governance seriously. However, governance, which has both structural elements (such as the composition and independence of the board of directors) and behavioral elements (such as the effectiveness and capability of the directors), has been slow to gain the attention of analysts and investors. One reason is that U.S. firms are generally perceived to be well run and well regulated and that the scandals involve a few bad apples. Another reason is that it has been hard to get good data on governance. This is beginning to change. Information on governance is becoming more widely available, and a number of research firms have begun to rate firms on their corporate governance. In Hooper's view this should result in institutional investors' taking corporate governance issues more into account.

Deutsche Bank Securities has produced its own rating system for firms' corporate governance, combining quantitative and qualitative factors in four areas: board structure, independence, and performance; shareholder treatment (e.g., the presence of anti-takeover devices); information disclosure; and corporate compensation. The resulting corporate governance scores vary widely across the firms in the S&P500. In Hooper's judgment, the dispersion means that investors cannot take for granted that being a U.S. firm or being in the S&P500 means that corporate governance standards are completely sound; the average firm has room to improve its corporate governance performance on this metric. In fact, on average, governance scores have been improving: From June 2001 to June 2003, 71 percent of the companies in the S&P500 showed improved scores. That's not surprising given the increased regulatory and media interest in governance over this period. More surprising to Hooper is that scores at 27 firms fell significantly over this period. The key factors leading to the deterioration were the adoption of poison pills and/or equity incentive plans that would lead to dilution in shareholder voting power.

There appears to be a weak positive correlation between corporate governance score and firm size as measured by market capitalization, but as Hooper points out, the causation could go either way. Larger firms could be instituting better governance structures, perhaps because of greater investor scrutiny or because more resources can be devoted to governance or better governance structures could lead to larger size over time. The governance score does not seem to be systematically related to which U.S. state a firm is incorporated in, but firms incorporated offshore in Panama, the Cayman Islands, or Bermuda have noticeably lower scores. The Deutsche Bank research indicates that good corporate governance is associated with a higher return on equity in 10 out of the 16 major industry sectors of the S&P500, consistent with Metrick's results. While corporate governance does not explain much of the overall volatility in stock prices across S&P500 firms, better governance does appear to be associated with a somewhat lower variance in a firm's stock price.

Hooper's conclusion is that investors and analysts have lagged in appreciating the importance of corporate governance to firm valuation, partly because of the lack of information and data on governance. The scandals of the last two years have raised awareness of the issue, and data are becoming increasingly available with which to rate firms on their corporate governance performance. Hooper believes, as a result, investors will increasingly be taking corporate governance into account when making their investment decisions.

SUMMARY

The 2003 Policy Forum generated lively discussion among the program speakers and audience participants on a number of issues that policymakers must confront in setting policy in uncertain times, which may be an apt description of our economy in the recent business cycle. Our hope is that the ideas raised will spur further research and foster a greater understanding of today's economy.

We will hold our fourth annual Philadelphia Fed Policy Forum, "Challenges and Opportunities in the Global Economy: Perspectives on Outsourcing, Exchange Rates, and Free Trade," on Friday, December 3, 2004. You will find the agenda on page 43.

Tony Santomero, Mike Dotsey, and Charles Plosser of the University of Rochester