THE ANNUAL REPORT
OF THE
COUNCIL OF ECONOMIC ADVISERS

LETTER OF TRANSMITTAL

MR. PRESIDENT:

The Council of Economic Advisers herewith submits its 2007 Annual Report in accordance with the provisions of the Employment Act of 1946 as amended by the Full Employment and Balanced Growth Act of 1978.
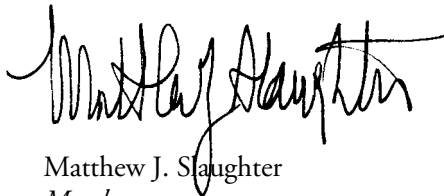
Sincerely,

Edward P. Lazear
Chairman

Katherine Baicker
Member

Matthew J. Slaughter
Member

CONTENTS

Overview

The U.S. economy continues to exhibit robust growth, with a strong labor market and moderate inflation (see Chapter 1). These accomplishments are supported by rapid productivity growth that makes our economy one of the most dynamic and resilient in the world. Productivity growth is a common thread that ties nearly all positive economic news together and plays a central role in our international competitiveness.

Much of this *Report* explores the role of productivity and productivity-related issues in the continuing expansion of the U.S. economy. Policymakers face a challenge: productivity growth is important for economic growth and many of the underlying issues that they are trying to solve, but there is no single cause of productivity and no single policy to spur its growth (see Chapter 2). Tax policy can be structured to encourage productivity growth (see Chapter 3). Entitlement programs, on the other hand, may indirectly weigh on productivity growth if not reformed (see Chapter 4). Open commerce and financial markets allow productivity to flourish (see Chapters 7-9). Economists discuss productivity growth using macroeconomic data, but its result is most importantly seen in increases in individual Americans' standards of living.

Chapter 1: The Year in Review and the Years Ahead

The economic expansion continued for the fifth consecutive year in 2006. This economic growth comes despite numerous headwinds, and results from inherent U.S. economic strengths and pro-growth policies. Chapter 1 reviews the past year and discusses the Administration's forecast for the years ahead. The key points are:

- Real GDP posted above-average 3.4 percent growth in 2006. The composition of growth changed, with more coming from exports and business structures investment, while residential investment flipped from contributing to GDP growth in 2005 to subtracting from it in 2006. Consumer spending remained strong.
- Labor markets continued to strengthen, with the unemployment rate dropping to 4.6 percent and payroll job growth averaging 187,000 per month. Real average hourly earnings accelerated to a 1.7 percent increase during the 12 months of 2006.
- Energy prices rose sharply in the first half of the year, but then declined just as sharply in the second half.

Chapter 2: Productivity Growth

Productivity growth rarely makes the headlines, but is important to the Nation because higher productivity growth improves the outlook for economic issues such as standards of living, inflation, international competitiveness, and long-run demographic challenges. Chapter 2 reviews the sources of the recent strength in productivity growth, highlighting the role that flexible markets and entrepreneurship play in explaining cross-country differences. It also explains the benefits of productivity growth and discusses how policymakers can further promote it. The key points are:

- Recent productivity growth has been primarily driven by efficiency growth (growth in how well labor and capital inputs are used) and by capital deepening (growth in the amount of capital that workers have available for use).
- Openness to international trade and investment, and improvements in the education and training of the U.S. workforce, will continue to be important to long-run productivity growth.
- Policies that encourage capital accumulation, research and development, and increases in the quality of our education system can boost productivity growth.

Chapter 3: Pro-Growth Tax Policy

Chapter 3 discusses the advantages of adopting a more pro-growth tax system. It reviews recent changes that have reduced tax distortions on capital investment decisions, and evaluates options to reduce such distortions further. The key points are:

- The goal of pro-growth tax policy is to reduce tax distortions that hamper economic growth. Most economists agree that lower taxes on capital income stimulate greater investment, resulting in greater economic growth, greater international competitiveness, and higher standards of living.
- The tax code contains provisions that discourage investment and create distortions that affect the level, distribution, and financing of capital investment.
- Estimates from research suggest that removing these tax distortions to investment decisions could increase real gross domestic product (GDP) by as much as 8 percent in the long run.
- Since 2001, temporary changes in the tax code have reduced the tax on investment. These pro-growth policies have stimulated short-run investment and economic growth. However, the temporary nature of the provisions eliminates desirable long-run economic stimulus.

Chapter 4:
The Fiscal Challenges Facing Medicare

Social Security, Medicare, and Medicaid are three entitlement programs in the United States that provide people with important economic security against financial risk. However, the projected long-term growth in entitlement spending is unsustainable because of the pressure it puts on future Federal budgets. It is crucial that reforms to these programs preserve the protection against financial risk that these programs provide without having negative effects on economic growth. Chapter 4 focuses on Medicare by examining the main reasons for its projected financial pressures and by discussing ways to improve the efficiency of the program and thus slow the growth of Medicare spending. The key points are:

- Medicare spending is growing quickly, primarily because of the demographic shift to an older society and the increases in per-beneficiary medical spending driven largely by new technologies.
- Rewarding providers for supplying higher-quality care and improving incentives for patients to choose higher-value care can both increase the efficiency and slow the growth of Medicare spending.

Chapter 5: Catastrophe Risk Insurance

Insuring economic losses arising from large-scale natural and manmade catastrophes such as earthquakes, hurricanes, and terrorist attacks poses challenges for the insurance industry and for Federal and State governments. Chapter 5 examines the economics of catastrophe risk insurance. The key points are:

- In insurance markets, as in other markets, prices affect how people weigh costs and benefits. Artificially low insurance prices can discourage people from adequately protecting against future losses. For example, subsidized property insurance prices may stimulate excessive building in high-risk areas, potentially driving up future government disaster relief spending.
- Government intervention in insurance markets can have unintended consequences, such as limiting the availability of insurance offered by private firms.
- Insurers manage catastrophe losses by being selective about which risks to insure, designing insurance contracts to provide incentives for risk-reducing behavior, and charging prices that are high enough to enable them to diversify risk over time or transfer risk to third parties. By managing and pricing risk more effectively, government insurance programs can reduce the burden they impose on taxpayers and minimize negative effects on private insurance markets.

Chapter 6: The Transportation Sector: Energy and Infrastructure Use

The transportation sector accounts for the majority of the petroleum consumed in the United States and—whether plane, train, ship, or automobile—almost all transportation is powered by petroleum. Understanding the petroleum market, and the ways in which consumers and firms respond to changes in world oil prices, is key to understanding the transportation sector. In addition to petroleum, the transportation sector also relies heavily on infrastructure. The key points of Chapter 6 are:

- Recent increases in the price of oil and the external costs of oil have led to renewed interest by markets and governments in the development of new alternatives. Government can play a role in ensuring that external costs are taken into account by markets, but ultimately markets are best suited to decide how to respond.
- Cars and light trucks are the largest users of petroleum. As a result, the fuel economy of the vehicles purchased and the number of miles that they are driven have a large effect on oil consumption.
- Congestion is a growing problem in American urban areas. Cities and States have shown a growing interest in and capacity for setting prices for road use during peak periods to reduce the full economic costs of congestion.

Chapter 7: Currency Markets

The need for international transactions provides the impetus for a huge, well-functioning market that facilitates currency conversions and allows global economic integration and trade to occur smoothly and quickly at low cost. Both by volume of trade and ease of making transactions, currency markets today are the world's deepest, most liquid markets. Currency markets range from common markets where parties simply exchange one currency for another to sophisticated markets where parties buy and sell currencies far into the future. The key points of Chapter 7 are:

- Foreign-exchange markets allow firms to trade goods and services across borders, and to manage the risks they face from fluctuations in the price of their domestic currency.
- As with any other good, the exchange value of a currency is determined by its supply, as well as the demand for the country's assets, goods, and services.
- Over much of the 20th century, countries tended to favor fixed exchange rates, but in recent decades there has been a shift toward freely floating exchange rates.
- Monetary and exchange-rate policies are tightly linked. A nation's government must decide between controlling its exchange rate and controlling its domestic inflation rate.

Chapter 8: International Trade and Investment

The United States derives substantial benefits from open trade and investment flows. Over many decades, increased trade and investment liberalization has been an important catalyst for greater productivity growth and rising average living standards in the United States. The key points of Chapter 8 are:

- Looking ahead, international trade liberalization in services presents significant opportunities for U.S. workers, firms, and consumers.
- Foreign direct investment (FDI) flows into the United States benefit the U.S. economy by stimulating growth, creating jobs, promoting research and development that spurs innovation, and financing the current account deficit.
- U.S. direct investment abroad is an important channel of global market access for U.S. firms. U.S. multinational companies have contributed to productivity growth, job creation, and rising average living standards in the United States.

Chapter 9: Immigration

The United States is a nation of immigrants and a nation of laws, and we value both historical legacies. Immigrants continue to make positive contributions to our Nation and our economy, yet our current immigration laws have proven difficult to enforce and are not fully serving the needs of the American economy. The key points of Chapter 9 are:

- International differences in economic opportunities and standards of living create strong incentives for labor migration. Once established, migration flows from a certain region tend to be self-perpetuating.
- Foreign-born workers make significant contributions to the American economy, but not all Americans gain economically from immigration. Foreign-born workers tend to be concentrated at the low end and the high end of the educational spectrum relative to native-born workers.
- Immigration policy plays a key role in determining the volume and composition of the foreign-born workforce. Comprehensive immigration reform can help ensure an orderly, lawful flow of foreign-born workers whose presence continues to benefit the American economy.

C H A P T E R   1

The Year in Review and the Years Ahead

The expansion of the U.S economy continued for the fifth consecutive year in 2006. Economic growth was strong, with real gross domestic product (GDP) growing at 3.4 percent during the four quarters of 2006. This strong economic growth comes in the face of numerous headwinds and resulted from the inherent strengths of the U.S. economy and pro-growth policies such as tax relief, regulatory restraint, and opening foreign markets to U.S. goods and services. Growth in the first quarter rebounded from the effects of the 2005 hurricanes, including a recovery in consumer confidence and consumer spending, and the rebuilding of oil and natural gas infrastructure in the Gulf of Mexico. Although growth slowed in the middle two quarters of the year, the overall pace of real activity was strong in the face of near-record inflation-adjusted prices of crude oil and a sharp decline in home construction. On the inflation front, energy prices fell substantially towards the end of the year, allowing overall consumer price inflation to moderate in 2006; however, price inflation increased for goods and services other than food and energy. In response to these output and inflation developments, the Federal Reserve continued raising the federal funds rate through June, and then held it constant for the rest of the year. The Administration forecast calls for the economic expansion to continue in 2007, but we must continue to pursue pro-growth policies such as those designed to keep tax relief in place, restrain government spending, slow the rate of health care inflation, enhance national energy security, and expand free and fair trade.

This chapter reviews the economic developments of 2006 and discusses the Administration's forecast for the years ahead. The key points of this chapter are:

- Real GDP posted strong 3.4 percent growth in 2006, up from the 3.1 percent 2005 pace. The composition of aggregate demand changed from preceding years. More growth came from exports and business structures investment, while residential investment flipped from contributing to GDP growth in 2005 to subtracting from it in 2006.
- Labor markets continued to strengthen, with the unemployment rate descending to 4½ percent in the fourth quarter , and payroll job growth averaging 187,000 per month.
- Energy prices, which rose through August and then declined, dominated the movement of overall inflation in the consumer price index. Core inflation (which excludes food and energy inflation) moved up from 2.2 percent during the 12 months of 2005 to 2.6 percent in 2006, with much of this upward trend due to an acceleration in the amount that

renters pay for apartments and other rental properties and the estimated rent on owner-occupied housing. Energy prices fell sharply from September through October, and core inflation fell toward the end of the year.

- Real average hourly earnings accelerated to a 1.7 percent increase during the 12 months of 2006, reflecting solid labor markets combined with tamer energy prices.
- The Administration's forecast calls for the economic expansion to continue in 2007 and beyond, although the pace of expansion is projected to slow somewhat from the stronger growth of recent years. The unemployment rate is projected to edge up slightly in 2007, while remaining below 5 percent. Real GDP growth is projected to continue at around 3 percent in 2008 and thereafter, while the unemployment rate is projected to remain stable and below 5 percent.

Developments in 2006 and the Near-Term Outlook

The economy went through a period of rebalancing during 2006, with faster growth in business structures investment and exports partially offsetting pronounced declines in homebuilding. At the same time, consumer spending continued to grow.

Consumer Spending and Saving

Consumer spending sustained its strong growth during the four quarters of 2006 (rising 3.7 percent in real terms), continuing its 15-year pattern of rising faster than disposable income. Several factors helped to keep spending elevated, and as a result, kept saving down (according to the official definition in the national income and product accounts (NIPA)). These factors included rising energy costs (through the third quarter), rising wealth, and falling unemployment rates. As a result, the personal saving rate fell to a negative 1.0 percent for the year as a whole—its lowest annual level during the post-World War II era. Despite the negative saving rate, Americans continue to build wealth in the form of *capital gains* (the rise in asset prices), which are not included in the definition of saving in the NIPAs. The declining saving rate continues a long-term trend which began in the 1980s.

Energy Expenditures

World demand for crude oil increased from 79.74 million barrels per day in 2003 to 84.18 million barrels per day during the first three quarters of 2006. The United States accounted for about one-eighth (0.5 million barrels per day) of this higher (4.4 million barrel per day) pace of crude oil consumption. Most

of this increase in world demand was accounted for by non-OECD countries (up 4.1 million barrels per day). Consumption of the non-U.S. OECD countries fell 0.2 million barrels per day. In the face of this increase in world oil demand, the supply available to U.S. consumers was restrained, and consumers paid higher prices to maintain their consumption.

With the rise in energy prices, nominal energy purchases rose sharply. That consumers altered their spending patterns only slightly contributed to the fall in the saving rate. Consumer energy prices increased 29 percent relative to nonenergy prices (according to the NIPA price indexes) from the fourth quarter of 2003 to the fourth quarter of 2006, while real consumption of energy per household fell only slightly, by 2.1 percent. Between 2004 and 2006, consumers appear to have maintained both energy and nonenergy consumption by reducing their saving. Consumers' response to persistently high energy prices is likely to emerge gradually, as consumers economize on energy consumption and possibly on nonenergy consumption.

Wealth Effects on Consumption and Saving

The rise in household wealth has also played a role in the decline of the saving rate. During the late 1990s and again during the past 3 years, a strong rise in household net worth coincided with a sizeable increase in consumer spending relative to disposable personal income (see Chart 1-1).

Chart 1-1 Consumption and Net Worth (Relative to Disposable Personal Income)
Consumption gains from 2004 to 2006 were partly supported by an increase in net worth (wealth), with rises in housing and stock market wealth accounting for most of this increase.

Sources: Department of Commerce (Bureau of Economic Analysis), Federal Reserve Board, and Council of Economic Advisers.
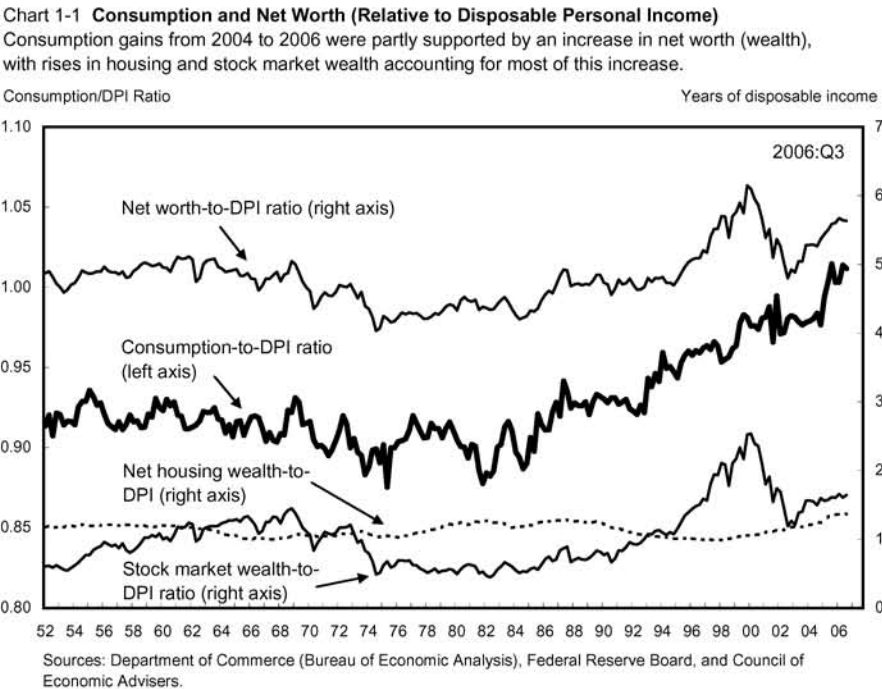
Despite the negative saving rate during 2006, Americans continued to build wealth because of capital gains. During the four quarters ending in the third quarter of 2006, the household wealth-to-income ratio increased 0.04 years, to 5.63 years of income. (The units of the wealth-to-income ratio are years because wealth is measured in dollars while income is measured in dollars per year. That is, total household wealth in the third quarter of 2006 represents the equivalent of 5.63 years of accumulated income.) More than half of the increase during these four quarters was accounted for by an increase in stock market wealth. Housing wealth (net of mortgage debt) also edged up relative to income over these four quarters, but by much less than its increases during the preceding 2 years. By the third quarter of 2006, the overall wealth-to-income ratio was well above the ratio over most of the past 50 years.

Personal and National Saving

Consumer responses to the rise in energy prices and increases in the wealth-to-income ratio lowered the personal saving rate to negative 1.0 percent in 2006. The *personal saving rate,* the rate at which households save, has been declining since the mid-1980s.

Corporate net saving takes the form of retained earnings which are not paid out to shareholders. (*Net* saving excludes funds used to replace worn out capital goods.) Retained earnings add to the wealth of corporate shareholders and supply funds for new investment. Corporate net saving rose to 3.8 percent of gross domestic income (GDI) during the first three quarters of 2006, its highest level since the 1960s. (GDI is the economy-wide sum of all sources of income and differs from GDP only by measurement error.) But even with these high levels of net corporate saving, *net private saving* (the sum of personal and corporate saving) was only 3.1 percent of GDI during the first three quarters of 2006, near its lowest level in the post-war period.

A still broader measure of net saving—*net national saving*—is the sum of government and private (personal plus corporate) net saving. When the Federal government runs a deficit (spends more than it collects in tax revenue), Federal saving is negative, as it was in 2006. Because the Federal deficit declined substantially in 2006, and because corporate saving rose, net national saving (which was negligible in 2005) rose to 2.0 percent of GDI during the first three quarters of 2006, its highest level since early 2002. *Gross* national saving, which includes funds for replacing worn out capital goods, is higher than net saving (13.8 percent versus 2.0 percent during the first three quarters of 2006), but shows similar historical fluctuations.

Projected Consumer Spending

Looking ahead, real consumer spending during the four quarters of 2007 is expected to grow less than 3 percent, down from an average of 3.5 percent during the past 3 years. This projected rate is slightly less than the projected

2007 growth of *real disposable personal income* (household income less taxes, adjusted for inflation), and so the saving rate is forecasted to edge up. During the longer term, real consumption is projected to increase at about the same pace as real GDP and real income.

Housing Prices

Nationally, housing prices increased less in 2006 than in 2005. An inflation-adjusted version of the housing price index (the nominal version of which is compiled by the Office of Federal Housing Enterprise Oversight from new home sales and appraisals during refinancing) increased at an average annual rate of 6.4 percent from 2000 to 2005, and then slowed to a 2.6 percent annual rate of increase in the first three quarters of 2006. (These inflation-adjusted prices are deflated by the consumer price index.) Looking back, the cumulative increase in inflation-adjusted housing prices during the 6 years from 1999 to 2005 is one of the largest on record, exceeded only by the period immediately following the Second World War. Since 1929, periods of rising real prices have been linked to increases in the share of the gross national product allocated to home construction (see Chart 1-2). The 6.4 percent annual rate of increase in the relative price of housing from 2000 to 2005 was associated with an increase in the residential construction share of GDP from 4.6 percent to 6.2 percent.

Chart 1-2 Residential Investment and House Prices
Real house price increases in 1944-46, 1977-79, and 1999-2005 led to rises in the share of GDP devoted to residential investment.
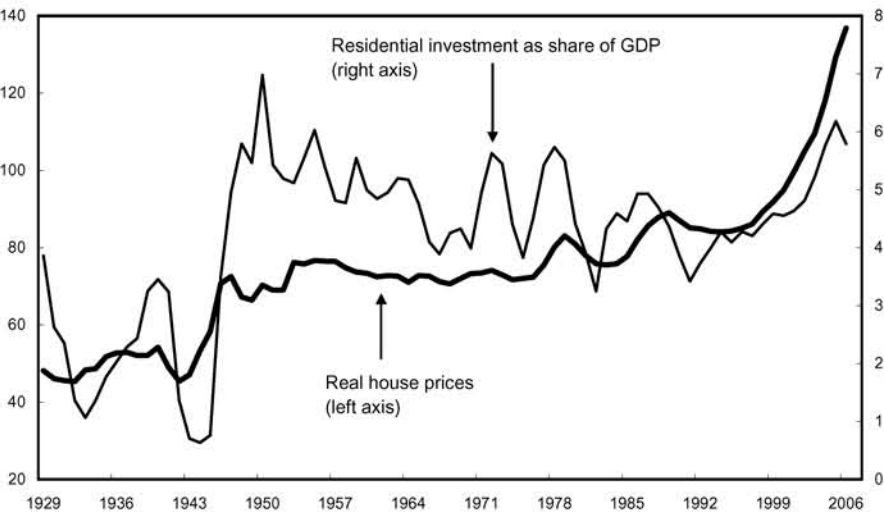
Source: Bureau of Economic Analysis; real house prices from 1929-1975 are courtesy of Robert Shiller; real house prices 1975-2006 are from the Office of Federal Housing Enterprise Oversight (OFHEO) and are deflated by the CPI-U-RS from the Bureau of Labor Statistics; 2006 real house price is the average of the 2nd and 3rd quarters.

Although relative housing prices (that is relative to the consumer price index (CPI)) increased in almost all metropolitan areas during the 5 years from 2000 to 2005, the increases were concentrated in a few high-profile markets; increases in most areas were only modest. For example, real prices in Los Angeles increased at a 14.3 percent annual rate, but real price increases in 71 percent of metropolitan areas were less than the 6.4 percent national average. Most house price changes reflect local conditions (such as local economic and population growth, tastes, and geographic and zoning limitations on construction). In areas with restricted supply, small changes in demand may translate into large price changes.

Although house-price increases during these 5 years were concentrated in a few markets, the decline in mortgage rates from 2000 to 2005 was one common factor that may have helped raise home prices across the nation. Because of the drop in mortgage rates, prices could increase 4.4 percent per year during this period without raising the monthly mortgage payment.

Residential Investment

Every major measure of housing activity dropped sharply during 2006, and the drop in real residential construction was steeper than anticipated in last year's *Report*. New home sales fell 27 percent from a peak in October 2005 through July 2006, a period when rates on conventional mortgages moved up about 70 basis points. (A basis point is one one-hundredth of a percentage point.) Sales then edged up during the 5 months from August through December, when mortgage rates dipped lower. Builders reacted sharply to the early-2006 drop in sales so that housing starts, which peaked at an annual rate of 2.27 million units in the beginning of the year, fell to slightly more than 1.6 million units by the end of the year. The drop in home construction activity subtracted roughly 0.7 percentage point from the annual rate of real GDP growth in the second quarter, and 1.2 percentage points in the second half of the year. Furthermore, even if housing starts level off at their current pace, normal lags between the beginning and completion of a construction project imply that residential investment will subtract from GDP growth during the first half of 2007.

During 2006, employment in residential construction fell, as did production of construction materials and products associated with new home sales (such as furniture, large appliances, and carpeting). Yet despite these housing sector declines, the overall economy continued to expand (see Box 1-1).

Box 1-1: Indirect Effects of the Housing Sector

Thus far, the sharp drop in homebuilding has had few consequences for the rest of the economy. Employment fell in sectors related to new home construction and housing sales. Despite these repercussions, overall payroll employment continued to increase, the unemployment rate continued to fall, and real consumer spending continued to move upward through the end of 2006.

Although residential investment fell sharply, real GDP growth during 2006 was sustained by increases in other forms of investment. As can be seen in the chart below, private nominal nonresidential construction (that is, business construction of office buildings, shopping centers, factories, and other business structures) grew rapidly in the first three quarters of the year and moved up a bit further in the fourth quarter. Nonresidential construction draws from some of the same resources (such as construction labor and materials) as the residential construction sector. The high level of residential investment during the past couple of years may have limited the growth of investment in nonresidential structures. While the case for housing crowding out other sectors is strongest for nonresidential investment, residential investment competes with all other sectors of production in credit and labor markets. A drop in the share of the economy engaged in housing could provide some room for other sectors to grow.

Private Construction

Although residential construction has fallen sharply from its peak, nonresidential investment continues to grow and absorb some of the resources formerly used in the residential sector.

Dollars (billions), seasonally adjusted at an annual rate

Source: Census Bureau.
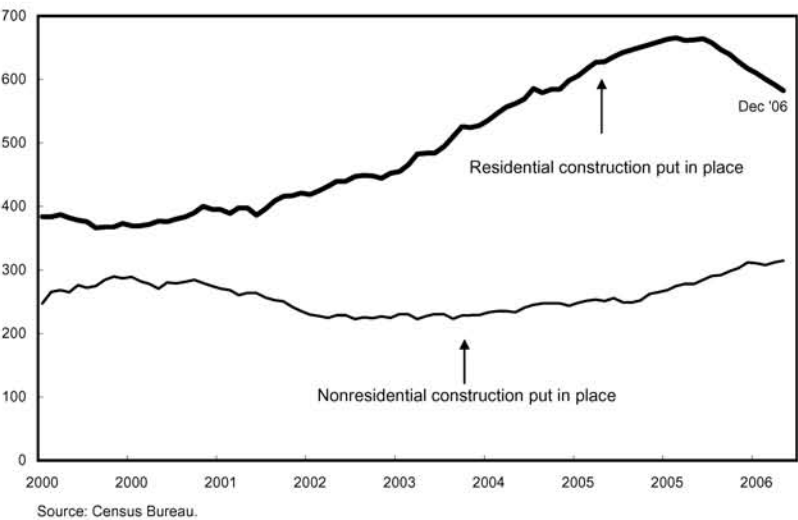
continued on the next page

The housing market could also affect the rest of the economy through the wealth channel. That is, declines in housing prices could reduce household net worth and thereby reduce consumption. The increase in housing prices during 2000–2005 contributed noticeably to the gain in the ratio of household wealth to income (shown earlier in chart 1-1) and supported growth in consumer spending. Some of this support may have been facilitated by homeowners taking out larger mortgages after their homes appreciated in value. In contrast, housing wealth decelerated in the second and third quarters of 2006, while the stock market accounted for most of the gain in the wealth-to-income ratio. Thus far, national measures of housing prices have not declined, and negative effects through the wealth channel have not occurred.

In addition to incomes and mortgage rates, the number of homes built is underpinned by demographics. Homebuilding during 2004 and 2005 averaged about 2.0 million units per year, in excess of the roughly 1.8-to-1.9-million unit annual pace of starts that is consistent with the pace of household formation implied by demographic models. As a result, the pace of homebuilding will tend to be drawn below this level for long enough so that the above-trend production of 2004 and 2005 will be offset by below-trend production. The construction of new homes has fallen rapidly, however, and this offset may well be complete sometime during 2007. Looking further ahead, the residential sector is not expected to make noticeable positive contributions to real GDP growth until 2008 and beyond.

Business Fixed Investment

During 2006, real business investment in equipment and software grew 5 percent, slower than the 7 percent average pace during the 3 previous years. Its fastest-growing components included computers, as well as machinery in the agricultural and service sectors. Investment in mining and oil field machinery was also strong, likely in response to elevated crude oil prices, and to the need to replace Gulf of Mexico facilities damaged by the 2005 hurricanes. Investment in heavy trucks has been solid throughout 2006 as trucking firms have been buying in advance of new environmental regulations (on particulate matter emissions issued in 2000 that became effective in 2007), which will raise heavy truck prices in 2007. Aircraft investment, however, declined sharply for the second consecutive year. Software investment posted a strong 7.9 percent gain in 2006, but since 2000, it has grown at only a 3.7 percent annual rate, a noticeable deceleration from the roughly 16 percent annual rate of growth during the 1990s.

The turnaround in investment in business structures (that is, nonresidential construction) during 2006 has been dramatic, with growth at 12 percent, up from an anemic 2 percent gain during 2005. Growth in 2006 was strongest for office buildings, multi-merchandise centers, lodging facilities, and recreational structures. Investment in petroleum and natural gas structures also grew rapidly, reflecting high petroleum and natural gas prices and the reconstruction of the Gulf of Mexico capacity. Investment continued to fall, however, in air transportation structures and medical buildings.

Business investment growth is projected to remain strong in 2007, somewhere in the neighborhood of the 9 percent annual rate of growth during the first three quarters of 2006. Continued growth in output combined with a tight labor market are expected to maintain strong demand for new capital equipment at the same time as corporations are flush with funds for these investments. The financial environment for these investments is favorable. *Cash flow* (the internally generated funds that are available for corporate investment) was at a record 10.3 percent average share of GDP in the first three quarters of 2006, while nonresidential investment (at 10.5 percent of GDP) was close to its historical average. In the longer run, business investment is projected to grow only slightly above the growth rate of real GDP.

Business Inventories

Inventory investment was fairly steady during 2006, and had only a minor influence on quarter-to-quarter fluctuations. Real nonfarm inventories grew at an average $44 billion annual pace during 2006, a 3.0 percent rate of growth that is roughly in line with the pace of real GDP growth over the same period. Coming off a long-term decline, the inventory-to-sales ratio for manufacturing and trade (in current dollars) remained relatively flat during the first half of the year, but began to pick up in August.

Inventory investment is projected to be approximately stable during the next several years, as is generally the case for periods of stable growth. The overall inventory-to-sales ratio is expected to continue trending lower.

Government Purchases

Real Federal consumption and gross investment grew 2.4 percent during 2006. This was the third consecutive year of growth at roughly 2 percent. Defense spending accounted for all of the increase during the four-quarter period, while nondefense purchases fell. The quarterly pattern of these Federal purchases has been volatile with sizeable increases in the first and fourth quarters of the year. Most of the first-quarter surge was in defense components.

Federal outlays (which include purchases, investment, and transfers such as Social Security) were boosted by a $111 billion appropriation in fiscal year

(FY) 2006 for reconstruction and relief efforts arising from the 2005 hurricanes. In addition, the supplemental defense spending package for on-going operations in Afghanistan and Iraq was $70 billion for FY 2006 and was passed in mid-June. An additional $70 billion emergency funding was provided in the regular defense appropriation act passed at the end of September 2006. Another supplemental appropriation for defense is likely for FY 2007.

Nominal Federal revenues grew 15 percent in FY 2005 and 12 percent in FY 2006. These rapid growth rates exceeded growth in outlays and GDP as a whole, and the U.S. fiscal deficit as a share of GDP shrank from 3.6 percent in FY 2004 to 2.6 percent in FY 2005 to 1.9 percent in FY 2006.

State and local government purchases rose 3 percent during 2006, up noticeably from rates below 1 percent during each of the 3 previous years. In the wake of the 2001 recession, this sector fell sharply into deficit in 2002. Revenues began to recover in 2003, and by the first half of 2006 the sector was out of deficit, allowing for an increase in state and local consumption and investment. This pattern of delayed response to downturns resembles the past several business-cycle recoveries.

Exports and Imports

Real exports of goods and services grew 9.2 percent during 2006, up from the 6.7 percent export growth over the four quarters of 2005. This accelera-tion reflects rapid growth among our trading partners. Real GDP among our OECD trading partners grew 2.9 percent during the four quarters of 2005, and is estimated to have grown at the same pace in 2006. In addition, the economies of some of our major non-OECD trading partners such as China, Singapore, and India are growing at rates of 7 to 10 percent per year, although these countries comprise only about 7 percent of our exports.

The fastest growth in U.S. goods and services exports was to India, but exports to China, Africa, and Latin America also grew rapidly. Despite the rapid export growth to these emerging economies, the European Union (EU) remains the major export destination, consuming nearly 25 percent of our exports. Within the EU, Great Britain's imports of American goods and serv-ices grew at a notable 18 percent annual rate during the first three quarters of 2006.

Real imports grew 3.1 percent in 2006, a slower pace than the 5.2 percent increase over the four quarters of 2005. Petroleum imports, which grew strongly in the fourth quarter of 2005 to replace production losses after the hurricanes, declined 10 percent during the four quarters of 2006. Real imports of nonpetroleum goods grew 5.3 percent over the same period, down slightly from the year-earlier pace.

The *current account deficit* (the excess of imports and income flows to foreigners over exports and foreign income of Americans) jumped to 7.0 percent of GDP in the fourth quarter of 2005, partly due to petroleum imports that replaced lost Gulf of Mexico production. The current account deficit then retraced some of its earlier increase in the first three quarters of 2006, when oil imports declined. It appears to have fallen further in the fourth quarter, reflecting the drop in prices of imported crude oil. Current account deficits mean that domestic investment continues to exceed domestic saving, with foreigners financing the gap between the two.

Employment

Nonfarm payroll employment increased 2.2 million during the 12 months of 2006, an average pace of about 187,000 jobs per month. The unemployment rate declined by 0.4 percentage point during the 12 months of the year to 4.5 percent. The average unemployment rate in 2006 (4.6 percent) was below the averages of the 1970s, the 1980s, and the 1990s.

Job gains were spread broadly across major sectors in 2006, with the natural resource and mining sector (which includes oil and natural gas extraction) experiencing the fastest growth rate (8.1 percent), likely due to increased demand for energy products. The service-providing sector accounted for 95 percent of job growth during the 12 months of 2006, a slightly larger contribution than would be suggested by its 83-percent share of overall employment. Within the service-providing sector, 24 percent of job growth was in professional and business service jobs. As noted, the service-providing sector accounted for almost all of the 2006 job gains. The goods-producing sector accounted for the remaining 5 percent of the gains (notably weaker than its 17-percent share of overall employment), a continuation of the long-term trend under which the goods-producing share of total employment has fallen in each of the past five decades. Within the goods-producing sector, employment growth during 2006 was concentrated in mining and construction, while manufacturing employment decreased for the ninth consecutive year.

Jobless rates fell among most major demographic segments of the population during the 12 months of 2006. The unemployment rate dropped for each of the four educational-attainment groups (less than high school, high school, some college, and college graduates). For the second consecutive year, the drop in the unemployment rate was most pronounced among those without a high school degree. After falling 0.8 percentage point during 2005 (when the overall rate fell 0.5 percentage point), the jobless rate in this group fell another 0.7 percentage point during the 12 months of 2006 (when the overall unemployment rate fell 0.4 percentage point). By race and ethnicity, the unemployment rate fell the

most during 2006 among Asians, Hispanics and blacks (1.4, 1.1 and 0.9 percentage points), in contrast to 0.2 percentage point for whites. By age, the jobless rate fell most among workers 25 to 34 years old. By sex, the jobless rate fell more among adult women than adult men.

Furthermore, the median duration of unemployment, an indicator that typically follows the business cycle with a substantial lag, declined from its December 2005 level of 8.5 weeks to a December 2006 level of 7.3 weeks, close to its historical average. The number of long-term unemployed (those out of work for more than 26 weeks) fell by 263,000 during the year.

The Administration projects that employment will increase at a pace of 129,000 jobs per month on average during the four quarters of 2007. In the long run the pace of employment growth will slow, reflecting the aging of the population and the diminishing rates of labor force growth. The Administration also projects the unemployment rate will average 4.6 percent over 2007, before edging up to 4.8 percent in 2008 and beyond.

Productivity

Labor productivity growth usually increases during the early stage of a business-cycle recovery but then falls somewhat as the cycle matures. Early in this most recent expansion, productivity grew at a remarkable 3.9 percent annual rate for the years 2002 and 2003 and then slowed to a 2.6 percent annual rate for the years 2004 and 2005. Overall productivity has grown at a vigorous 3.1 percent annual rate from the business-cycle peak in the first quarter of 2001 until the third quarter of 2006.

Although 1995 has been regarded as a watershed year for productivity because of the acceleration of productivity from a 1.5 percent to a 2.4 percent annual rate of growth, the further acceleration to a 3.1 percent annual rate of growth during 2001 to 2006 is striking, especially given a flat or diminished contribution from *capital deepening* (the increase in capital services per hour worked). (The time periods referred to are those shown in Table 1-2 later in this chapter.) The 1995–2001 acceleration may be plausibly accounted for by a pickup in capital deepening and by increases in organizational capital, the investments businesses make to reorganize and restructure themselves, in this instance in response to newly installed information technology. In contrast, capital deepening does not explain any of the post-2001 increase in productivity growth. The post-2001 acceleration in productivity therefore appears to be accounted for by factors that are more difficult to measure than the quantity of capital, such as continuing improvements in technology and business practices. (See *Chapter 2, Productivity Growth* for an extended discussion of this.)

Rather than assuming that the recent remarkable pace of productivity growth will continue, the Administration believes it is prudent to build a budget based on a forecast somewhat lower that the 3.1-percent pace of productivity growth since 2001. Productivity growth is projected to average 2.6 percent per year during the 6-year span of the budget projection— roughly equal to the average annual pace during the past decade.

Prices and Wages

As measured by the consumer price index (CPI), overall inflation fell from 3.4 percent during the 12 months of 2005 to 2.5 percent during 2006 (Chart 1-3). The drop in overall CPI inflation was almost entirely due to the deceleration of energy prices from a 17.1-percent increase in 2005 to a 2.9-percent increase in 2006. Food prices increased 2.1 percent during 2006, similar to the pace of the previous year. Core CPI prices (that is, excluding food and energy) increased 2.6 percent during 2006, up from a 2.2-percent increase a year earlier.

Chart 1-3 **Consumer Price Inflation**
Core CPI inflation moved up during the first 9 months of 2006, but then edged down. Energy inflation added to overall inflation through August, and pulled it down through November.
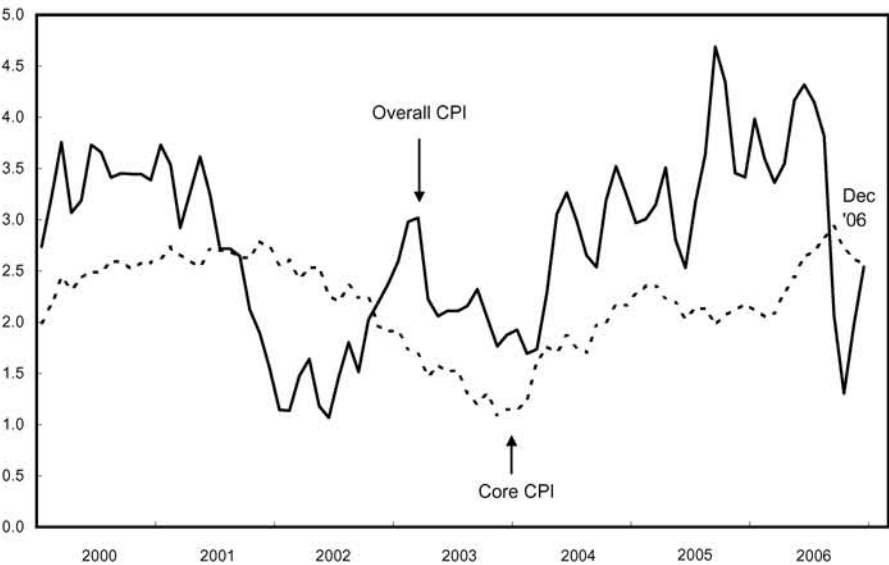
Source: Department of Labor (Bureau of Labor Statistics).

After rising sharply during 2004 and 2005, prices of petroleum products slowed to a 6.1 percent increase during the 12 months of 2006, as the sharp rise through August was reversed later in the year. Prices of natural gas, which had risen sharply during 2005, fell 14 percent during 2006. As of mid-January 2007, prices in futures markets suggested that crude oil prices will rise modestly during 2007, while natural gas prices will increase substantially.

The 0.4 percentage point acceleration of core CPI prices was accounted for primarily by rent of shelter (which consists primarily of rent paid by renters and by the rent on owner-occupied dwellings), which accelerated to a 4.3 percent rate of increase during the 12 months of 2006 from 2.7 percent in 2005. Some of the acceleration in core CPI prices may also have been a delayed reaction to the rapid increase in energy prices from mid-2003 to mid-2006, as the higher energy prices were absorbed into the prices of every service and commodity that requires inputs of energy or transportation. Econometric estimates (although imprecise) suggest that perhaps a quarter of a percentage point of the increases in the core CPI during the past year may be attributable to the past increases of these energy inputs. The Administration projects that the CPI will increase at a 2.6 percent annual rate during 2007 and 2008, about the same as the 2006 pace of the core CPI.

Hourly compensation (which is about 61 percent of nonfarm business output) has increased a bit faster in 2006 than in 2005. Nominal hourly compensation for workers in private industry increased 3.2 percent in 2006, up from 2.9 percent during the 12 months of 2005 according to the Employment Cost Index (ECI). All of this increase was from growth in wages and salaries (3.2 percent in 2006 versus 2.5 percent during 2005) while hourly benefits grew more slowly (3.1 percent versus 4.0 percent).

Another measure of hourly compensation published by the Department of Labor and derived from the National Income and Product Accounts has increased somewhat faster (at 4.3 percent) than the 3 percent increase in the ECI during the four quarters through the third quarter of 2006.

Unit labor costs have put little—if any—upward pressure on inflation thus far, and it appears unlikely that they will over the next year. Unit labor costs have increased at the same pace as the GDP price index, a 2.9 percent rate during the four quarters through the third quarter of 2006. The Administration expects the growth rate of hourly compensation to increase during 2007, as this nation's rapid productivity gains are shared by workers. But even with this acceleration in compensation, the expected strong pace of productivity growth will likely keep unit labor costs from putting upward pressure on inflation during 2007.

Moderate growth of hourly compensation and solid growth of productivity together with strong aggregate demand has driven the profit share of gross domestic income to its highest level since 1966.

Non-supervisory production-worker wages (which cover 82 percent of the private workforce) increased 4.2 percent (in nominal terms) during the

12 months through December 2006—an acceleration of 1.1 percentage points from the pace a year earlier. Real hourly wages of production workers increased 1.7 percent, a 2.1-percentage point acceleration from the pace a year earlier. The acceleration in real earnings reflects both the 1.1-percentage point increase in nominal wages and a 1 percentage point deceleration in consumer prices.

Among the many available measures of inflation, the Administration forecast focuses on two: the CPI and the price index for the GDP. The CPI measures prices for a fixed basket of consumer goods and services. It is widely reported in the press, and is used to index Social Security, the individual income tax, Federal pensions, and many private-sector contracts. The GDP price index covers prices of goods and services produced in the United States including consumption, investment, and government purchases. In contrast to the CPI, its weights are not fixed but move to reflect changes in spending patterns. Of the two indexes, the CPI tends to increase more rapidly in part because it measures a fixed basket of goods; the GDP price index increases less rapidly because it allows for households and businesses to shift their purchases away from items with increasing relative prices and toward items with decreasing relative prices. Among the differences, the GDP price index (which includes investment goods) places a larger weight on computers, which tend to decline in price (on a quality-adjusted basis). In contrast, the CPI places a much larger weight on rent and energy.

The "wedge," or difference between the CPI and the GDP measures of inflation, has implications for Federal budget projections. A larger wedge (with the CPI rising faster than the GDP price index) raises the Federal budget deficit because Social Security and Federal pensions rise with the CPI, while Federal revenue tends to increase with the GDP price index. For a given level of nominal income, increases in the CPI also cut Federal revenue because they raise the brackets at which higher income tax rates apply and affect other inflation-indexed features of the tax code.

During the 25 years from 1981 to 2005, the wedge between inflation in the CPI-U-RS (a historical CPI series designed to be consistent with current CPI methods) and the rate of change in the GDP price index averaged 0.32 percent per year. The wedge was particularly high during 2005 when the CPI increased 0.6 percentage point faster than the GDP price index. The wedge during 2005 reflected the 35 percent increase in crude oil prices, which have a larger weight in consumer prices (via their effect on refined-petroleum products) than in GDP as a whole. Because domestic production accounts for only about 35 percent of U.S. oil consumption, the weight of oil prices in GDP is roughly one-third of its weight in consumption. This effect unwound during the fourth quarter of 2006 when oil prices declined, causing the wedge to fall to -0.6 percentage point during the four quarters of 2006. From 2008 forward, the wedge is projected to average 0.3 percentage point.

Financial Markets

The Wilshire 5000 (a broad stock market index) increased 13.9 percent during 2006, while the Standard and Poor 500 (an index of the 500 largest corporations) increased 13.6 percent. This was the fourth consecutive year of stock market gains following 3 years of declines. The market has now recovered most of its losses since the March 2000 peak, at least in nominal terms.

Despite increases in short-term rates, yields on 10-year notes remained low, increasing only 9 basis points during the 12 months of 2006. The low level of long-term interest rates was due in part to low and stable long-run inflation expectations.

The Administration forecast of short term interest rates is roughly based on financial market data as well as a survey of economic forecasters. As of November 13, 2006, the date that the economic forecast was finalized, trading in financial futures suggested that market participants expected short-term rates to fall over the next several years, and the Administration's interest rate projections reflect those views. The Administration projects the rate on 91-day Treasury bills (5.1 percent on November 13) to remain flat in 2007 before edging down in 2008 and 2009. The short-term rate is projected to fall to 4.1 percent by 2012. At that level, the real rate on 91-day Treasury bills would be close to its historical average.

The yield on 10-year Treasury notes on November 13 was 4.61 percent, 48 basis points below the discount rate on the 91-day Treasury bills—a noticeable reversal of the usual pattern which shows higher rates for long-term yields. The Administration expects the 10-year rate to increase above the 91-day rate during 2007, eventually reaching a more normal spread of about 1.2 percentage points by 2010. An increase of a similar magnitude appears to be expected by market participants (as evidenced by higher rates on 20- and 30-year Treasury notes than on notes with 10-year maturities). As a result, yields on 10-year notes are expected to increase somewhat further, reaching a plateau at 5.3 percent from 2010 onward.

The Long-Term Outlook Through 2012

Coming off a fifth year of expansion, the U.S. economy is settling into a period of steady growth. Having reached a high level of resource utilization by year-end 2006, growth is likely to slow in 2007 and then will expand through 2012 at around 3.0 percent. Inflation will remain low and is expected to edge a bit lower, and the labor market will remain firm (Table 1-1). The forecast is based on conservative economic assumptions that are close to the consensus of professional forecasters. These assumptions provide a sound basis for the Administration's budget projections.

TABLE 1-1.— *Administration Forecast* [1]

Year	Nominal GDP	Real GDP (chain-type)	GDP price index (chain-type)	Consumer price index (CPI-U)	Unemploy-ment rate (percent)	Interest rate, 91-day Treasury bills [2] (percent)	Interest rate, 10-year Treasury notes (percent)	Nonfarm payroll employ-ment (millions)	Nonfarm payroll employ-ment (average monthly change, Q4-to-Q4 thousands)[3]
	Percent change, Q4-to-Q4				Level, calendar year				
2005 (actual)	6.4	3.1	3.1	3.7	5.1	3.1	4.3	133.5	160
2006.................	5.9	3.1	2.7	2.3	4.6	4.7	4.8	135.3	151
2007.................	5.5	2.9	2.5	2.6	4.6	4.7	5.0	137.0	129
2008.................	5.5	3.1	2.3	2.6	4.8	4.6	5.1	138.6	139
2009.................	5.3	3.1	2.2	2.5	4.8	4.4	5.2	140.2	126
2010.................	5.2	3.0	2.1	2.4	4.8	4.2	5.3	141.5	113
2011.................	5.0	3.0	2.0	2.3	4.8	4.1	5.3	143.0	118
2012.................	5.0	2.9	2.0	2.3	4.8	4.1	5.3	144.3	107

[1] Based on data available as of November 13, 2006.

[2] Discount basis.

[3] If the effect of the BLS benchmark adjustment were included, monthly job growth would average 202 and 191 thousand in 2005 and 2006 respectively. The level of payroll employment would be 133.7 and 136.2 million in these 2 years.

Sources: Council of Economic Advisers, Department of Commerce (Bureau of Economic Analysis and Economics and Statistics Administration), Department of Labor (Bureau of Labor Statistics), Department of the Treasury, and Office of Management and Budget.

Growth in GDP over the Long Term

The Administration projects that, following a slight pickup of growth from 2007 to 2008, real GDP will increase at a slowly diminishing rate from 2008 through 2012. Indeed, real GDP is projected to decelerate from a 3.1 percent rate of growth during the four quarters of 2008 to 2.9 percent by 2012. The average growth rate during this interval is roughly in line with the consensus of private forecasters for those years. After 2007, the year-by-year pace is close to the estimated growth rate of potential real GDP, a measure of the rate of growth of productive capacity. (An economy is said to be growing at its potential rate when all of its resources are utilized and inflation is stable. The supply-side components of potential GDP growth are presented in Table 1-2 and are discussed below). The unemployment rate is projected to edge up in 2007 (from its 4.5 percent level in the fourth quarter of 2006) and to plateau at 4.8 percent in 2008. As discussed below, potential GDP growth is expected to slow in the near term as productivity growth reverts toward its long-run trend (about 2.6 percent per year), and to slow further during the 2007-to-2011 period as labor force growth declines due to the retirement of the baby-boom generation.

The growth rate of the economy over the long run is determined by its supply-side components, which include population, labor force participation,

the ratio of nonfarm business employment to household employment, the length of the workweek, and labor productivity. The Administration's forecast for the contribution of the growth rates of different supply-side factors to real GDP growth is shown in Table 1-2.

TABLE 1-2.— *Supply-Side Components of Real GDP Growth, 1953–2012*
[Average annual percent change]

Item	1953 Q2 to 1973 Q4	1973 Q4 to 1995 Q2	1995 Q2 to 2001 Q1	2001 Q1 to 2006 Q3	2006 Q3 to 2012 Q4
1) Civilian noninstitutional population aged 16+ [1]	1.6	1.4	1.2	1.2	1.0
2) Plus: Civilian labor force participation rate	0.2	0.4	0.1	-0.3	-0.2
3) Equals: Civilian labor force [2]	1.8	1.8	1.4	1.0	0.8
4) Plus: Civilian employment rate	-0.1	0.0	0.3	-0.1	0.0
5) Equals: Civilian employment [2]	1.7	1.8	1.7	0.9	0.8
6) Plus: Nonfarm business employment as a share of civilian employment [2][3]	-0.1	0.1	0.4	-0.7	0.1
7) Equals: Nonfarm business employment	1.6	1.9	2.0	0.2	0.8
8) Plus: Average weekly hours (nonfarm business)	-0.3	-0.3	-0.1	-0.2	0.0
9) Equals: Hours of all persons (nonfarm business)	1.3	1.6	1.9	0.0	0.8
10) Plus: Output per hour (productivity, nonfarm business)	2.5	1.5	2.4	3.1	2.6
11) Equals: Nonfarm business output	3.8	3.1	4.3	3.0	3.4
12) Plus: Ratio of real GDP to nonfarm business output [4]	-0.2	-0.2	-0.5	-0.3	-0.4
13) Equals: Real GDP	3.6	2.8	3.8	2.7	3.0

[1] Adjusted by CEA to smooth discontinuities in the population series since 1990.

[2] BLS research series adjusted to smooth irregularities in the population series since 1990.

[3] Line 6 translates the civilian employment growth rate into the nonfarm business employment growth rate.

[4] Line 12 translates nonfarm business output back into output for all sectors (GDP), which includes the output of farms and general government.

Note: 1953 Q2, 1973 Q4, and 2001 Q1 are NBER business-cycle peaks. Detail may not add to total because of rounding.

Sources: Council of Economic Advisers, Department of Commerce (Bureau of Economic Analysis), and Department of Labor (Bureau of Labor Statistics).

As can be seen in the fourth column of the table, the mix of supply-side factors determining real GDP growth has been unusual since the business-cycle peak at the beginning of 2001. The high rate of productivity growth (3.1 percent at an annual rate, shown in line 10) has been partially offset by the decline in the participation rate (line 2) and the workweek (line 8). Also notable is the large and puzzling decline in the ratio of nonfarm business employment to household employment (line 6). This unusual decline reflects the slow growth of employment as measured by the payroll survey (which asks employers to report the number of jobs) relative to the more rapid growth of employment as measured by the household survey (which estimates the number of employed persons through a sample of households). This disparity

has been reduced somewhat by the just-issued benchmark revision to payroll employment, but has yet to be satisfactorily explained.

The participation rate fell, on net, from 2001 to 2006 (although it ticked up in 2006), and is projected to trend lower through 2012. The recent behavior stands in contrast to the long period of increase from 1960 through 1996. Looking ahead, the participation rate is projected to decline, reflecting the aging of the baby-boom cohorts, leading to more retirements and a likely increase in the share of people on disability pensions (see Box 1-2).

Box 1-2: Long-Term Prospects for Labor Force Participation

The overall rate of labor force participation is projected to decline as the baby-boom cohorts advance into age brackets with much lower participation rates. Participation in the labor force (by working or by looking for a job) declines as people age through their 50s and 60s, as can be seen in the following chart.

Labor Force Participation Rates By Age
This age-participation profile follows the same birth cohorts as they get older. Participation rates for men and women edge down during their 50's and fall sharply during their 60's.
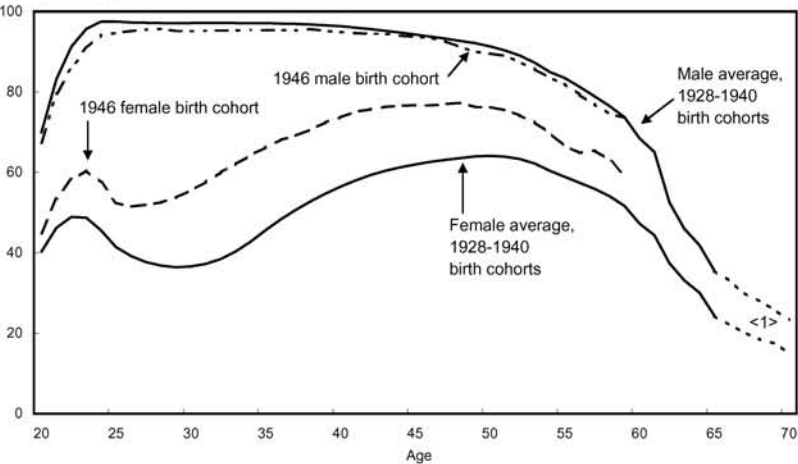
<1> Not all 13 cohorts are included in participation estimates for age 65 and older because some cohorts are still too young. These participation rates have been adjusted to account for the reduced number of cohorts.
Source: Department of Labor (Bureau of Labor Statistics) with interpolations by Council of Economic Advisers.

This chart shows the estimated average lifetime age-participation profile for the 13 cohorts born from 1928 to 1940. Men's participation is high (exceeding 90 percent) from age 24 through age 50, but then declines thereafter, dropping to 83 percent by age 55 and 36 percent by age 65. The rate of labor force exit is particularly rapid around 62, the

continued on the next page

Box 1-2 — *continued*

age at which one becomes eligible for early Social Security retirement benefits. In fact, about 40 percent of those eligible elect to begin collecting Social Security annuities at age 62, although this does not necessarily mean that they exit the workforce.

The difference between the age-participation profile of this 1946 cohort (the dotted lines) and those of its elders illustrates how participation rates have evolved over time. Female participation rates have moved sharply upward—in a roughly parallel shift. In contrast, male participation rates have changed little over time, moving down only slightly.

The current age distribution of the U.S. population is shown by the bars in the following chart, and the black line shows an estimate of the age distribution of the population in 2012. The large baby-boom cohorts (who were born between 1946 and 1964) are now 42 to 60 years old, and their aging will shift a sizeable fraction of the population into age brackets with lower participation rates, thus decreasing the share of the population in the high-participation ages.

U.S. Population By Age
Because of the aging of the baby-boom generation, the U.S. population in 2012 will have many more people of ages 55-65, and fewer of ages 35-45.
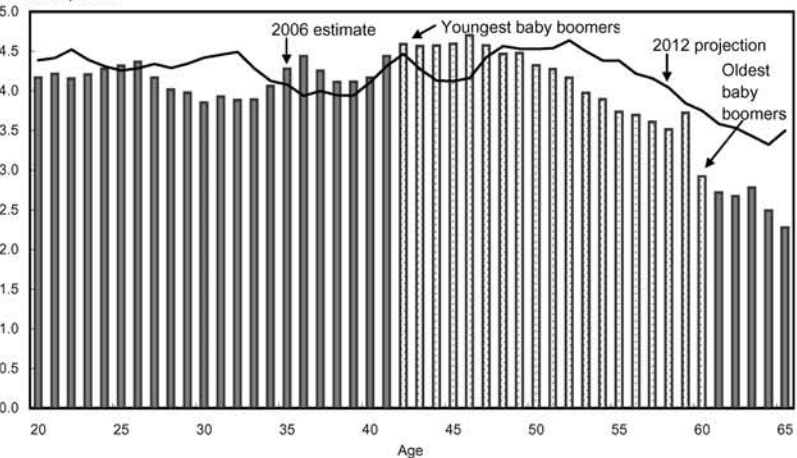
Note: October 2006 estimates from Census Bureau. The 2012 population is projected using growth factors from Census Bureau's 2004 Interim Projections applied to 2006 population estimates.
Sources: Department of Commerce (Census Bureau) and Council of Economic Advisers.

An extrapolation that moves the participation rate of each cohort along a path that parallels the 1928–1940 reference cohort and projects how the aging of the population translates into participation rates suggests an average participation rate decline of roughly 0.3 percent per year. A decline of this magnitude would alter a wide range of labor-market behaviors. In response to the emerging shortage of experienced workers, real wages are likely to increase and workweeks are likely to lengthen. Labor productivity is likely to increase as employers invest in labor-saving capital. And more immigrants may enter the U.S. labor force. The largest effect of the baby–boom retirements, however, is likely to be an endogenous effect on the labor force participation rate itself as developments in pay and pension arrangements evolve to induce higher participation rates among experienced workers than our extrapolation would suggest.

The Composition of Income over the Long Term

The Administration's economic forecast is used to estimate future government revenues, a purpose that requires a projection of the components of taxable income. The income-side projection is based on the historical stability of labor compensation as a share of gross domestic income (GDI). During the first half of 2006, the labor compensation share of GDI was 56.7 percent (according to the preliminary data available when the projection was finalized), slightly below its 1963–2005 average of 58.1 percent. From this jump-off point, the labor share is projected to slowly rise to 57.8 percent by 2012.

The labor compensation share of GDI consists of wages and salaries (which are taxable), non-wage compensation (employer contributions to employee pension and insurance funds—which are not taxable), and employer contributions for social insurance (which are not taxable). The Administration forecasts that the wage and salary share of compensation will be approximately flat between 2007 and 2012. Employer contributions to defined-benefit pension plans rose by almost 1 percentage point of total compensation between 2001 and 2002, boosting the growth of non-wage compensation. Contributions leveled off and then edged lower in subsequent years.

The capital share of GDI is expected to edge down from its currently high level before eventually reaching its historical average in 2012. Within the capital share, private depreciation is expected to increase (as a result of the strong growth of investment during the past 3 years). Profits during the first

three quarters of 2006 were about 12.2 percent of GDI, well above their post-1959 average of roughly 9 percent. Book profits (also known in the national income accounts as *profits before tax*) are expected to decline as a share of GDI.

The GDI share of other taxable income (rent, dividends, proprietors' income, and personal interest income) is projected to edge up slightly over the next 2 years.

Conclusion

With the rapid-growth period of the expansion fading into the past, the economy is currently going through a period of rebalancing, where higher growth of nonresidential investment and exports are offsetting the lower rates of housing investment. The economy is projected to settle into a steady state in which real GDP grows at about 3 percent per year, the unemployment rate creeps up towards a noninflationary level (of 4.8 percent) and inflation remains moderate and stable (about 2.2 to 2.6 percent on the CPI). Consumer spending is projected to grow in line with disposable income, and business investment and exports are projected to grow a bit faster than GDP as a whole. Economic forecasts are subject to error, and unforeseen positive and negative developments will affect the course of the economy over the next several years. Given the economy's fundamental strengths, however, prospects for continued growth in the years ahead remain good. Nonetheless, much work remains in making our economy as productive as possible. Later chapters of this *Report* explore how pro-growth policies such as tax reform, fiscal restraint, open commerce, and enhancing our energy security can enhance our economic performance.

Productivity Growth

News about economic issues focuses on topics such as inflation, international competitiveness, standards of living, and long-run demographic challenges. Productivity growth rarely makes the headlines. Why is productivity growth important to the nation? Because higher productivity growth improves the outlook for all of these issues. It helps keep inflation in check, makes it easier for American businesses and workers to compete, raises standards of living, and reduces the difficulty of meeting long-run demographic challenges by increasing the total amount of resources available.

Over the past 10 years, gross domestic product (GDP) per capita has grown faster in the United States than in almost every other advanced industrialized country. The United States owes its recent strong per capita growth in large part to strong labor productivity growth. A continuation of this productivity growth is essential to increasing real wages and maintaining the high standard of living in the United States.

To remain competitive, U.S. businesses must hold costs down by getting the most out of the inputs they use—that is, they must increase labor productivity. Similarly, for U.S. workers to earn higher wages than workers in other countries while competing in a global economy, U.S. labor productivity must exceed that of lower-wage countries.

Labor productivity growth also holds the key to dealing with the economic and fiscal challenges of a rapidly aging population. The total amount of goods and services produced in a country, measured by GDP, can grow only if productivity or hours of work increase. As the baby boomers (those born between 1946 and 1964) reach retirement, growth in total hours of work across the U.S. economy will slow, and the United States will have to depend increasingly on productivity growth to drive increases in GDP. While labor force growth will slow, the elderly population will expand relatively quickly. Strong GDP growth must continue in order to maintain the standards of living for both the working age and the dependent populations.

The amount that U.S. workers produce has grown at remarkable rates in recent years. Since 1995, productivity growth has averaged over 2.5 percent per year, compared to an average growth rate of about 1.4 percent per year over the preceding 20 years. Most other major industrialized countries suffered a slowdown in productivity growth between 2000 and 2005, but in the United States, growth accelerated to about 3 percent, the fastest productivity growth of any G7 country—Canada, France, Germany, Italy, Japan, the

United Kingdom, and the United States—over that period. Given that the United States' productivity was already among the highest and that these countries have similar access to technological improvements and financial markets, the sudden increase in U.S. productivity growth relative to other developed countries is especially impressive.

Table 2-1 illustrates how small differences in productivity growth rates can, over time, have large effects on the level of productivity and hence on the standard of living. When productivity doubles, twice as much output can be produced using the same level of labor. The table lists four different productivity growth rates that correspond to averages for different U.S. historical time periods, along with the number of years it would take to double the standard of living at that rate of growth. If productivity continues to grow at the rate from the most recent period (3.1 percent), the U.S. standard of living will double in about 23 years; at the slower productivity growth rate experienced during the 1973–1995 period (1.4 percent), doubling would take more than twice as long.

TABLE 2-1.— *Implied Doubling Rates for the Level of Productivity Using Historical Growth Rates*

	Productivity growth rate	Doubling time (in years)
1950 to 1973	2.6%	27.0
1973 to 1995	1.4%	49.9
1995 to 2000	2.5%	28.1
2000 to 2005	3.1%	22.7

Source: Department Labor (Bureau of Labor Statistics), Council of Economic Advisers calculations.

This chapter reviews the sources of the recent strength in productivity growth, highlighting the role that flexible markets and entrepreneurship play in explaining cross-country differences. It also explains the benefits of productivity growth and discusses how policymakers can further promote it. Key points are:

- Recent productivity growth has been primarily driven by *efficiency growth* (growth in how well labor and capital inputs are used) and by *capital deepening* (growth in the amount of capital that workers have available for use).
- Efficiency growth comes from developing new methods of production and new products. Entrepreneurship and competition make key contributions to such innovation.
- Investment in information technology (IT) capital and innovative new ways of using it have been important sources of productivity growth in many industries with particularly high growth rates.

- Openness to international trade and investment is especially important for fostering competition and thus productivity growth.
- Increases in the education and training of the U.S. workforce have been and will continue to be important to long-run productivity growth.
- Policies that encourage capital accumulation, research and development, and increases in the quality of our educational system can boost productivity growth over the long run.

The Basics of Productivity Growth: Framework and Recent Facts

Labor productivity measures the goods and services produced per hour of work. In the United States, the most commonly used measure of labor productivity is that for the nonfarm business sector, which excludes all levels of government, nonprofit institutions, households, and farms. Because output from nonbusiness entities is particularly difficult to measure, nonfarm business labor productivity is thought to best measure how labor productivity varies over time. For international comparisons of productivity, total output per hour worked is often used because data on hours by sector are not always readily available.

Factors That Increase Labor Productivity

What increases labor productivity? Research on this question usually divides changes in labor productivity into three sources: capital deepening, increases in skill, and efficiency gains.

Capital Deepening

Capital deepening happens when businesses invest in more or better machinery, equipment, and structures, all of which make it possible for their employees to produce more. Matching employees with better capital increases the number of goods employees produce in each hour they work. Examples of capital deepening include the purchase of more sophisticated machine tools for workers in the manufacturing sector, or a faster computer system for a travel agent. A business may add capital when it increases its workforce—for example, a travel agency might buy additional computers when increasing the number of travel agents it employs—but that does not constitute capital deepening if the amount of capital available *per worker* does not increase.

Farming provides a classic example of the benefits of using more and better capital. In 1830, it took a farmer 250 to 300 hours of work to produce 100 bushels of wheat; in 1890, with the help of a horse-drawn machine, the

time dropped to between 40 and 50 hours; in 1975, with the use of large tractors and combines, the 100 bushels could be produced in just 3 to 4 hours. While it is most likely that farmers were more educated in 1975 than they were in the 1830s, the change in the farmers' skills alone could not be the source of this dramatic efficiency gain; an important source is the use of better capital. Changing from a hoe to the tractor would be categorized as capital deepening, and the resulting increase in output is capital deepening's contribution to productivity growth.

Increases in Skill

Just as a worker who is paired with a better machine can produce more goods, a worker who learns a skill needed for production can produce more output in less time. For example, a worker who takes a class on how to use a computer increases the skill with which she uses the computer; the computer is no faster, but the worker's increased skill increases her output per hour worked and hence boosts her productivity. Workers increase their skills through additional education, training, on-the-job experience, and so on.

Efficiency Gains

Businesses achieve efficiency gains—more output with the same amount of input—when they devise better ways of organizing and using the equipment they own and the people they employ. Efficiency gains include both *process innovations,* which increase productivity by reducing the capital or labor needed to produce a unit of output, and *product innovations,* which increase productivity by increasing the value of output. For example, when Henry Ford began mass-producing Model T's, the Model T itself was a product innovation, while the moving assembly line was a process innovation. The combination of improved process and product allowed the Ford Motor Company to reduce its production costs and become more competitive.

A more recent example of process improvements that led to direct efficiency gains may also be helpful in illustrating this concept. Managers at a 3M tape-manufacturing plant increased productivity by reorganizing part of their production process. By moving machines such as glue coaters and tape slitters closer to the packing equipment and robotic transporters, 3M substantially increased labor productivity at its plant. The reorganization reduced the need to move output around the plant, and cut the length of the production cycle. In addition, with all the packing supplies located in one place, managers could see when they had more than they needed and could cut costs by reducing excess inventories of supplies. This improvement is an efficiency gain because the plant produced more output without increasing capital or labor. This example is typical of the innovative process: companies purchase and install new machines—from computers to conveyor belts—but it takes time and further innovation to learn how to take full advantage of the new machines.

Entrepreneurship (developing new ways of doing business and making risky investments to implement them) and competition partially determine the degree to which innovation contributes to labor productivity. If a business comes up with a new product or a new way of organizing production and spends the resources to try it out, and if the new way improves on the old, the business ends up with a higher level of profit and an incentive to expand. Innovation by one business is likely to have little direct effect on a nation's productivity growth, but competition forces other businesses to either come up with innovations of their own or to cede market share. When this happens, capital investment and labor flow to businesses with better methods of production, and productivity increases as a result.

Entrepreneurship occurs on both small and large scales; many large multinationals spend large sums on research and development in order to innovate and expand, but individual entrepreneurs who operate on a small scale may also innovate. The entry and growth of new businesses, combined with the exit of older, less productive businesses, has been found to be responsible for a substantial share of efficiency growth.

Productivity Growth in Recent Years

Chart 2-1 illustrates how increases in skill, capital deepening, and efficiency gains have contributed to productivity growth in recent years. It is important to note that the relative sizes of these contributions are only approximate and that some increases in the quality of labor and capital may be counted as efficiency gains. For example, economists can accurately measure education levels of the labor force, but on-the-job training is also commonplace and measuring the impact of this training on skill levels is difficult. Similar issues arise in adjusting for the quality of capital, particularly during periods of rapid technological changes. The net result is likely an understatement of skill increases and capital deepening, and a resulting overstatement of efficiency gains.

Chart 2-1 contrasts three periods, 1990–1995 (when U.S. productivity growth was relatively slow), 1995–2000 (when the pace of productivity growth quickened), and 2000–2005 (shows the most recent growth rate). Over these 15 years, skill increased at a fairly steady pace of about 0.3 percent to 0.4 percent per year. The sources of this increase are increased rates of college attendance and the increased experience of the workforce. Increases in skill have been an important source of long-run increases in labor productivity, and help explain why the United States has high income levels relative to other countries. Continuing a steady increase in skill is vital to maintaining solid productivity growth into the future, a topic discussed at more length in Chapter 2 of the 2006 Economic Report of the President.

But even when educational attainment among the young rises substantially, the skill level of the workforce as a whole evolves slowly. Because skill has

Chart 2-1 **Sources of Labor Productivity Growth**
Capital deepening and efficiency gains account for the increase in productivity growth since 1995.

Percent
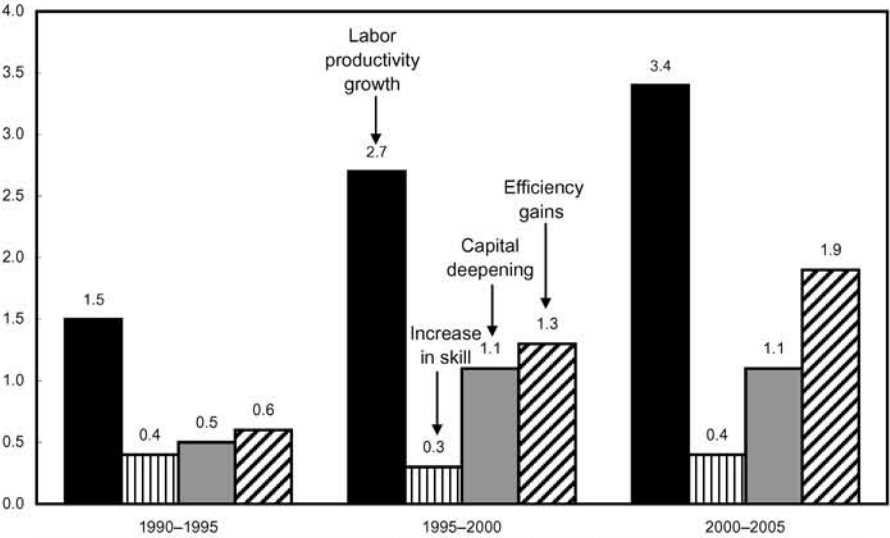
Note: Based on productivity for private business sector. Estimates reflect latest available data, which do not include benchmark GDP revisions for 2003–2005.
Source: Department of Labor (Bureau of Labor Statistics).

increased at a relatively steady rate, it cannot be the source of the recent acceleration in productivity growth. Instead, capital deepening and efficiency gains have been the key productivity-raising factors. Between 1995 and 2005, increases in the quality and quantity of the U.S. capital stock accounted for 1.1 percent per year in productivity growth in the United States, more than doubling the contribution of capital to productivity growth relative to the 1990 to 1995 period. The surge in productivity in the late 1990s resulted not just from a rapid increase in the number of machines used in U.S. production, but also from large quality improvements to the capital stock. Many of these improvements came from the revolution in information technology, which is commonly accepted as the initiating force behind the acceleration. But investment in IT capital alone was not the whole story. Firms needed to develop processes that best used the new capital. In many ways, the first increase in productivity growth (the higher growth rate between 1995 and 2000) was due to increased capital, while the second boost (in the period between 2000 and 2005) occurred as firms became better and better at using the new technology.

Productivity Growth and Worker Earnings

The previous section looked at the sources of recent productivity gains, but did not discuss what productivity gains mean for a worker's paycheck. This section examines how productivity growth affects average compensation and which groups have gained the most over time.

Productivity and Average Earnings

The economic gains from productivity growth reach workers directly through growth in employee compensation, where compensation includes wages and the contributions that employers make for benefits such as health insurance and for government programs such as unemployment insurance and Social Security. Chart 2-2 shows that over long periods of time, productivity and real compensation grow at about the same rate. Real wages have grown somewhat more slowly than compensation and thus productivity over the last 20 years. The reason for this difference is that non-wage compensation, particularly employer contributions for health insurance, has accounted for an increasing share of compensation over this time period.

Chart 2-2 Productivity and Real Compensation Grow Together
Productivity and compensation often diverge temporarily but grow together over the long run.
Index 1992 = 100
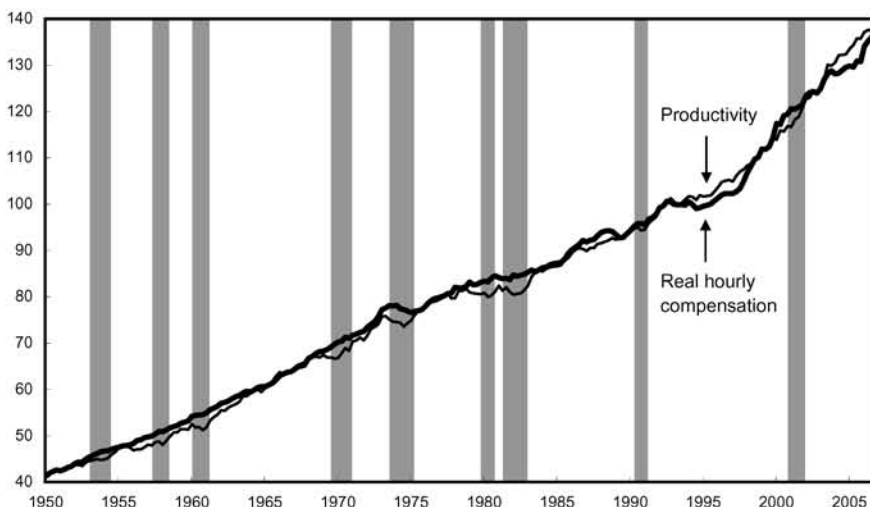
Note: These data cover all persons (including supervisory workers and proprietors) in the nonfarm business sector. Real hourly compensation is hourly compensation deflated by the price deflator for nonfarm business output. Shaded areas denote recessions.
Source: Department of Labor (Bureau of Labor Statistics).

Productivity growth is not a smooth process. Chart 2-2 shows that even in the recent time period, 1995 to 2005, when average productivity growth has been high, there are short periods of time where productivity growth appears to slow sharply or accelerate rapidly. Such changes in productivity growth are not uncommon. In addition, productivity sometimes grows faster than compensation, while sometimes compensation grows faster. Such short-term divergence in growth rates follows regular patterns and has been repeated many times. At times when productivity growth is particularly high, compensation growth tends to lag behind for a period of time before catching back up.

Why does compensation tend to lag behind productivity growth? When productivity growth is high, economic growth can happen without substantial employment growth. In other words, as productivity grows, businesses are able to expand output in response to increased demand without hiring more workers; the efficiency gains imply that each individual worker produces more output in the same amount of time. As the economy continues to expand, businesses once again begin to hire new employees, and the increased demand for workers begins to push up wages and compensation. Increased demand for workers leads to a period in which compensation growth exceeds productivity growth, and the two variables then converge for a while.

When productivity grows faster than compensation, businesses' profits tend to rise because the value of the goods and services they sell rises faster than their payroll costs. As a result, profits tend to rise during periods of rapid productivity growth. As tight labor markets bid up employee compensation, the increase in labor costs cuts into profits, and profits return to normal levels. In this process, profits vary more dramatically than employee compensation, falling much more sharply during recessions and then growing much more quickly in the early parts of the recovery. Because profits represent returns to earlier investments, very high profits in some years may not represent unusually large returns on investment because they may be offset by years of losses or unusually small profits.

Productivity and Income Differences

The productivity and compensation numbers used in this chapter describe averages, but over the last 30 years, the economic gains for some groups have not kept up with those averages, while the gains for other groups have been well above the average. These uneven gains have led to growing disparity (or inequality) in compensation and wages. The same competition for workers that makes average employee compensation track productivity growth over the long term will occur for particular groups of employees within the overall labor force. The compensation for groups whose productivity has increased relative to the rest of the labor force will increase relative to average compensation. A number of studies have shown that factors associated with higher

productivity—such as education and work experience—have also been increasingly associated with higher wages. This is consistent with the view that growing compensation disparity has been driven by faster growth in productivity for skilled workers than for the less skilled.

In the 1980s, the increase in disparity was seen both in falling wages at the bottom of the wage distribution and rising wages at the top. Since then, wages in the bottom half of the distribution have either been flat or have grown modestly while disparity has continued to increase in the upper part of the distribution. For example, between 1990 and 2005 the wage at the 10th percentile grew 13 percent while the median wage grew 10 percent, so the difference between them narrowed somewhat. The wage at the 90th percentile of the distribution grew 18 percent over that period, widening the gap between the upper tail of the distribution and the median.

Why have wage levels grown increasingly disparate? Changes in technology that increase the productivity advantages associated with skill—often termed skill-biased technical change—appear to be the most likely cause. That is, technological advances increased the productivity of skilled workers more than the productivity of the less skilled, leading employers to want to hire more skilled workers. In doing so, employers bid up the wages of skilled workers, widening the difference in pay associated with skill.

Why does skill-biased technical change appear to be the most reasonable explanation for this trend? The main reason is that the price that employers pay for skilled workers trended upward even while the supply of skilled workers continued to grow. For example, although the fraction of the workforce that is college educated has grown consistently over the past 30 years (an increase in supply), the additional wages needed for an employer to hire a college-educated worker have also grown (an increase in price). Absent a shift in demand, increases in supply should drive down prices, so a price increase implies that demand has shifted toward skilled workers as well.

Do improvements in the way goods and services are produced necessarily lead to greater disparity in pay? If changes in technology have increased disparity, does that mean that technological change is always bad for those who are in the lower portion of the wage distribution? There are two reasons to doubt that this is true. First, economists studying earlier periods have found that wage disparity actually narrowed in the first half of the 20th century, providing evidence that, in some periods, change has favored less skilled workers as opposed to skilled workers.

A second and more fundamental reason that productivity growth does not leave a whole class of workers behind in the long run is that if changes in technology raise the pay of relatively skilled workers, they also increase people's incentives to invest in acquiring skills. Many of the factors that increase an individual worker's productivity depend on the worker's decisions to invest in

developing new skills. When the rewards to gaining skills increase, workers have increased incentive to acquire additional skills. For example, over the past 30 years, there has been a substantial widening in the difference between pay for workers with a bachelor's degree and pay for those with only a high school diploma. For men, this difference grew from 50 percent in 1975 to 87 percent in 2004.

If this widening in pay differences represents an increase in the amount a worker gains by getting a college education, then it gives individuals a greater incentive to make such an investment in education. Over the last 10 years, there has been an increase in the percentage of people who choose to go to college rather than enter the workforce directly out of high school. In 1992, the size of the workforce with some college education was roughly the same as the size of the workforce with a high school diploma or less. By 2006, the workforce with at least some college had become 50 percent larger than the workforce with no college. Other levels of education, such as master's and doctoral degrees, have shown similar increases in the rewards for obtaining such a degree and in the number of people choosing to make that investment. From 1987 to 2003, wages for those with an advanced degree increased faster than for those of any other education group, and since the mid-1990s, the share of people age 30–39 with an advanced degree has increased by 38 percent. Thus increased demand for skilled workers has been followed by an increase in supply, which raises the average skill level in the economy and leads to higher average productivity.

Understanding the Acceleration in U.S. Productivity: Industry Analysis

Understanding why productivity growth in the United States has increased requires knowing what factors in the economy have changed. Chart 2-1 demonstrated that most of the recent increase came about through greater capital deepening and efficiency gains. What the chart did not tell us is why businesses increased their rates of capital investment to bring about capital deepening and why efficiency gains have been higher in the past decade than they were for much of the previous two decades.

Productivity growth for the economy as a whole comes from investment and innovation in a wide variety of businesses. A lot can be learned about the sources of growth by looking at which kinds of investments have grown most quickly, as well as which industries have had the fastest productivity growth. The average rate of productivity growth hides substantial differences across industries. In particular, the surge in productivity in the late 1990s appears to be a story of growth in industries making and using IT capital. Chart 2-3

illustrates that efficiency growth since 2000 has been particularly strong in the high-tech sector, but that it has also been strong in the distribution sector, which includes retail and wholesale trade, transportation, and warehousing. Finance and business services also showed strong efficiency growth and hence strong productivity growth. Manufacturing, which has made small investments in IT capital compared to the other sectors shown, has had the slowest recent growth in efficiency.

The strong productivity growth in the distribution and financial services sectors highlights one of the most striking differences between the pre- and post-1995 periods. From the 1970s through 1995, productivity growth in goods-producing industries was generally greater than that in service-providing industries. However, since 1995, productivity growth in service-providing industries has exceeded the growth in goods-producing industries (such as manufacturing).

Given this difference, one of the most important insights into the recent period of productivity growth comes from understanding why service-sector productivity growth accelerated after a long period of slow growth. As discussed above, capital deepening and efficiency growth accounted for most of the acceleration of productivity growth for the U.S. economy as a whole over the last decade.

Chart 2-3 **Efficiency Growth Highest in Sectors That Made Large IT Investments in 1990s**
The finance and business and the distribution sector made large IT investments in the 1990s and had large efficiency gains in 2000–2004.

Annual average percent change, 2000–2004
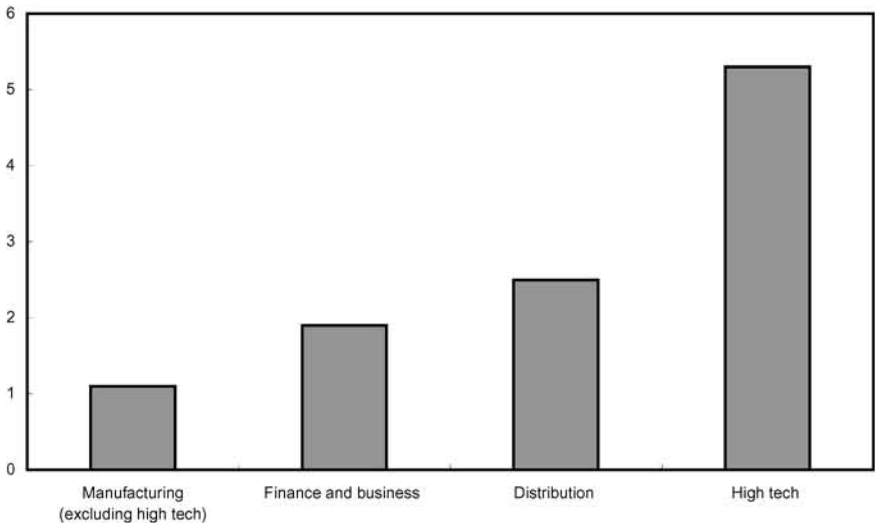
Source: Corrado et al, "*Modeling Aggregate Productivity at a Disaggregate Level: New Results for U.S. Sectors and Industries*," Federal Reserve Board, July, 2006.

In examining productivity growth rates over the recent period, researchers have found it useful to characterize investments by whether they involve a purchase of IT equipment, which is usually defined as computer hardware, software, and telecommunications equipment. Box 2-1 discusses some of the potential mechanisms, such as intangible capital accumulation, through which IT capital leads to productivity improvements.

Box 2-1: Intangible Capital and IT Investment

While information technology clearly accounts for a sizable share of productivity growth since 2000, the mechanisms through which it induced this growth are not as clear. The assumption has been that since efficiency growth has been the largest contributor to productivity in this recent period, IT gains are embedded in this growing efficiency. However, hidden within these increases in efficiency may also be capital growth not captured in official measures.

Standard measures of capital primarily count physical capital, but businesses expend resources on many other activities that aim to increase the value of future output. Some examples are research and development spending, revamping a business's organization, advertising aimed at improving consumers' perceptions of a business's brand, or developing a secret recipe. These kinds of activities are often called intangible investment because they build up assets that are valuable to firms but are not easily measured.

Conceptually, these activities qualify as capital investment, but they are not currently included in official capital measures because they are hard to measure. Why does this matter when discussing productivity? Expanding the definition of capital by including intangibles would change the shares of the factors contributing to labor productivity growth, increasing the share attributed to capital deepening and reducing the share attributed to efficiency gains. This shift would not only call into question the finding that IT investment contributed to productivity mainly through efficiency gains, but would also help explain why productivity did not accelerate with early waves of IT investments. Indeed, it is consistent with the hypothesis that for businesses to take full advantage of their IT investments, they needed to develop innovative business practices. Only when they made intangible investments to complement their IT investments did productivity growth really take off.

The industries that produce IT equipment had particularly rapid efficiency growth, resulting in falling prices accompanied by rapid increases in the speed and power of IT equipment. These industries directly brought up the average rate of productivity growth for the economy, but their advances also had significant indirect effects by driving a surge in IT equipment investment in other industries. The increase in capital deepening in the 1990s was led by large investments in IT equipment, but productivity gains from these investments did not immediately emerge.

In the 1995 to 2000 period, industries with above-average investment in IT equipment had significantly larger increases in their productivity growth rates than did other industries. For example, the retail trade and financial services industries had much higher productivity growth over the 1995 to 2000 period than in the preceding period, and had well-above-average investment in IT equipment. Box 2-2 indicates that much of the retail trade productivity gains occurred because of supply chain improvements made possible by information technology. Research estimating the contribution of IT-related forces—including both productivity growth in IT-producing industries and the share of productivity growth accounted for by IT investment in other industries—shows that information technology accounted for more than half of productivity growth from 1995 to 2000.

Box 2-2: Information Technology, the Supply Chain, and Productivity Growth in Retail Trade

The retail trade sector shows how IT investment, innovation, competition, and flexible markets interact to affect productivity growth. Retailers have made heavy investments in information technology and have had rapid productivity growth, but changes in the way that retailers use information technology—both in their stores and with their suppliers—were necessary to generate this surge in productivity growth. The focus here is on two types of innovations: changes in the organization of the supply chain of consumer goods and changes in the way retailers organize store operations.

Manufacturers and retailers of consumer goods have increased their use of electronic data interchange, allowing manufacturers to help retailers manage inventories and avoid stockpiling and shortfalls. Electronic data interchange also allows for automatic ordering, billing, and payment. Retailers benefit from lower costs of carrying inventory and reduced resources spent managing it, and manufacturers benefit

Box 2-2 — *continued*

from being able to smooth out production. Because these changes have enabled retailers to more reliably stock a wide variety of goods, consumers have benefited from increased product variety. Making these changes required an investment in IT equipment by manufacturers and retailers, and required them to change the way they exchanged information and interacted.

Large retailers also made internal changes that significantly increased productivity. One change was an increase in the scale of stores. Other important changes involved the use of information technology and improved management practices. Examples include an increased use of software to manage the flow of goods and staffing levels in stores, and more cross-training of employees to make better use of store labor. Rapid expansion of the largest firm put competitive pressure on other retailers, leading them to cut costs and, in many cases, to emulate the process improvements introduced by the industry leader.

Why Has Productivity Growth Accelerated in the U.S. While Slowing in Other Countries?

The United States has experienced the fastest acceleration of productivity growth among major industrialized countries since the early 1990s. Chart 2-4 shows that, after lagging behind most of the countries in the G7 between 1990 and 1995, the United States has been the country with the fastest growth in GDP per hour worked in the G7 between 2000 and 2005. Only the United States and Japan had faster productivity growth in the most recent period than they did in the early 1990s, and only the United States has shown consistent acceleration over this time period.

Since all of these countries have, in principle, approximately the same access to information and global markets, why have the other major industrialized countries not been able to post productivity gains as large as those in the United States and Japan? The major advances in this period appear to have come from opportunities that developed from the rapid advancement in information technology. While all developed countries had access to IT capital, the existing economic environment in the United States put it in position to quickly make the most of these opportunities. International openness to investment and trade combined with highly flexible and lightly regulated markets and an environment that fosters innovation appear to be at least part of the answer.

Chart 2-4 Average Annual Productivity Growth Has Fallen for Most G7 Nations Since 1990
Only the United States has shown consistent increase in productivity growth over this period.

Annual percent change
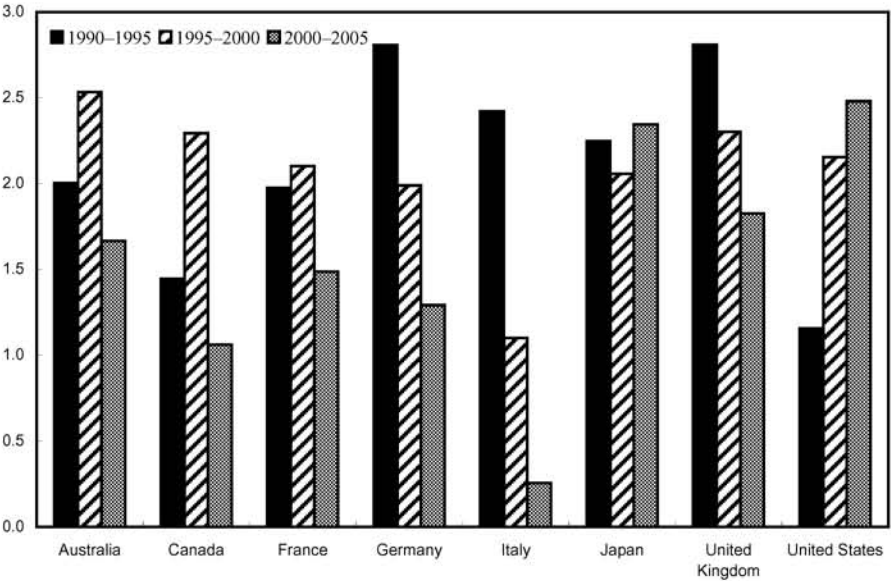
Source: Organization for Economic Cooperation and Development.

International Openness

As discussed earlier, capital deepening has played a significant role in U.S. productivity growth. Over the past 10 years, the United States is second only to Canada in its annual growth rate of real private investment. Real investment in the United States over this period increased at an annual rate of 5.1 percent, nearly double the average rate of the other G7 countries (excluding Canada). The United States has been able to accomplish this level of investment because of its open and transparent investment environment.

While capital deepening played an important role in the productivity gains experienced in the late 1990s, so did advances in information technology. To benefit from the IT boom, firms had to invest large amounts in computers, software, and employee training. From 1995 through 1999, U.S. investment in information-processing equipment and software increased at an average rate of around 20 percent per year, and total investment grew faster than in any other country in the G7. To help fund these investments, the United States received substantial flows of financial capital from abroad during this period. While the United States might have invested in IT capital without access to international financial markets, and while Europe may not have invested more even if it was more open to international capital flows, it is

almost certain that the United States was able to use its open investment environment to finance the increase in IT capital.

Access to international financial markets tends to lower borrowing costs and enable a country to increase capital investment rates without increasing domestic savings. This outcome would not be possible if businesses had access only to domestic financing.

International openness has also contributed in other ways to recent efficiency gains in the United States. Since the early 1990s, the United States has increased its openness to international trade. From the North American Free Trade Agreement (NAFTA) (signed into law in 1993) to the Trade Act of 2002 and the renewal of Trade Promotion Authority in the same year, the United States has worked to break down trade barriers. Lower trade barriers have in turn increased the level of international competition in product markets. Some U.S. companies have suffered from the increased competition; some have benefited. The increased competition forces firms to seek new ways of doing business to remain competitive, and because of this, international trade may contribute to growth in innovation.

Flexible Labor Markets

Efficiency gains resulting from more flexible and competitive labor markets have been another important reason why the United States was able to benefit from recent shifts in technology. The United States ranks first among G7 countries in the World Bank's Rigidity of Employment Index, indicating very flexible labor markets relative to other G7 countries. Japan, for example, ranks fourth among G7 countries, while France ranks last. The index averages measures of the difficulty of hiring a new worker, restrictions on expanding or contracting the number of working hours, and the difficulty and expense of dismissing a worker. While other countries are tied with the United States on the latter two measures, the United States owes its first place rank to the ease with which American employers can hire new employees.

Flexible labor markets allow workers to flow to high-productivity and high-wage industries. Hiring and severance costs tend to increase unemployment by making firms reluctant to hire new workers. They encourage labor hoarding, a practice in which firms hold on to workers not currently needed for production in order to avoid the costs of hiring new workers when the firm's workforce needs to expand. Labor hoarding lowers the level of productivity and reduces the average growth rate of productivity, as firms find it more difficult to respond to innovations and shifts in demand.

Flexible labor markets improve productivity growth because they allow firms to more easily adjust the size and scope of their operations in response to economic developments. For example, after an increase in efficiency, a firm may become more competitive and decide to expand output and so need to

hire more workers. The firm may also wish to change the mix of workers it employs. Flexible labor markets allow these transitions to occur at a low cost.

Low Costs of Starting a Business

Low costs of business entry with relatively few administrative hurdles have also contributed to greater efficiency gains in the United States. A recent study by the World Bank shows that the United States, at 5 days, ranks behind only Canada and Australia in terms of the number of days required to start a business, and has the fourth lowest administrative costs to start a new business. New businesses provide both a ready supply of new ideas and a source of competition that forces larger businesses to innovate. Both of these factors have likely given the United States an edge in taking advantage of new opportunities made possible by IT advances. As with flexible labor markets, the ease of starting a new business helps with the level and the growth rate of productivity. Over long periods of time, starting new businesses keeps the economic environment competitive, which spurs innovation and helps push inefficient firms out of the market place.

Policy Implications

What can the United States do to promote further productivity growth? First, the most important way to encourage capital deepening is to maintain the smallest possible difference between the before-tax and the after-tax rates of return to investments. Capital deepening makes workers more productive and leads to higher wages in the long run. Making the tax cuts on capital gains and dividends permanent would help in this regard. Chapter 3 of this report discusses policy options affecting the taxation of capital.

Second, policies must encourage investment in skills. One way to do this is to keep the tax rates on wage income low. If individuals see little return to going to college, vocational school, or graduate school because of high tax rates on moderate- to high-wage earners, their incentives to invest in skill will be dampened. Chapter 3 further discusses how tax policy affects investment in skill. Strengthening K-12 education, reducing our dropout rates, and ensuring that all children receive high-quality education will increase the skills of our workforce and better prepare our citizens for further skill investment as adults. The President's efforts over the past several years to improve education and training with the No Child Left Behind Act, community college initiatives, and job training reforms will help. Furthermore, because learning begets learning, the returns should continue into the distant future.

Third, we must remain open to foreign investment. Openness to foreign capital has given the United States the flexibility it needs to deepen its capital

stock and improve its productivity without necessitating a corresponding increase in domestic savings. To maintain current growth rates we must keep pushing for freer trade, especially in the area of services, which has become a significant part of our economy. Chapter 8 of this report discusses policies to increase our international openness.

Fourth, we must encourage innovation and entrepreneurship. The President has outlined a competitiveness initiative that increases public investment in basic research—an important complement to private sector innovation—and strengthens math and science education to provide the skills needed for technological innovation.

Conclusion

Maintaining a solid productivity growth rate is of great importance to maintaining and increasing U.S. standards of living. The surge in productivity growth since about 1995 has come from heavy business investment in information technology, accompanied by large efficiency gains from innovation and competition. The United States has gained more from rapid advances in information technology than the other major industrialized countries because its culture of entrepreneurship and its flexible markets for products, capital, and labor have allowed American businesses to make the most of these changes.

Pro-Growth Tax Policy

The word "investment" has different meanings to different people. In finance, investment means the purchase of financial products or other assets, such as mutual funds or gold, with an expectation of favorable future returns. For businesses, it can mean the purchase of a physical good, such as a durable machine or inventory, with the hope of improving future business. In economics, investment is defined as any use of resources intended to increase future production output or income. In particular, capital investment is money spent on physical capital such as buildings, equipment, or machinery, or on human capital such as education or job training. Because a larger capital stock makes labor more productive, investment is a primary driver of greater economic growth and higher standards of living.

If governments pursue policies that involve the least amount of government interference necessary for a well-functioning capital investment market, this will encourage an efficient amount of investment. One type of policy that is key to encouraging an efficient level of investment is *pro-growth tax policy*. One of the goals of pro-growth tax policy is to finance government services in a way that minimizes the effect of taxes on the capital investment decisions of households or businesses. By taxing investment returns too heavily or by providing tax advantages to certain types of investment, a tax system can discourage overall investment, as well as prevent capital from being used efficiently. A tax system that affects investment decisions in these ways is called "distortionary" because it creates incentives for people to base their saving and investment decisions on taxes, rather than making those decisions based solely on where they can use their resources most productively.

This chapter discusses the advantages of adopting a more pro-growth tax system. It reviews recent changes that have reduced tax distortions on capital investment decisions, and evaluates options to further reduce such distortions. It draws the following four main conclusions.

- The goal of pro-growth tax policy is to reduce tax distortions that hamper economic growth. Most economists agree that lower taxes on capital income stimulate greater investment, resulting in greater economic growth, greater international competitiveness, and higher standards of living.
- The current tax code contains provisions that discourage investment and create distortions that affect the level, structure, and financing of capital investment. These distortions dampen capital investment and contribute to an inefficient allocation of capital throughout the economy.

- Estimates from research suggest that removing these tax distortions to investment decisions could increase real gross domestic product (GDP) by as much as 8 percent in the long run.
- Since 2001, temporary changes in the tax code have reduced the tax on investment. These pro-growth policies have stimulated short-run investment and economic growth. However, the temporary nature of the provisions eliminates desirable long-run economic stimulus.

Rationale for Pro-Growth Tax Policy

All societies must decide on the amount of government services that best provides for the welfare of the citizenry. When deciding how to finance a given amount of government services, two features of the tax system must be determined—the appropriate tax base and the appropriate tax rate. The goal of pro-growth tax policy is to define a tax base and choose tax rates that finance government expenditures with the least distortionary effect on the economy. A tax system is distortionary when it creates incentives for people to make spending, saving, or investment decisions that are different from the decisions they would make in the absence of taxes. For example, by taxing the sale of apples and not oranges, a tax system would encourage people to consume more oranges and fewer apples than they otherwise would. Similarly, by taxing a family's out-of-pocket health spending but not employer-paid health insurance premiums, the tax system encourages inefficient consumption of health care by households. (See Box 4-1 in Chapter 4, The Fiscal Challenges Facing Medicare, for a discussion of the President's proposal to reform the tax treatment of health insurance.) By comparison, a tax system that taxes investment can create incentives that favor consumption over saving, investment in certain types of capital over others, or certain methods of financing capital investment. In the absence of distortionary taxes, people would have made those decisions based solely on the best and most productive use of those resources.

Defining the Tax Base

Most economists agree that the choice of the appropriate tax base is between taxing some measure of income or taxing some measure of consumption. Broadly defined, *income* is the increase in an individual's ability to consume during a period of time. Income can include labor earnings (both cash and benefits), interest payments, rents, royalties, dividends, increases in asset values, alimony, and pension payments. An important dimension of income taxation is that saving and investment are included in the tax base. Using income as the tax base is equivalent to taxing *potential* consumption. In effect,

this taxes all resources that people *put into* the economy. A tax system with an income base is distortionary because taxes affect decisions on when, how, and how much to save and invest. For example, in taxing household saving, future consumption (financed by saving) becomes relatively more expensive compared to current consumption. As a result, households tend to consume more and save less than they otherwise would if saving were not taxed.

By contrast, consumption is defined as the *actual* amount that people and businesses spend buying goods and services today. When a tax system has a consumption base, it only taxes what people *take out* of the economy. While there are several possible measures of a consumption tax base—retail sales, value-added, and consumed income, among others—all of these measures share the attribute of excluding saving and investment from the tax base. Such a tax system is considered "neutral" and efficient because it neither encourages nor discourages savings and investment decisions; it allows people to decide whether to consume now or to invest in the future based on market prices instead of on how to avoid paying taxes. Relative to an income tax, the consumption tax base results in a larger, more efficient stock of capital, which in turn makes workers more productive. Output and wages rise, resulting in higher standards of living. As a result, many economists feel that consumption is a better base for pro-growth tax policy.

Our current tax system has a hybrid tax base, with elements of both income and consumption tax bases. Some, but not all, of the return to saving and investment is excluded from the tax base through various provisions. For example, individual retirement accounts (IRAs), employer-sponsored retirement savings plans, lower tax rates on capital gains and dividends, and accelerated depreciation for certain types of investment are some of the provisions in the current tax code that provide at least a partial consumption tax base. Recent estimates suggest that about 65 percent of the return to household financial assets is taxed under an income tax base, with the remainder receiving consumption tax treatment.

Choosing the Tax Rates

A marginal tax rate tells how much tax is paid on an additional, or *marginal*, dollar of income. When assessing the effect of marginal tax rates on investment, it is the effective tax rate rather than the statutory tax rate that matters. A *statutory marginal tax rate* is a legal definition of the amount of extra income needed to pay taxes due from an additional dollar of taxable income in any year. By contrast, an *effective marginal tax rate* estimates the extra share of the total return from an investment needed to cover tax liabilities over an investment's useful life. A tax system with high effective tax rates on labor and capital income will dampen economic growth by reducing incentives to work and invest in capital formation.

Pro-growth tax policy, whether through adopting a consumption base, lowering statutory tax rates on saving and investment, or allowing individuals to fully deduct the cost of investment from taxable income, stimulates new investment by lowering the effective tax rate on investment income. Individuals and businesses will undertake more projects because lowering the effective marginal tax rate reduces the pretax rate of return necessary to make new projects profitable. In addition, lowering the effective tax rate on the return to capital investment enhances the competitive position of the United States in today's increasingly global economy. This is because a lower effective tax rate raises the after-tax return to U.S.-based investment relative to foreign investment, making U.S. investment relatively more attractive to both domestic and foreign investors.

The U.S. Tax System— Previous Distortions and Recent Reforms

The United States tax system has become increasingly distortionary and inefficient, with hundreds of highly targeted tax provisions that erode the potential for tax system neutrality and greater economic growth. A major source of inefficiency is the treatment of capital investment, both for physical capital and for human capital. The profusion of provisions has resulted in a system where taxes can be the primary determinant in whether to undertake new investment, what form the investment should take, and how to finance the investment.

Since 2001, several pro-growth tax policy changes have been enacted which have reduced the distortionary effect of taxes on investment decisions. This section discusses investment distortions in the tax system prior to 2001 and analyzes how changes since that time have reduced distortions and stimulated economic growth. Overall, the pro-growth policies enacted since 2001 have helped lessen the impact of the recession and have led to greater investment and overall economic growth.

Tax Treatment of Physical Capital Investment

This section discusses how two features of the tax system result in "tax wedge" distortions that contribute to physical capital investment inefficiency: depreciation schedules that result in an inefficient level and allocation of capital, and the double taxation of corporate profits that affects the level, form, and financing of business investment.

The Tax Wedge

The tax system creates a "tax wedge" for investment, making the pretax return on investment higher than the after-tax return on investment. This is important because investors require the pretax return to cover both the opportunity cost and the tax cost of investment. If the tax wedge is large, fewer projects will be undertaken because the after-tax return for some projects will be below the opportunity cost of investment. For example, consider an investment with a pretax return of 10 percent and an after-tax return of 7.5 percent, meaning the tax wedge is equal to 25 percent of the pre-tax return. If investors decide they require an 8 percent after-tax return in order to cover the opportunity cost of the investment, taxes will stop the otherwise profitable project from being undertaken. By lowering the effective tax rate on investment, the pretax return is unaffected but the after-tax return will rise. For example, if the effective tax rate is reduced to zero, then the tax wedge is eliminated and the after-tax return rises to 10 percent. Note that the tax wedge does not need to be eliminated for our hypothetical project to be financed—the effective tax rate only needs to be reduced to the point where the after-tax return is 8 percent. However, completely eliminating the tax wedge removes taxes from the investment decision. Two main contributors to the tax wedge on investment returns are depreciation schedules and the double tax on corporate profits.

Depreciation Schedules

A primary source of the inefficiency created by the tax wedge is the depreciation schedules that treat investments very differently depending on their business sector, asset life, and source of financing. *Depreciation schedules* tell how much of an investment's acquisition cost can be deducted from the taxpayer's taxable investment income in any year. There are two distortions associated with the tax depreciation system. First, spreading the deduction for the acquisition cost over a number of years lowers the present value of the total tax deduction relative to fully deducting the cost in the year purchased. By lowering the present value of the deduction, the depreciation system raises the tax cost and the total effective cost of investment. This makes some projects unprofitable and reduces the economy-wide level of investment. Second, the depreciation system distorts the allocation of investment among various sectors of the economy because the depreciation schedules lead to sectoral differences in effective marginal tax rates. Under an income tax system, the amount of investment cost counted each year should ideally equal the true economic depreciation of the asset. For example, if an asset loses 10 percent of its useful value per year, then an ideal income tax depreciation schedule would allow 10 percent of the cost to be excluded from income each year. When tax

depreciation is not the same as economic depreciation, the tax system distorts investment decisions regarding the allocation of capital investment.

A common method of measuring the relative distortions caused by the depreciation system is to calculate the effective marginal tax rates on different types of investment. Under current law, different types of investments are depreciated under various depreciation schedules ranging from 3 to 39 years. Because acquisition costs are deducted from taxable income at different rates, the amount of tax paid—and the effective marginal tax rate—varies by depreciation class. Table 3-1 shows the effective tax rates on different assets for different types of investments, with computer investment facing the highest effective marginal tax rate and petroleum infrastructure investment facing the lowest. Because marginal investments should provide the same after-tax rate of return, the depreciation schedule distorts the allocation of capital by discouraging investment in assets with high effective marginal tax rates.

Even if we adopted a tax system with tax depreciation equal to economic depreciation, there would still be a notable tax wedge that would distort investment decisions. To completely remove the investment distortions of depreciation schedules would require adopting a consumption tax base. With a consumption tax, all investment costs are fully deducted (fully expensed) from taxable income in the period in which the acquisition occurs. This has the effect of reducing the tax wedge to zero if there are no other taxes on investment returns. This means that the tax system is neutral to the level and allocation of capital investment because taxes do not affect the decision to invest and all types of investment are treated equally.

The Double Tax on Corporation Profits

The double tax on corporate profits—which is inconsistent with either an income tax or a consumption tax—also has a pronounced effect on investment

TABLE 3-1.— *Effective Marginal Tax Rates on Capital Income of Corporations by Asset Type*

Asset type	Effective marginal tax rate (%)
Computers and peripheral equipment	36.9
Inventories	34.4
Land	31.0
Automobiles	29.7
Educational buildings	28.4
Residential buildings	23.8
Medical equipment and instruments	20.4
Light trucks (incl. utility vehicles)	18.2
Household appliances	17.5
Aircraft	14.5
Railroad equipment	11.4
Petroleum and natural gas structures	9.2

Source: Congressional Budget Office.

decisions. First, corporations pay tax on net corporate earnings at a maximum marginal rate up to 35 percent. Second, individual investors are taxed on the returns they earn on corporate equity. These returns can take the form of a *capital gain*, the difference between the purchase price and the sale price of an asset, or a *dividend*, which is a share of corporate profits distributed to shareholders after corporate income tax has been paid.

The total tax on corporate income is calculated by combining these two layers of tax. Prior to 2001, the tax on individual investment returns (capital gains and dividends) created incentives for investors to favor projects that paid returns in the form of capital gains or interest payments instead of dividends because long-term capital gains were taxed at a maximum statutory rate of 20 percent, while dividend payments were subject to a maximum individual statutory rate of 39.6 percent (both tax rates do not take state and local taxes into account).

For corporate income distributed to shareholders as dividends, the double tax on corporate profits could approach the level of confiscation. For example, given a maximum statutory marginal tax rate of 35 percent for corporations and 39.6 percent for individuals, the combined effective marginal tax rate on distributed corporate profits could have been as high as 61 percent! Instead of paying out corporate profits as dividends, a corporation could retain and reinvest the after-tax profit, leading to an increase in its stock value. Prior to 2001, when a long-term capital gain was realized, the combined effective tax rate on corporate profits was about 42 percent, after accounting for the deferral of tax on the accrued gains. All else equal, investors tended to favor investment returns in the form of capital gains.

The high effective tax rate on equity-financed investment also created incentives that favored debt (taking out loans or issuing bonds) when financing new projects. As shown in Chart 3-1, while the economy-wide effective tax rate prior to 2001 was 20.4 percent, the effective tax rate on business sector investment was 29.8 percent. Chart 3-2 shows that the effective tax rate on equity-financed investment was 45.2 percent and the effective tax rate on debt-financed investment was almost zero. The reason for this large difference in effective rates is that corporations can deduct interest payments for loan and bond payments from taxable income, but must include dividend payments and retained earnings in taxable income. Individual investors then must pay taxes on the interest payments from their debt holdings and the investment returns (capital gains and dividends) from their equity holdings. This tax treatment results in a system where the return to corporate debt is taxed once but the return to corporate equity is taxed twice. The resulting overreliance on debt-financed investment could lead to greater bankruptcy risk during temporary industry or economy-wide downturns, as well as to a misallocation of resources in the economy.

Chart 3-1 Effective Marginal Tax Rates on New Investment

New investment can face highly disparate tax treatment depending on the sector.

Percent
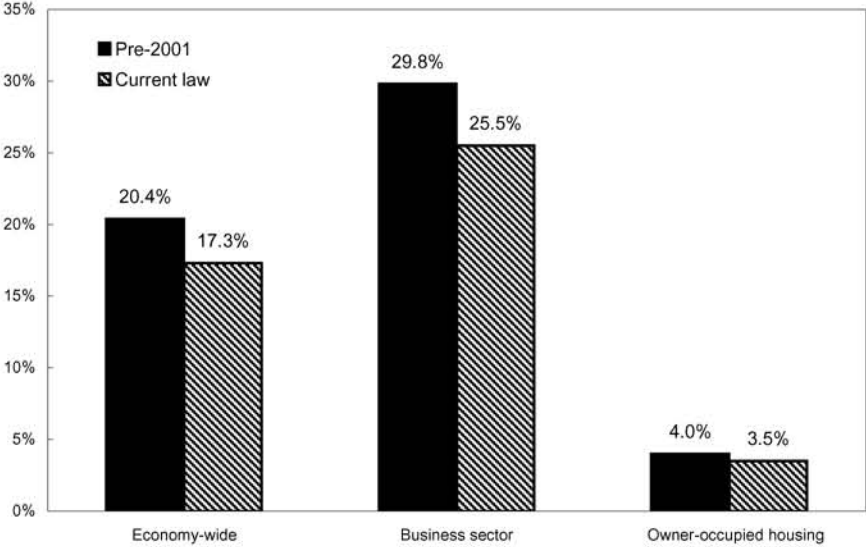
Source: Department of the Treasury (Office of Tax Analysis).

Chart 3-2 Effective Marginal Tax Rates on New Business Investment

Tax treatment of new investment in the business sector varies by type of financing.

Percent
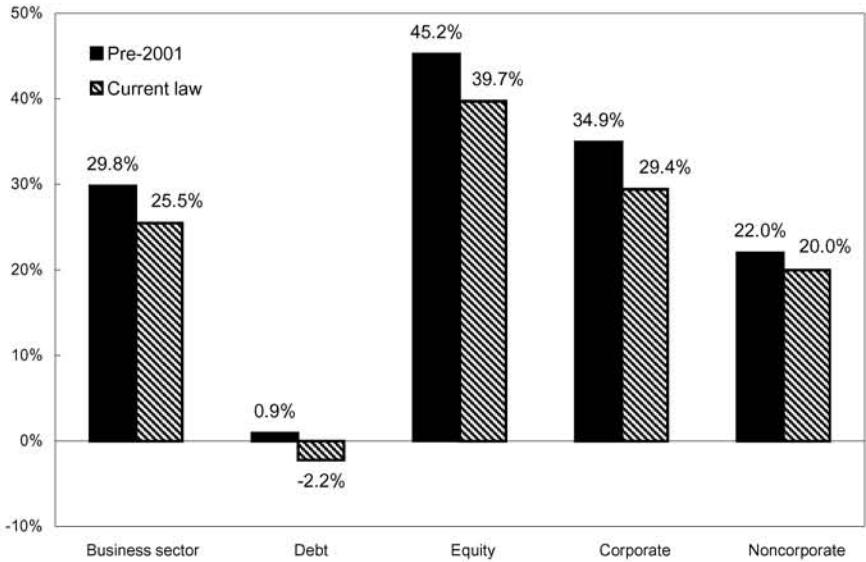
Source: Department of the Treasury (Office of Tax Analysis).

Tax Treatment of Human Capital Investment

Human capital investment (such as education and worker training) is an important input in the production of final goods and services, and investing in human capital is a cost of earning income. Prior to 2001, the tax treatment of education and training expenses was mixed. Some costs were fully deducted against taxable income, while others were subject to varying degrees of taxation. In addition, the treatment varied depending on whether the investment was paid for by businesses or households.

At the household level, most human capital investment was fully deducted because the tax system does not tax the opportunity cost of education—the foregone wages of working instead of attending school. For other human capital investment costs, there was a complicated set of rules, with the tax treatment primarily determined by the income of the individual taxpayer undertaking the investment. Some costs could also be deducted under both income and payroll (Social Security and Medicare) taxes.

The opportunity cost of working was fully deductible under both the income and payroll tax. Other costs fully deductible under both taxes were scholarships, fellowships, and reduced tuition. Costs that were fully deductible under just the income tax included education costs paid through Coverdell Education Savings Accounts (Coverdell ESAs), interest payments on student loans, and Treasury bond interest. These costs were excluded from income tax so long as they were used for tuition and related expenses such as fees, books, supplies, and the equipment required for courses of instruction.

At the firm level, human capital investment received more efficient tax treatment than physical capital investment. Consider a $50,000 investment in office equipment. For many businesses, this cost was not fully deductible. Instead, the cost was recovered through depreciation provisions, with a fraction of the cost deducted from taxable income over a 7-year period. Alternatively, the firm and workers could have agreed to reduce cash compensation by $50,000 and invest the money in job training. In this case, the firm would have deducted the cost of training from taxable income as an ordinary business expense and workers would not have claimed the cost as taxable income for income or payroll taxes. In this way, the investment cost was fully deductible in the year the training occurred, resulting in no tax distortions to the firm's human capital investment decision.

In addition to allowing partial deductibility of human capital investment, the tax system had two human capital investment tax credits available for use by households. In 2000, the Hope credit provided a tax credit of up to $1,500 per eligible student for the first 2 years of post-secondary education. To qualify for this credit the student had to be pursuing a degree or other recognized educational credential. The Lifetime Learning credit provided a tax credit of 20 percent of the first $5,000 in household education expenses per

year. This credit was available for any post-secondary education investment for an unlimited number of years, regardless of whether the student was pursuing a degree or educational credential.

Tax credits differ somewhat from tax deductions. A *tax credit* directly reduces the amount of tax you have to pay. By contrast, *tax deductions* reduce the amount of income subject to tax. Tax credits can provide investment incentives that are equivalent to partial or full deductions and can also be more generous than full deductions. For example, consider a person who has qualified education expenses of $5,000 and receives a $1,000 Lifetime Learning credit. If this person is paying taxes at a 20 percent effective marginal tax rate, then the credit is equivalent to being able to fully deduct the education cost from taxable income. If the person is paying taxes at a higher marginal tax rate, then the credit is equivalent to a partial deduction. For example, if the student is paying tax at a 31 percent marginal tax rate, then the credit is equivalent to being able to deduct about $3,200 of the investment cost. Similarly, if the student is paying tax at less than 20 percent, then the credit provides more than a full deduction (i.e., a tax subsidy).

Overall, the tax system in place prior to 2001 can be characterized as relatively inefficient with respect to investment in physical and human capital. Changes to this system were and are still necessary to eliminate distortions that keep the economy from reaching its full potential.

Pro-Growth Changes Since 2001

A number of pro-growth tax initiatives have been proposed and signed into law by President Bush since 2001. The initiatives enacted include provisions aimed at reducing the double taxation of corporate profits by lowering the tax rate on dividends and capital gains; temporary bonus depreciation; expansion of deductibility of higher education costs; and several smaller provisions aimed at encouraging investment. Taken together, these reforms reduced the effect of taxes on investment decisions.

Reducing the Double Tax on Corporate Profits

The Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA), proposed and signed by President Bush, reduced the double tax on corporate profits by lowering the top individual tax rate on dividends and capital gains to 15 percent through 2008. These changes promoted economic growth by increasing capital in the corporate sector and improving the allocation of capital throughout the economy. As shown in Chart 3-3, in the 9 quarters preceding JGTRRA, real private nonresidential investment fell at an average annual rate of about 7.5 percent and annual real GDP growth averaged 1.1 percent. In the 13 quarters after JGTRRA was enacted, real private nonresidential investment grew at an average annual rate of about

6.9 percent, with annual real GDP growth averaging 3.6 percent. While it is too early to estimate the full effect of pro-growth tax policy on GDP, recent estimates suggest that without the tax cuts the economy would have had as many as 3 million fewer jobs and real GDP would have been as much as 3.5 to 4 percent lower by the end of 2004.

Several studies indicate that prior to JGTRRA, corporations had been steadily reducing dividend payments. The reason is that the tax system resulted in a strong tax bias in favor of retained earnings and capital gains. Since passage of JGTRRA, there has been an increase both in the average amount of corporate dividend payments (Chart 3-4) and in the percent of firms paying dividends (Chart 3-5). Reducing the double tax on corporate profits also slightly reduced tax-motivated incentives for debt finance because it reduced the effective marginal tax rate on equity finance. As seen in Chart 3-2, the effective marginal tax rate on equity-financed corporate investment is now about 40 percent, a drop of about 12 percent from the pre-2001 effective tax rate. While this rate is still substantially higher than the effective tax rate on debt-financed corporate investment, the relative reduction reduced the distortion between debt and equity finance.

A major challenge facing this pro-growth change is the impermanence of the capital gains and dividend tax reductions. Originally scheduled to expire at the end of 2008, both provisions were recently extended until the end of 2010 in the Tax Increase Prevention and Reconciliation Act of 2005 (TIPRA). For

Chart 3-3 **Real Private Nonresidential Fixed Investment**
In the 13 quarters after the President signed JGTRRA into law, real private nonresidential investment grew at an average annual rate of 6.9 percent.

1-quarter percent change (annualized)
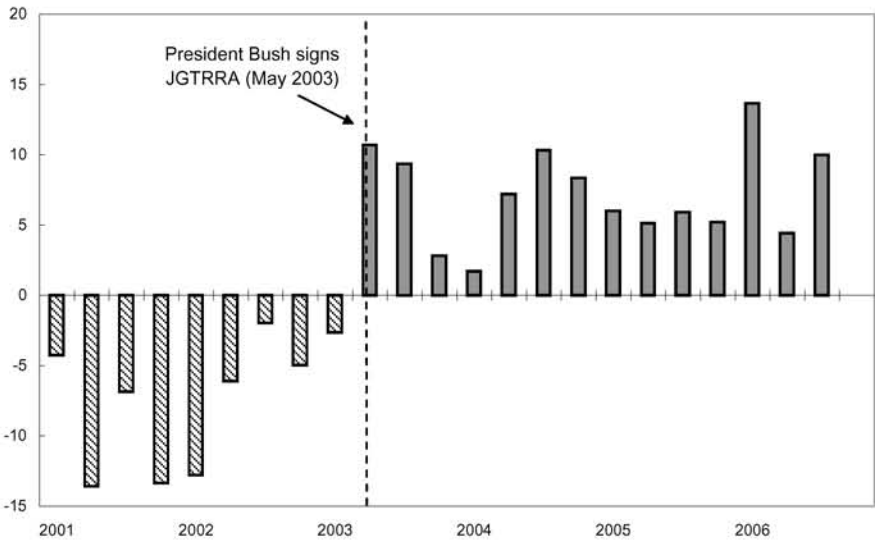
Source: Department of Commerce (Bureau of Economic Analysis).

these changes to have lasting effects on investment and economic growth, these pro-growth policies should be made permanent.

Chart 3-4 **Dividends per Share**
Since passage of JGTRRA, there has been an increase in the average amount of dividend payments.

4-quarter percent change

President Bush signs
JGTRRA (May 2003)
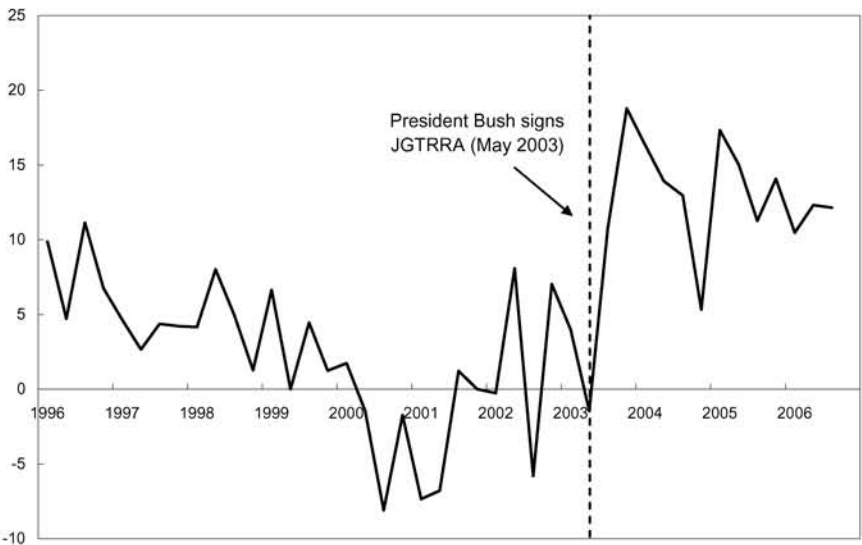
Source: Standard & Poor's.

Chart 3-5 **Percent of Firms in the S&P 500 Paying Dividends**
The percent of firms paying dividends has increased following passage of JGTRRA.

Percent

President Bush signs
JGTRRA (May 2003)
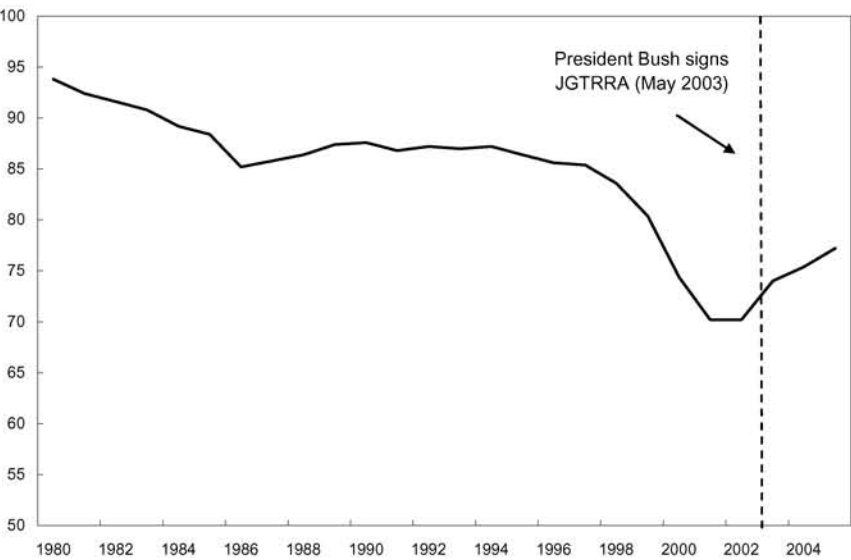
Source: Standard & Poor's.

Increasing the Deductibility of Capital Investment

Another pro-growth change proposed and signed into law by President Bush was the Job Creation and Worker Assistance Act of 2002 (JCWAA). This act included a provision for *temporary bonus depreciation*, which allowed taxpayers an additional first-year depreciation deduction of 30 percent from taxable income. In 2003, JGTRRA included a modification to the JCWAA bonus depreciation provision, allowing taxpayers to take a first-year depreciation deduction of 50 percent from taxable income. Both provisions were temporary and expired at the end of 2004 because the purpose of these provisions was to provide a temporary investment stimulus to speed economic recovery and promote short-term economic growth. By allowing investors to deduct more of the cost of investment from taxable income in the year of acquisition, these provisions had the effect of lowering by one-half or more the effective marginal tax rate on qualifying investment.

Removing Distortions to Human Capital Investment

President Bush proposed and signed into law a number of provisions that reduced tax distortions affecting human capital investment decisions. Among these provisions were statutory changes that allow households to deduct (within limits) higher education costs; an expansion of the deductibility of student loan interest payments; and an expansion of the full deductibility of employer-provided education expenses to include workers pursuing graduate school education. Other changes include an increase in the amount of money a household can contribute to a Coverdell ESA; the removal of tax considerations from higher education costs paid through qualified tuition programs (Section 529 plans); an increase in the amount of costs eligible for the Lifetime Learning credit; and an expansion of eligibility for these various education provisions.

Other Changes

Other changes that have been signed into law by President Bush over the past 5 years are tax credits aimed at encouraging research investment; an expansion of full deductibility of the acquisition cost of tangible property for small business (called Section 179 expensing); full deductibility of brownfields projects; and full deductibility of certain oil exploration costs. Some of these changes stimulated investment and greater short-run economic growth. Unfortunately, the temporary nature of many of these provisions reduces their potential to stimulate long-run efficiency gains to investment and economic growth.

Incremental Approaches to a More Pro-Growth Tax System

Many economists agree that adopting a broad-based consumption tax would benefit the economy. There is a substantial body of research that estimates the economy-wide growth effects of this broad pro-growth tax reform. The estimated effects can vary widely depending on the type of model used and the policy change considered. For example, when considering the transition to a pro-growth consumption tax, estimates of the short-run increase in the capital stock range from about 1 percent to about 14 percent, with estimates of the long-run increase in the capital stock ranging from about 0 percent to about 32 percent. As a result of *capital deepening* (the increase in capital per worker), the long-run increase in real gross domestic product is estimated in the range of about 2 percent to about 8 percent (about $260 billion to about $1.1 trillion in 2006 GDP).

In the absence of such broad reform and the transition to a consumption tax base, there are two primary alternatives for adopting a more pro-growth tax system. One is to allow investors to completely deduct (fully expense) or substantially deduct (partially expense) the cost of their investments in the year in which the investments are made. The other alternative is to lower the statutory tax rate on investment income by reducing or eliminating the tax rate on corporate income, capital gains and dividends, or a mixture of both. Both of these approaches would reduce the amount of tax paid on an investment return, lowering the pretax rate of return necessary to undertake new investment. If one of the objectives of pro-growth tax policy is to move incrementally to a more efficient, consumption-based tax system, then expensing does a better job than rate reductions of meeting this objective. Indeed, full expensing of investment is a necessary component of a consumption tax base. By contrast, reducing the statutory corporate tax rate or eliminating the tax on capital gains and dividends could be accomplished under the existing hybrid tax system.

There are a number of reform options that contain elements of these approaches. One option is a value-added tax (VAT) that replaces all or part of the corporate income tax; another, the Growth and Income Tax (GIT), proposed by the President's Tax Reform Panel, would lower effective marginal tax rates on new investment. Other options focus on household saving as a means to remove investment distortions. However, compared to a VAT or the GIT, these options would provide relatively less stimulus for domestic growth within a rapidly expanding global market. The reason is that focusing on savings incentives tends to ignore the full effects that capital has on the economy. By reducing taxes on investment, the economy develops more capital, increasing labor productivity and wages. In addition, reducing

effective tax rates on investment attracts more foreign investment because U.S.-based investment would offer relatively higher after-tax rates of return. (See Chapter 8, International Trade and Investment, for a discussion of the benefits to the U.S. of foreign investment.) Expanding savings incentives can provide capital deepening, but it will not encourage greater investment by foreign investors who do not receive the benefits of the reform. This section focuses on pro-growth options that would have the greatest impact on economic growth.

Expensing of Investment

Allowing investors to fully deduct the cost of an investment from taxable income is called *full expensing* of investment. As shown in Box 3-1, in the absence of other taxes, full expensing reduces the tax paid on the normal return to capital investment to zero, completely removing taxes from the investment decision. This happens for two reasons. First, all assets face the same effective tax rate—zero—so that taxes no longer influence the decision about where or in what to invest. This results in a more efficient allocation of capital. Second, with full expensing there is no difference between the pretax and after-tax rates of return to investment. As a result, taxes do not discourage capital formation.

It is important to note that full expensing is equivalent to not taxing the ordinary, normal return (or opportunity cost) of new investment. As shown in Box 3-1, the reason is that full expensing is equivalent to an interest-free loan on the value of foregone tax liability. To see this result, consider the example in Box 3-1. Under the income tax, the firm pays $35 in tax on the cost of the investment, whereas under full expensing the tax liability on the cost of the investment is zero. Assuming that the pretax return of 10 percent equals the normal opportunity cost of funds, the deferral of tax liability is worth $3.50 to the firm, which is exactly equal to the tax on the investment return. Because the opportunity cost of this loan is equal to the normal return to the investment, full expensing of investment costs is equivalent to excluding the normal return portion of capital income from taxation. However, returns in excess of the opportunity cost (called *supra-normal* returns) are still subject to taxation. For our example, if the total return of 10 percent is composed as a normal return of 6 percent and a supra-normal return of 4 percent, then the deferral of tax liability is worth $2.10 to the firm. This is equivalent to the firm paying $1.40 in tax, which is a tax of exactly 35 percent on the $4.00 supra-normal return.

Partial expensing of investment occurs when something less than 100 percent of an asset's purchase price is excluded from taxable income in the year the asset is purchased. Partial expensing reduces, but does not eliminate, the amount of tax paid on the return to capital investment because costs

**Box 3-1: Investment Returns Under Different Tax Systems:
A Numerical Example**

Suppose a firm undertakes an investment in a new machine that costs $100 and that earns a pretax rate of return of 10 percent. Assume that the machine does not depreciate in value and that the firm sells the machine for $110 after 1 year. Under a system with a corporate income tax and no expensing, the after-tax cost of the machine is $100 because the firm receives no deduction from taxable income when it purchases the machine. At the end of the year the firm deducts the cost of the machine from the firm's total income and has a net income of $10. With a corporate tax rate of 35 percent, the firm pays $3.50 (35 percent of $10) in tax to the government. This leaves the firm with $6.50 in after-tax income, and results in an after-tax rate of return of 6.5 percent on its investment of $100. The corporate income tax creates a 3.5 percentage point tax wedge between the pretax rate of return (10 percent) and the after-tax rate of return (6.5 percent) on the investment.

With full expensing, the firm deducts the cost of the machine from taxable income at the time of purchase. This means the firm's after-tax cost of the machine is only $65. As before, the firm then sells the machine at the end of the year for $110. Under full expensing, the entire $110 is included in taxable income because the firm deducted the cost of the machine when it was purchased. This means the firm pays $38.50 (35 percent of $110) in taxes and makes an after-tax profit of $6.50. The firm earns an after-tax rate of return of 10 percent on the $65 investment, which equals the pretax rate of return. Because the firm is not taxed on the investment's return, the result is an effective marginal tax rate of zero.

In contrast, consider what happens when the government lowers the corporate tax rate to 25 percent but allows no expensing. The firm sells the machine at the end of the year for $110 and pays tax of $2.50 (25 percent of $10). As such, the firm's after-tax rate of return is 7.5 percent and the tax wedge between the pretax and after-tax rate of return is 2.5 percentage points. Lowering the corporate tax rate reduces the disincentive to invest but does not eliminate it unless the statutory tax rate is reduced to zero. By comparison, reducing the statutory corporate marginal tax rate to 25 percent would be equivalent, in terms of the effective tax rate, to about 38 percent partial expensing of investment costs.

Income Tax versus Pro-Growth Tax: A Numerical Example

Cost of machine	$100	
Pre-tax rate of return	10%	
Value of asset in 1 year	$110	
Corporate rate tax	35%	

Income tax:

Net taxable income		
= Selling price - Cost of asset	$110 - $100	$10
Taxes owed		
= Corporate tax rate * Profit	35% * $10	$3.50
After-tax return		
= Net income - Taxes owed	$10 - $3.50	$6.50
After-tax rate of return		
= After-tax return / Cost of machine	$6.50 / $100	6.5%
EMTR on investment income*		
= Tax paid / Investment income	$3.50 / $10	**35%**

Pro-growth tax:
Expensing

New cost of machine		
= Old cost of machine * (1 - corp rate)	$100 * (1 - 35%)	$65
Net taxable income		$110
Taxes owed	35% * $110	$38.50
After-tax return	$110 - $38.50 - $65	$6.50
After-tax rate of return	$6.50 / $65	10%
EMTR on investment income	$0 / $10	**0%**

Corporate rate cut (new rate=25%)

Net taxable income	$110 - $100	$10
Taxes owed	25% * $10	$2.50
After-tax return	$10 - $2.50	$7.50
After-tax rate of return	$7.50 / $100	7.5%
EMTR on investment income	$2.50 / $10	**25%**

*Note: EMTR refers to the effective marginal tax rate.

in excess of those expensed are still subject to the tax depreciation schedules, resulting in an inefficient allocation of capital.

There are several advantages to adopting full expensing as part of the current tax system. First, full expensing reduces the tax wedge between the pretax and the after-tax rates of return on investments, resulting in a more efficient level and allocation of capital throughout the economy. Second, if coupled with the repeal of capital gains and dividends taxes, full expensing completely removes taxes from equity-financed investment decisions. Third, full expensing reduces distortions that affect the financing of new investment by reducing incentives to debt-finance investment. Fourth, expensing is an integral part of many major tax reform proposals, such as a transition to a VAT, a consumed income tax, or the GIT. Overall, full expensing greatly simplifies the tax system and is an important step in the transition to a full consumption tax.

There are two important issues that must be resolved when adopting expensing as part of the tax system. The first issue is *transition costs*, which pertain to how the tax system will treat existing capital, called "old capital," at the time of the change. This is important because expensing can place a potentially heavy tax burden on the owners of existing capital. This tax burden arises because of the difference in the treatment of new capital (which can be expensed) and old capital (which does not benefit from expensing). As shown in Box 3-1, the after-tax rate of return on new investment rises with full expensing. The increase makes new investment projects relatively more attractive to investors than purchasing existing capital projects. Consequently, the relative value of the existing capital at the date of the change must fall in order for old capital to earn the same after-tax rate of return as an investment in new capital. The decline in value is equivalent to an unavoidable tax on existing capital and is considered a transition cost of full expensing.

The second issue is the treatment of interest payments under full or partial expensing. If expensing is to result in taxes being neutral in investment decisions, interest payments must be taken out of the tax system. Otherwise expensing could result in negative tax rates and overinvestment in capital. Removing interest from the tax base means that borrowers cannot deduct interest payments from taxable income. Similarly, lenders would not include interest payments in taxable income. The elimination of interest deductibility would help to equalize the tax treatment of different types of financing and would reduce tax distortions in investment decisions. However, excluding financial transactions from taxation could create difficulties for financial services businesses and result in opportunities for *tax arbitrage*—forming or consolidating businesses to take advantage of the difference in tax rates as the basis for profit. The taxation of financial services under a consumption tax is a perennially thorny problem that has yet to admit of an easy solution.

Reducing Statutory Tax Rates

An alternative to expensing of investment is to reduce statutory tax rates on investment income. Unless the tax rate is reduced to zero, however, lowering the statutory tax rate will not completely eliminate distortions affecting capital investment decisions. As discussed in Box 3-1, the effect of lower statutory rates on investment is similar to that of partial expensing of investment. Lowering the statutory tax rate on investment can take many forms—lowering the corporate tax rate, lowering individual tax rates, reducing or eliminating the tax rate on capital gains and dividends, or some combination of these. All of these alternatives have the effect of reducing tax distortions on investment decisions, but the economic effects will differ according to which tax rates are reduced.

One of the biggest misconceptions about pro-growth tax policy is that reducing the statutory corporate tax rate only benefits corporations. The main problem with this argument is that corporations are pure legal entities that cannot themselves bear the burden of taxes. It is households, in their role as owners and users of corporate capital, who benefit from the reduction in corporate tax rates. As discussed in Box 3-2, corporate tax burdens are distributed across all households. The long-run effect of reducing the corporate tax rate is to increase the capital stock, making labor more productive. Ultimately, reducing corporate taxes benefits labor through higher wages and benefits capital owners through higher after-tax returns.

An important goal of pro-growth tax policy is to promote a tax system that does not create distortions that affect the structure of business formation or business investment. By reducing statutory tax rates for corporations or households in an uncoordinated way, the tax system can create incentives that favor certain forms of business. For example, consider reducing the maximum effective corporate tax rate below the maximum effective individual tax rate. This would make it relatively more attractive for businesses to incorporate rather than form as a sole proprietorship or partnership (which pay tax using individual rate schedules). Consolidating the business and individual tax bases would reduce or remove taxes from consideration in business decisions.

Reducing individual tax rates can also reduce tax considerations from capital investment decisions. Perhaps the most direct way to stimulate greater individual saving and investment is to reduce or eliminate the tax rate on capital gains and dividends. This is important because even with full expensing, the effective tax rate on investment is positive as long as there are taxes on capital gains and investment income. Consider two effects from the recent reduction in taxes on capital gains and dividends. First, there was an overall reduction in taxes on corporate income, which stimulated greater investment. Second, the changes reduced the tax distortion that favored returns in the form of capital gains. Prior to JGTRRA, the double tax on corporate income was as high as 42 percent and 61 percent for corporate

income distributed as capital gains and dividends, respectively. After JGTRRA, the double tax on corporate income fell to about 40 percent and 45 percent for capital gains and dividends, respectively. As shown in Charts 3-3 to 3-5, following JGTRRA, real private nonresidential investment rose substantially, and there was an increase in the average amount of dividend payments and the percent of firms paying dividends.

Comparison of Effects of Different Pro-Growth Policies

The primary objective of pro-growth tax policy is to stimulate new investment. New investment leads to a larger capital stock, increases in productivity, higher wages, and economic growth. Full expensing of investment does a

better job than rate cuts in meeting this objective. As noted above, rate cuts reduce but do not eliminate the effect of taxes on new investment decisions. In addition, a tax rate reduction applies to all investments, new and old alike. By contrast, full expensing is carefully targeted towards removing tax considerations from new investment decisions.

One method of comparing policies is to estimate "bang for the buck" measures that show the amount of investment stimulus per dollar of tax cost. These measures are derived by using sophisticated macroeconomic models to simulate the effect of pro-growth policy changes, assuming that each policy change has the same budget effect. As shown in Table 3-2, full expensing provides investment incentives that are 3.5 times as large per dollar of revenue cost compared to reductions in corporate tax rates. The reason for this difference is that much of the revenue cost from statutory rate reductions is from reducing taxes on existing capital. Because expensing applies to new capital only, the potential for economic growth is much greater with expensing than for reductions in the statutory tax rates that have the same revenue cost.

As discussed above, a major issue with expensing is the transition cost imposed on existing capital. It is possible that during the transition to full expensing, the government could provide tax relief to the owners of existing capital. However, the revenue cost of providing this type of transition relief would require rate increases or other tax changes that could reduce the incentive to invest in new capital projects. Estimates of the cost of transition relief range from about 1 percentage point to about 6 percentage points of the long-run increase in real GDP, depending on how and for how long transition relief is paid. Thus it is possible that providing transition relief to owners of existing capital could eliminate all of the efficiency gains from adopting a more pro-growth tax system.

TABLE 3-2.— *Effective Marginal Tax Rates on Investment*

	Effective marginal tax rate on investment	"Bang for the Buck": investment incentive relative to revenue cost (present value)
Current law	17%	
Policy change:		
100% expensing	0%	70%
30% expensing	13%	70%
Corporate tax rate lowered to 25%	15%	20%
Tax rate on dividends and capital gains lowered to 10%................	16%	20%

Source: Department of the Treasury (Office of Tax Analysis).

Conclusion

The goal of pro-growth tax policy is to finance a given level of government services in a way that minimizes the drag imposed on the economy by tax distortions on investment decisions of households and businesses. Of particular importance is the effect a tax system may have on capital investment decisions. Taxing capital in a way that distorts investment decisions can affect the level, allocation, and financing of new projects. Reducing the tax on capital income will lead to a larger capital stock and higher standards of living. With more capital available, labor becomes more productive and real wages rise.

An incremental approach to pro-growth tax policy would be a transition to a tax system that allows full expensing of capital investment. Research indicates that we could expect up to a 8-percent increase in long-run real GDP from adopting the pro-growth policy of full expensing. Full expensing provides relatively more bang for the buck because it targets new investment, whereas rate cuts benefit old and new capital alike.

Reducing or eliminating distortionary capital taxation leads to a more efficient level and allocation of capital throughout the economy. This increase in efficiency in turn results in higher productivity, GDP, and standards of living. While there have been recent changes to a more pro-growth tax system, the temporary nature of the provisions reduces the long-run impact of these policy changes on investment and economic growth. Making these changes permanent would ensure a tax system that minimizes tax distortions to investment decisions that can keep the economy from reaching its long-run potential.

The Fiscal Challenges Facing Medicare

Social Security, Medicare, and Medicaid are three vital entitlement programs in the United States that provide people with important economic security against the financial risk associated with retirement, disability, and medical expenses. In 2006, the Federal Government spent $1.1 trillion on these entitlement programs; this amount is projected to grow to $1.5 trillion by 2012. In the absence of reforms to either raise more revenue or restrain future spending, excess growth in entitlement spending will need to be offset by reductions in discretionary spending, putting significant pressure on other important programs. As history has shown, there is no uncontroversial way to reform these entitlement programs. Reforms to increase tax revenue will have negative effects on the economy. At the same time, it is crucial that any spending reforms preserve the protection against financial risk that these programs provide. Thus, improving the efficiency of these programs is crucial to slowing the growth of entitlement spending.

This chapter focuses on Medicare. It begins with a brief overview of the program and then examines the main reasons for the projected financial pressures facing Medicare. It concludes with a discussion of ways to improve the efficiency of Medicare spending and thus the long-term financial outlook of this important program. The key points in this chapter are:

- The projected long-term growth in entitlement spending, including Medicare, is unsustainable because of the pressures it places on future Federal budgets and by implication, on the economy.
- Medicare spending is growing quickly, primarily because of the demographic shift to an older society and the increases in per-beneficiary medical spending driven largely by new technologies.
- Rewarding providers for supplying higher quality care and improving incentives for patients to choose higher value care can both increase the efficiency and slow the growth of Medicare spending.

Entitlement Spending and Medicare

Social Security, Medicare, and Medicaid are *entitlement programs;* that is, individuals who are eligible for these programs are entitled to particular benefits. Social Security provides income to seniors, the disabled, and surviving spouses and dependents. Medicare provides health insurance to retirees and the disabled. Medicaid provides health insurance to certain lower income

groups. Workers and their spouses are entitled to receive Social Security and Medicare benefits if they make sufficient payroll contributions while working, and citizens and qualified aliens are entitled to Medicaid benefits if they meet certain income and other demographic criteria.

Chart 4-1 shows spending on Social Security, Medicare, and Medicaid in 2006 as a percent of the total Federal budget. The $549 billion in Federal spending on Social Security benefits was 21 percent of total Federal outlays. The $330 billion in federal spending on Medicare benefits was 12 percent of outlays. The $191 billion in federal spending on Medicaid was 7 percent of outlays. Because Medicaid is jointly funded by the Federal and State governments, State governments also spent about $139 billion on Medicaid.

For those not covered by Medicare or Medicaid, the federal government also helps with the purchase of private health insurance coverage in a variety of ways, including the exclusion of employer contributions towards health insurance premiums from personal income taxes. These tax expenditures are included in the Federal budget and are estimated to equal $133 billion in 2006. The President's 2008 budget includes a proposal to replace the existing exclusion for employer-provided health insurance with a flat standard deduction to all families who purchase health insurance that meets minimum requirements for catastrophic coverage, in order to improve the efficiency and equity of these tax expenditures. The President's policy proposal is described in Box 4-1.

Chart 4-1 2006 Government Outlays
Entitlement spending consumes 40 percent of Federal Government outlays.
Billions of dollars

Non-Defense Discretionary $525 19%

Social Security $549 21%

Department of Defense $510 19%

Medicaid $191 7%

Medicare $330 12%

Interest $219 8%

Other Mandatory $372 14%

Note: Shaded areas indicate entitlement programs.
Source: Office of Management and the Budget.
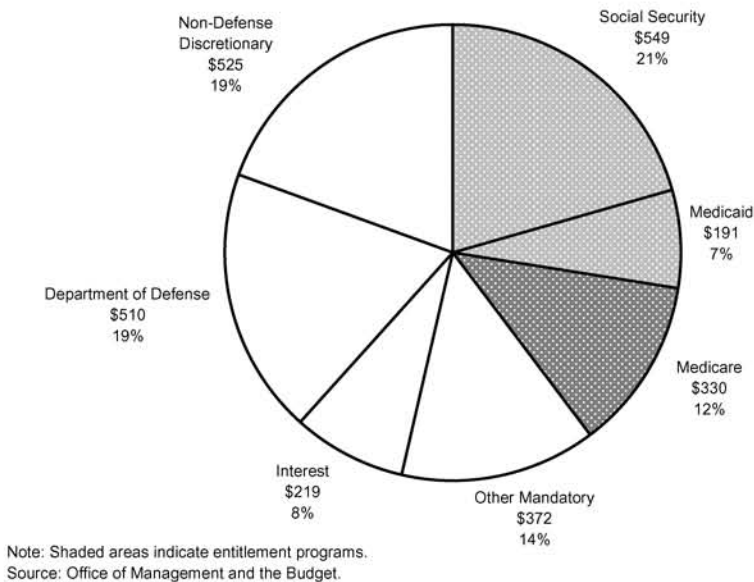
Box 4-1: The President's Proposal to Improve the Tax Treatment of Private Health Insurance

The current tax treatment of private health insurance coverage is both inequitable and inefficient. Employer contributions (and in most cases, employee contributions) toward private health insurance coverage are exempt from income and payroll taxes. This is inequitable because it does not offer the same tax break to families that do not have access to employment-based insurance and instead purchase a private plan in the individual health insurance market. It is also inefficient because it provides a larger tax break to families with more generous health insurance policies, which in turn can drive the inefficient use of medical care of low value. For more detail about these inefficiencies, see Chapter 4 of the 2006 *Economic Report of the President.*

The President's 2008 Budget has proposed reforming the current open-ended tax exclusion for employment-based health insurance coverage, effective in 2009, with a flat $15,000 standard deduction for health insurance to all families (or $7,500 for individuals), whether that insurance was obtained through their employer or on their own. The amount of this standard deduction would be independent of the actual amount spent on the premium, so families who obtain insurance policies for less than $15,000 (but satisfying a set of minimum requirements for catastrophic coverage) would still be able to exempt the full $15,000 of compensation from income and payroll taxes. The annual increase in the standard deduction for health insurance would be linked to the Consumer Price Index, and the policy would be roughly budget neutral.

This policy would reduce inequity in the tax code by providing the same tax treatment of health insurance purchases to families with or without access to employment-based health insurance. Those who are currently insured in the individual health insurance market would see a reduction in taxes commensurate with those insured in the group market, and those who are currently uninsured would be given a strong incentive to purchase coverage. For instance, for an uninsured family of four with $50,000 in income facing a 15 percent marginal income tax rate and a 15.3 percent total combined payroll tax, the value of the $15,000 exclusion would be worth about $4,500, and would thus offset the cost of roughly half of a health insurance plan costing $9,000.

This policy would also reduce the inefficiency of the current tax treatment of employment-based health insurance. An insured wage-earning family of four with $50,000 in income currently receives a tax break of about $3,000 toward a $10,000 policy but about $6,000 toward a

Spending on Social Security, Medicare, and Medicaid is projected to increase and claim an even more significant share of the federal budget in the future. Examining total spending as a fraction of gross domestic product (GDP) is especially relevant because this measures the portion of the overall economy devoted to each particular program. For instance, Social Security spending was 4.2 percent of GDP in 2005 and is projected to be 6.3 percent of GDP in 2080. Total Medicare spending was 2.7 percent of GDP in 2005 and is projected to be 11.0 percent of GDP in 2080. Total health care spending in the United States by private and public sources combined was 16.0 percent of GDP in 2005, equaling almost $2.0 trillion or $6,697 per person. Although national health expenditures have grown at a slower rate than the previous year for the prior 3 years, health spending has still consistently grown at a faster rate than general inflation.

While Social Security, Medicare, and Medicaid share some common features, each also poses its own opportunities and challenges, warranting detailed specific analysis. Chapter 5 of the 2002 *Economic Report of the President* examined Medicaid coverage for low-income families, Chapter 6 of the 2004 *Economic Report of the President* examined Social Security, and Chapter 4 of the 2006 *Economic Report of the President* examined health care spending generally. This chapter focuses primarily on Medicare.

The Basics of Medicare

A primary motivation behind the passage of Medicare in 1965 was that many of the elderly at the time had no health insurance. Medicare was structured to mimic the prevalent form of private health insurance at the time, Blue Cross and Blue Shield. Blue Cross plans covered inpatient hospital services, and Blue Shield plans covered physician and hospital outpatient services. The "Blues" were the basis for separate Part A and Part B plans that reimburse hospitals and physicians on a fee-for-service basis, respectively. Seniors who have worked at least 40 quarters in qualified employment are automatically

enrolled in Part A at age 65. Seniors who lack 40 quarters of employment can buy into Part A by paying a monthly premium. People under the age of 65 with certain disabilities or end-stage renal disease are also eligible for Medicare. Enrollment in Part B is optional and requires a premium contribution, although there is a penalty for not immediately enrolling and the amount is higher for individuals making more than $80,000 per year. The Centers for Medicare and Medicaid Services (CMS) administers the Medicare program by implementing the statutes that determine the form of payments to hospitals, physicians, and outpatient providers.

Most outpatient prescription drugs were not covered by Medicare until the implementation of the Medicare Modernization Act (MMA) of 2003, which created Part D of Medicare. Like Part B, Part D is optional, requires a premium contribution, and has a penalty for late enrollment. Unlike Part B, however, Part D is administered by private health insurance plan sponsors. Seniors have the alternative option of enrolling in a private Medicare Advantage insurance plan if one exists in their region. These are private health insurance plans that provide Part A, Part B, and, in most cases, Part D services. These plans often provide additional benefits to seniors at lower costs. The Medicare Advantage program is described in more detail in Box 4-2.

Box 4-2: The Medicare Advantage Program

Approximately 16 percent of Medicare beneficiaries are enrolled in private managed-care health plans, including primarily health maintenance organizations (HMOs) but also preferred provider organizations (PPOs) and private fee-for-service plans. These Medicare Advantage plans contract with Medicare to provide the services covered by Part A and Part B and usually offer additional benefits such as relatively lower cost sharing and additional covered services. Enrollment into these plans is voluntary but requires that a local plan is available. As of 2006, all Medicare beneficiaries had the option of enrolling in a Medicare Advantage plan, including plans that provide prescription drug coverage.

Prior to 1997, Medicare HMOs received a *capitated* payment based on 95 percent of the average Medicare beneficiary spending in the county, adjusted only for age, gender, Medicaid enrollment, and disability status. Studies suggest that healthier beneficiaries were more willing to enroll in these plans, because HMOs typically place restrictions on care. As a result, the program increased total Medicare expenditures because the payments to the HMOs were generally higher than the actual costs of their enrollees in the fee-for-service program.

   The 1997 Balanced Budget Act eliminated the direct link between plan payment rates and local fee-for-service expenditures and sought to expand the types of plans available to beneficiaries beyond the urban areas where they had generally been available. The 1997 Balanced Budget Act also mandated the use of risk adjustment to vary the payments to insurers based upon the health status of its enrollees by 2000. As a result, incentives to engage in wasteful competition for relatively healthier enrollees were mitigated so that insurers would instead engage in competition to provide higher value care at a lower cost for all enrollees. Because of some of the limits on the growth in payments in the 1997 Balanced Budget Act, many private insurers withdrew from the Medicare market. Enrollment declined by about 25 percent from 1999 to 2003.

   The 2003 Medicare Modernization Act expanded the Medicare Advantage program in two important ways (in addition to changing the name from "Medicare+Choice" to "Medicare Advantage"). First, the 2003 Medicare Modernization Act increased the payment levels to the plans to encourage participation across all Medicare Advantage plans. Second, the 2003 Medicare Modernization Act created new regional preferred provider organizations that offer a uniform deductible and an upper limit on out-of-pocket spending to increase both the number of choices available to Medicare beneficiaries (especially in rural areas) and special needs plans to target certain beneficiaries (such as those with dual eligibility, those with chronic conditions, and the institutionalized).

   Medicare spending is financed by a combination of payroll taxes, general revenue, and premiums paid by beneficiaries. Part A of Medicare is financed by a Hospital Insurance (HI) payroll tax of 2.9 percent. The HI payroll tax is split evenly between employees and employers, but economists generally believe the employer tax is ultimately paid by workers in the form of relatively lower wages. Part A is a pay-as-you-go system in which payroll taxes on current workers' wages finance the benefits of those currently retired. If the payroll tax revenues exceed spending for the year, the difference is placed into the HI Trust Fund. If taxes are lower than spending, money is withdrawn from the HI Trust Fund. Parts B and D constitute the Supplementary Medical Insurance component of Medicare and are financed by general Federal government revenues and beneficiary premiums, which are set to equal approximately 25 percent of total Part B and Part D spending, respectively.

Nations around the world provide various forms of social insurance for their elderly populations. One of the purposes of health insurance is to ensure that people are protected against the financial risk associated with uncertain medical spending. Economists generally attempt to justify government intervention into private market outcomes by suggesting potential market failures that may exist in the absence of any government intervention. Many economists would justify the existence of Medicare (and its government provision of health insurance for the elderly and disabled) with three potential explanations. The first potential explanation is that many people may lack sufficient information to plan properly for the financial hardships that would otherwise arise from expensive medical treatment when they age or become disabled. Medicare requires workers to pay a premium during their working years toward future costs and thus the program can be considered a form of forced savings. In this way, Medicare is similar to Social Security, which requires people to set aside some of their wages now in exchange for a promise of income at retirement. But this reason alone is insufficient to explain the provision of health insurance as opposed to additional income.

A second potential explanation for government intervention in the provision of health insurance for seniors is to avoid having seniors in poor health pay considerably more toward their health care. In the United States, most people participate in health insurance plans through their place of employment. Most people lose these plans upon retirement. (Private retiree health insurance plans only cover what Medicare does not.) Because about 40 percent of people at age 65 have at least one serious preexisting chronic health condition, initiating coverage in a private individual health insurance market after retirement (under the assumption that the Medicare program did not exist) would force insurers to charge higher premiums to those in poor health. Younger people face uncertainty that they may develop a chronic condition in the future (and thus they would face variable premiums in the absence of Medicare). This suggests that there may be efficiency gains from providing future insurance coverage with pooled contributions. (Private health insurance markets handle this intertemporal uncertainty of developing a chronic health condition with "guaranteed renewal at class average rates" provisions that ensure that premiums do not vary with the onset of illness for those with coverage.)

A third potential explanation for government intervention in the provision of health insurance is related to the redistribution of resources toward low-income people. Economic theory suggests that unconditional transfers of wealth are generally more efficient than in-kind transfers of goods or services for achieving any desired redistribution. In an ideal world, the poor would use some of this transferred wealth to purchase health insurance. However, if the poor believe that society will provide them with additional resources in the

event of an uninsured loss, they may have an incentive to forego buying insurance. This precommitment problem, sometimes called the "Samaritan's Dilemma," has been demonstrated to be alleviated by the direct provision of health insurance rather than a direct transfer of wealth. This economic argument, however, justifies the subsidization of, or requirement for, insurance but does not justify a government-run plan.

Increases in Medicare Spending over Time

Projections of Future Medicare Spending and Revenue

Sources of Spending

Since Medicare was created in 1965, total spending on all of its programs has grown steadily. As noted above, total Medicare spending was 2.7 percent of GDP in 2005 and is projected to be 11.0 percent of GDP in 2080. These values for Medicare spending, however, actually understate the total spending for Medicare beneficiaries because the private payments for cost sharing are not included. For instance, in 2006, Part A requires individuals to pay $952 of the cost of each hospitalization (this $952 is called a *deductible*), and Part B generally requires them to pay 20 percent of the Medicare-approved payment (this 20 percent is called *coinsurance*) in addition to a deductible. Some beneficiaries pay Medicare deductibles and coinsurance amounts from their own pockets, while others obtain private insurance to cover these costs. Some of this private coverage is included in employer-sponsored retirement benefits, while some is provided by directly purchased Medigap plans. Some low-income Medicare beneficiaries are also eligible for Medicaid. For these *dually eligible* people, Medicaid covers most of these cost-sharing amounts required by Medicare.

Chart 4-2 shows historical and projected private and public spending for Medicare-covered services as a percentage of GDP for 1966 through 2050. Including private spending by Medicare beneficiaries and Medicaid spending on Medicare beneficiaries presents a more complete picture of beneficiaries' total consumption. In 2006, beneficiaries bore about 37 percent of Medicare-related spending, and about 63 percent was financed by payroll taxes and general revenues. However, these amounts shown here do not include the portion of Medicaid spending on long-term care services, such as nursing homes, because this type of care is not covered by Medicare. More detail about coverage of long-term care is provided in Box 4-3.

Chart 4-2 Total Healthcare Spending by Medicare Beneficiaries, 1966-2050
Government and private Medicare spending has grown rapidy and is projected to continue growing.

Source: Council of Economic Advisers analysis of data compiled from the 2006 Medicare Trustees' Report and the Medicare Current Beneficiary Survey.
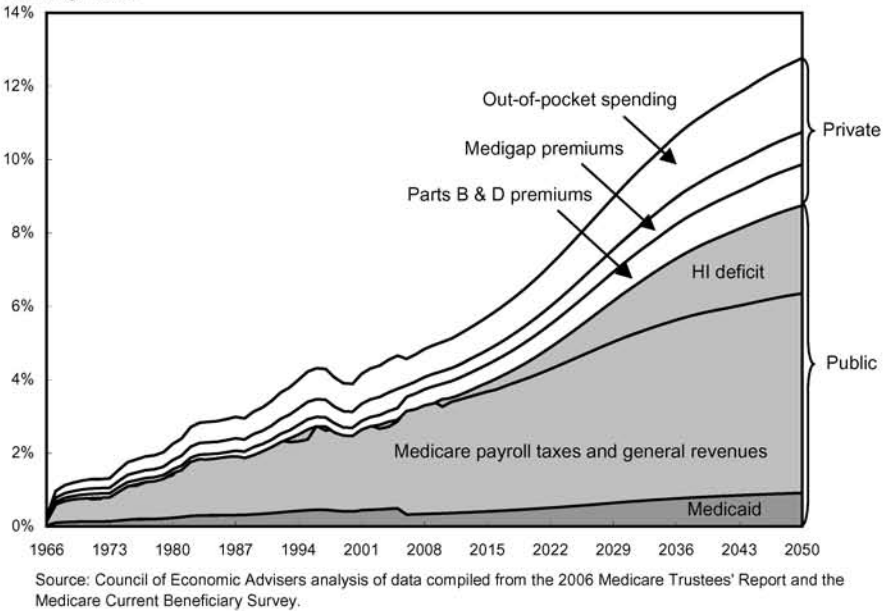
Medicare Solvency

The Medicare program does not have enough projected revenue to cover projected future spending. Under current projections made by the Medicare Actuaries and presented in the 2006 Medicare Trustees Report, the Medicare HI Trust Fund is projected to be exhausted in 2018. The projected 75-year deficit for the Medicare HI Trust Fund is 3.51 percent of taxable payroll. That is, the Medicare HI payroll tax would have to be immediately increased from 2.90 percent to 6.41 percent to cover all projected spending over the next 75 years. Alternatively, a reduction in Medicare Part A expenditures by 51 percent would be necessary to make the Medicare Trust Fund solvent. As a comparison, this Medicare deficit is relatively larger in magnitude than the Social Security Trust Fund deficit. An increase in the Old Age, Survivors, and Disability Insurance (OASDI) payroll tax from 12.4 percent to 14.4 percent or a reduction in Social Security benefits by 13 percent is projected to make the Social Security program solvent over 75 years.

The Medicare Supplementary Medical Insurance (SMI) program is considered to be solvent by the Medicare Trustees only because Part B and Part D spending is required by law to be financed by general revenues. However, the consequences of increased spending on Medicare SMI may be

Box 4-3: Long-Term Care

Nine million people use long-term care (LTC) to alleviate the hardships accompanying old age or disability. LTC is medical care required over a long period of time by someone with a chronic illness or disability. An estimated 70 percent of people who reach the age of 65 will need some form of LTC before they die. Medicare does not have a large LTC component, as it only covers post-acute care in skilled nursing facilities and some home health care, which total less than 20 percent of all LTC. Private, noninsured spending covers about 25 percent of LTC expenditures, while private insurance pays for less than 10 percent. Many Medicare beneficiaries obtain LTC after they have depleted their assets and become eligible for Medicaid. Medicaid LTC eligibility is often tied to receiving Supplemental Security Income and having very few assets, but states have the discretion of easing eligibility criteria. Medicaid covers over 45 percent of all LTC expenditures. About one-third of Medicaid expenditures go to LTC.

The average price for 1 year in a nursing home is $70,000. This cost is high enough to strain even middle-income families, yet few people prepare financially for potential LTC expenses. Studies generally attribute failure to purchase LTC insurance to a lack of awareness about the potential costs of LTC, the benefits of coverage, and a misperception that Medicare covers all LTC. Adverse selection in the market (by those who expect to use long-term care being more likely to purchase insurance) results in very high premiums and relatively fewer insurance companies offering LTC policies. Many seniors forgo obtaining private coverage and instead become Medicaid-eligible by sheltering their assets through income annuities, trusts for their children, and asset transfers to family members. In response to these loopholes, States and the Federal government have tightened Medicaid eligibility. Because of the pressure LTC places on State budgets, many policymakers believe that changes should be made to LTC administration.

Encouraging the purchase of private long-term care insurance may be a valuable step in reducing Medicaid spending on LTC while protecting seniors from poverty. For example, New York currently has a 20 percent tax credit available toward the purchase of LTC insurance. Such a subsidy should generally make LTC insurance more attractive to middle-aged people. Medicaid *spend-down insurance*, which permits people who purchased and used LTC insurance to keep some assets and still qualify for Medicaid, could also increase the attractiveness of private LTC coverage.

just as dire. Without large reductions in Medicare SMI spending or increases in taxes, either Federal budget deficits will grow rapidly or dramatic reductions in spending for other Federal programs will have to be made.

Spending on Medicaid is also funded by general revenues. The elderly and disabled covered by Medicare account for about one-quarter of Medicaid enrollees, but they account for about two-thirds of Medicaid spending, mainly because of spending on acute and long-term care. An additional challenge for funding Medicaid is the inverse relationship between the proportion of the population eligible for benefits and the tax base available to fund the program. During economic downturns, lower personal income causes State governments with balanced-budget requirements to face the strain of both a decrease in tax revenue and a higher number of residents who meet the low-income eligibility threshold and are thus in need of assistance.

Implications for Reform

In light of the mounting fiscal pressures on entitlement spending, it is critical to increase the efficiency of spending on benefits. Reforms of the Medicare program should aim to reduce the growth of spending by redirecting resources toward the highest value uses and away from inefficient care of low value. Controlling cost growth while preserving the vital financial and health protections offered by the program is particularly important in light of the large negative consequences of raising taxes. An increase in the payroll tax rate would decrease incentives to work, increase efforts to receive compensation in forms not subject to taxation, and be a drag on economic growth.

As noted above, Medicare taxes on current workers' wages essentially fund an insurance pool from which benefits are paid on behalf of retired or disabled workers. A pay-as-you-go system of intergenerational transfers is consistent with the basic idea behind insurance if the aggregate amount paid into the pool (in the form of taxes on workers) equals the aggregate amount of expected benefits to be paid from the pool. In private insurance markets, policyholders must have confidence that future claims will be covered by the insurer. To help alleviate consumer concerns, government regulations often place solvency requirements on insurers that require them to have enough assets to cover their liabilities. Thus, for Medicare's pay-as-you-go financing mechanism to function as a social insurance program, younger generations must have confidence that the government will indeed meet its future insurance obligations to them. The rapid increase in Medicare spending over time clearly threatens the confidence that younger generations have in the solvency of the program. Indeed, a recent survey found that almost two-thirds of workers are "not too confident" or "not at all confident" that Medicare "will continue to provide benefits of at least equal value to the benefits received by retirees today".

The next section of this chapter examines the reasons behind this projected growth in Medicare spending. The average annual growth rate of Medicare spending is projected to be 2.8 percentage points higher than GDP growth per year between 2006 and 2040. Part of this increase in spending is due to growth in the number of Medicare beneficiaries, and part of this increase in spending is due to growth in real (inflation adjusted) Medicare spending per beneficiary.

Reasons for the Changes in Medicare Spending over Time

Increases in the Number of Medicare Beneficiaries

The proportion of the United States population covered by Medicare has increased over time. This has resulted from the normal eligibility age remaining fixed at 65 combined with the aging of the population. The aging of the population is due to both increased life expectancy and decreased fertility. In 1965, 65-year-old retirees could expect to live for 14.7 more years; by 2006, they could expect to live for 18.6 more years. In 1965, the fertility rate was 96.3 births per 1,000 females aged 15 to 44; by 2004, it had fallen to 60.7 births. (These changes in demographics have a similar effect on Social Security.)

The worker-per-beneficiary ratio illustrates the portion of the population which provides revenue to cover the needed spending on Medicare beneficiaries. In 1965, there were about 4.6 workers for each Medicare beneficiary. In 2005, there were about 3.8 workers for each Medicare beneficiary. In 2050, there are projected to be only 2.2 workers for each Medicare beneficiary.

In addition to being affected by long-term increases in longevity and decreases in fertility, the worker-per-beneficiary ratio during the upcoming years is also affected by the aging of the baby boom generation, which is made up of those born between 1946 and 1964. (The baby boom generation can be viewed as a temporary change in fertility rates.) The baby boom generation explains the relatively steady worker-per-beneficiary ratio between 1975 and 2005 and the dramatically decreasing ratio between 2010 and 2040. After 2050, most benefits owed to the baby boom generation will have been paid, and the worker-per-beneficiary ratio is projected to be relatively steady though 2080 as long as current assumptions hold.

Unlike Medicare, the full retirement age for Social Security is 65 for those born in 1937 and earlier, and will rise slowly to 67 for those born in 1960 or later. However, the effect of increasing the eligibility age for Medicare would not have a very large effect on total Medicare spending, because Medicare

spending increases with age as people become less healthy. For instance, while people ages 65 and 66 represent about 9 percent of the Medicare population, they are the recipients of only about 4 percent of total Medicare spending.

Increases in Spending per Beneficiary

Real growth in Medicare spending per beneficiary has averaged about 4 percent per year between 1996 and 2006, roughly 2 percentage points greater than real per capita growth in GDP. For the Medicare Trustees Report, the Medicare actuaries assume that the annual growth rate of Medicare spending per beneficiary during the period between 25 and 75 years from now will decrease to equal the growth rate of GDP per capita plus an average of 1 percentage point. In addition to this so-called "intermediate" assumption, these actuaries also consider a "low-cost" assumption, in which annual Medicare spending growth equals per capita GDP growth and a "high-cost" assumption, in which annual Medicare spending growth equals per capita GDP growth plus 2 percentage points.

One way to evaluate the affordability of these projected increases in Medicare spending is to consider the effect of applying this growth rate to overall medical spending in the United States and examine the resulting growth in consumption of all other goods and services in the future economy (that is, nonmedical consumption). One study estimated that applying the intermediate assumption of long-term medical spending growth, equal to the growth rate of per capita GDP plus 1 percentage point, would still result in positive real growth in the level of nonmedical consumption over the next 75 years. However, the high-cost assumption of long-term medical spending growth, equal to the growth rate of per capita GDP plus 2 percentage points (and, as noted above, roughly equal to the growth rate of Medicare spending in recent history), would cause the level of real nonmedical consumption to increase only until year 2040 and decrease thereafter. During the period between 2010 and 2040, an average of over 60 percent of the annual increase in income would be allocated toward health care spending.

Research suggests that most of the increase in medical spending over time has been driven by the advent of new technologies. New technologies make available new treatments, some of which are more effective than others. Research also suggests that the increased medical spending has, on average, resulted in improvements in health with additional value exceeding the additional costs. For instance, the real cost of treating heart attacks increased by about $10,000 for Medicare beneficiaries between 1984 and 1998, driven by technological advances such as catheterization and angioplasty. Life expectancy for heart-attack patients increased by about 1 year during this same period. Although it is difficult to measure the value of human life and

it is not clear that this relationship is causal, an estimate of the value of these added health benefits is about $70,000, far in excess of the added costs.

Economists have suggested that an increase in medical spending over time is not necessarily problematic, in and of itself, so long as the marginal benefits exceed the marginal costs. A simple cross-national comparison of the fraction of GDP devoted to health care spending suggests that the United States is a high-expense outlier relative to other developed countries. However, it is plausible that the marginal benefits of improved health are dependent on income, so that as a country's GDP increases, it may be rational for that country to devote a relatively higher share of its GDP to health care. This perspective suggests that it may make sense for the United States to spend more than other countries because it has higher per capita income and health care can be a valued use of those higher resources.

Improving the Efficient Allocation of Resources in Medicare

The remainder of this chapter considers ways to improve the efficiency of spending in the Medicare program, in order to slow the projected growth in spending. Policymakers face the challenge of enacting policies that limit inefficient health care spending but do not limit efficient health care spending or the development of beneficial new technologies. This section begins by providing several examples of sources of inefficiency in health care spending and concludes by suggesting several ways to improve the incentives that providers and Medicare beneficiaries face. Improving the efficiency of health care spending is critical to improving both the long-term fiscal strain on the Medicare program and the quality of care to patients, and it is likely that a multipronged approach will be necessary.

Inefficient Health Care Spending

While some of the greater health care spending may be attributed to technological improvements that enhance the quality of care and to increases in national wealth, there are also many findings that are consistent with some degree of inefficiency associated with relatively higher health care spending. Health outcomes in the United States are often not substantially better than those in other developed countries that spend far less on health care. The Rand Health Insurance Experiment found that increased medical spending led to only limited health improvements. The Dartmouth Atlas of Health Care shows wide variations in Medicare spending within the United States without associated variation in health or health outcomes.

It may, at first, appear to be difficult to reconcile the research findings that new technologies over time produce valuable health benefits with the research findings that higher spending does not yield better outcomes. It is likely that there is significant overconsumption of health care that provides little marginal benefit. Consider a costly new technology that provides very large health benefits to specific patients in need. Suppose, however, that it is also consumed by patients who benefit very little from the treatment. If the benefits to "appropriate" patients are very large, the increase in spending over time on both "appropriate" and "inappropriate" patients combined can still imply that the new technology is cost effective. However, because some "inappropriate" patients also receive the treatment, some of the variation in spending is due to inefficiency. If this characterization is accurate, the technology is not as cost effective as it should be.

This overconsumption of health care is frequently thought of as being caused by poor incentives such as overly generous health insurance coverage. That is, patients often face marginal prices for costly treatments that, due to insurance coverage, are lower than the true marginal costs of treatment. (More detail on optimal forms of private health insurance and the effect of increasing cost sharing by consumers is provided in Chapter 4 of the 2006 *Economic Report of the President*.) The presence of generous health insurance may also influence the research and development of certain technologies with questionable cost effectiveness.

There is also evidence of significant underuse of valued health care. For example, there is a large body of medical literature demonstrating the cost effectiveness of beta blockers for patients recovering from a heart attack. Due to their effectiveness, they are prescribed in over 90 percent of cases. However, studies have shown that persistence in use of beta blockers declines rapidly even in the first year of treatment. Moreover, the U.S. Preventive Services Task Force recommends that all women over 40 receive mammograms every 1 to 2 years, that all adults over 50 receive regular colorectal screenings to detect colon cancer, and that all adults over 50 receive annual immunizations against influenza. Compliance, however, is low: 68 percent of women receive recommended mammograms, 35 percent of adults receive recommended colorectal cancer screenings, and 65 percent of adults over 65 receive annual influenza vaccines.

These data suggest that there are two main ways in which the efficiency of Medicare spending could be improved, because there is both a relationship between the insurer and beneficiaries and a relationship between the insurer and providers. One is to encourage the use of cost-effective care that is currently underconsumed. Medicare now covers an initial preventive physical examination and many preventive screenings, but there are still potential improvements to be made. Policies to achieve this goal should aim to improve

the incentives for health care providers and insurers to provide high-quality care. A second way to improve the efficiency of Medicare spending is to discourage the use of ineffective care that is currently overconsumed. Policies to achieve this goal should aim to improve the incentives that Medicare beneficiaries face regarding their consumption of care. More detail on these policies is provided in the next two sections.

Better Incentives for Health Care Providers and Insurers

Medicare generally pays providers of the same service the same fee, regardless of the quality of care. If hospitals and physicians were paid amounts that reflected objective measures of the quality of care provided, with differential payments tied to higher quality and more efficient care, ideally many problems of underuse and misuse of care could be reduced. In practice, while "pay for performance" holds a great deal of promise, it may be difficult to fully implement because of the complexity of producing objective measures of quality. For instance, tying payments to process measures—such as rewarding cardiac physicians based on the proportion of their heart attack patients using beta blockers—may cause providers to place too much emphasis on limited aspects of providing high-quality care. Alternatively, tying payments to outcomes measures—such as rewarding cardiac surgeons whose patients have lower post-discharge mortality rates—may cause providers to face perverse incentives to avoid treating high-risk patients most in need. Adequate pay-for-performance measures will require sophisticated techniques to control for underlying differences in patient health, which highlights the importance of developing systems to collect detailed information about the kind of care that patients receive. With the advent and adoption of better health information technology and the development of rigorous and well-tested measures, using pay-for-performance techniques to reimburse providers may become a vital contributor toward higher quality and more efficient care.

High-quality health care may also be encouraged by providing patients with valuable information so they may compare various providers to one another. Competition among health care providers may improve incentives to provide high-value care in two ways: higher quality and lower price. If patients have access to the providers' price and quality information, they will have incentives to choose those providers with the highest value of care, and physicians and hospitals will have strong incentives to reduce their fees and improve the quality of care to attract more patients. There are two parts of Medicare where this kind of information is available and these incentives are in place. Private Medicare Advantage plans have strong incentives to offer higher quality care at lower beneficiary premiums to encourage enrollment. The new Part D prescription drug benefit provides information about the

price of prescriptions by plan and by pharmacy, provides access to customer service information by plan, and also benefits from price competition among insurers. More detail on the structure of and experience with the new Medicare Part D benefit is provided in Box 4-4.

Box 4-4: Medicare Part D Prescription Drug Benefit

The Medicare Part D prescription drug benefit went into effect January 1, 2006, as a result of the 2003 Medicare Modernization Act. Prior to that date there was almost no coverage for outpatient prescription drugs in Medicare, except in Medicare Advantage plans. (Part B does cover drugs in certain instances.) Part D beneficiaries may now enroll in their choice of plans in their region. In 2007, the 34 regions will offer between 45 and 66 standalone prescription drug plans at different prices with varying levels of coverage at or above the minimum benefit package. If an individual seeks greater benefits, they will generally pay a higher premium. Individuals with incomes below 150 percent of the Federal Poverty Level who meet eligibility requirements receive additional assistance in the form of reduced premiums, deductibles, and coinsurance. The premium subsidies are on a sliding scale to better target those with the lowest incomes. By June of 2006, over 38 million Medicare beneficiaries had some form of prescription drug coverage.

One important feature of the Part D program is the competitive premium bidding process by insurers. Each year insurers submit premium bids for the following year to Medicare. These premium bids are weighted by enrollment to determine the weighted average bid; this amount is referred to as the *benchmark premium*. The basic premium that nonpoor Medicare beneficiaries pay for a specific plan is the difference between the plan's bid and 75 percent of the weighted average bid (that is, the federal direct subsidy). Some low-income beneficiaries are automatically enrolled in plans whose premiums are at or below the regional enrollment-weighted average. Thus, there are significant incentives for insurers to submit low bids. Early projections suggested that the average premium in 2006 would be $37 per month, but premiums ultimately averaged $24 per month. In 2007, the average premium is expected to remain about the same.

Competitive bidding appears to be a successful model for providing low costs to both beneficiaries and the government without government interference in determining drug prices. Satisfaction with the Part D program is high. Several surveys have shown that at least 75 percent of enrollees are pleased with the Part D benefit.

Better Incentives for Medicare Beneficiaries

In addition to the competition induced by the new Part D benefit, its pricing structure and associated subsidy for premiums provide good incentives for Medicare beneficiaries to obtain relatively more efficient forms of insurance coverage. Because the Federal subsidy toward the prescription drug plan is generally a fixed proportion of the average premium bid each year, beneficiaries receive the additional benefits of choosing plans that are less generous than the average benchmark plan. Thus, beneficiaries appropriately receive the full marginal benefits from either a higher amount of cost sharing or a more restrictive list of covered medicines. This mechanism for having Medicare beneficiaries pay lower amounts for less generous coverage therefore improves the incentives for insurers to design more optimal products.

A potential downside to this mechanism for determining beneficiary premiums, however, is that it could lead to relatively higher premiums for people with higher expected expenses due to chronic health conditions if these high-risk people gravitate toward plans with relatively more generous benefits. As a result, these plans' higher premiums would reflect a relatively sicker pool of people covered by the plan, in addition to the underlying value of more generous benefits. However, these potential problems can be alleviated by the use of *risk-adjusted* payments to plans, as described in Box 4-2.

This mechanism for determining the premium contribution toward different plans, currently in place for Part D, could potentially be applied to the entire Medicare program. Providing beneficiaries with a choice of comprehensive plans and having the premium contribution for each plan vary in relation to a benchmark plan has potential for improving the efficiency of overall Medicare spending. A key difference between Medicare Part D and the entire Medicare program, however, is the combination of the government-run fee-for-service and Medicare Advantage components of the latter. This benchmark mechanism is likely to be successful only if the same premium contribution is made toward both the fee-for-service component of Medicare and the private Medicare Advantage plans, putting them on equal footing. Just as described above, this mechanism for determining premium contributions would cause beneficiaries to receive the appropriate marginal benefits when choosing plans with levels of coverage that are less generous than the benchmark plan. It could therefore help to allow beneficiaries to determine the optimal forms of out-of-pocket cost sharing and the optimal adoption of new technologies over time. These two specific issues are explored below.

Premiums versus Out-of-Pocket Payments

The level of out-of-pocket cost sharing that would induce beneficiaries to consume the optimal level of care is difficult to determine. The share of out-of-pocket spending that will lead to an efficient amount of care would be set

at the level at which the marginal cost of being exposed to more financial risk through relatively more cost sharing is less than the marginal benefits from reducing the overconsumption of medical care resulting from relatively more cost sharing. In practice, it is difficult to quantify these competing interests. Nevertheless, Medicare currently may be missing this balance at both the high-cost and low-cost extremes. Medicare currently does not provide protection against certain catastrophic health care costs (except in some Medicare Advantage plans). For example, there is increased beneficiary cost sharing after a hospitalization exceeds 60 days, and a cessation of benefits after 120 days. While these upper limits on benefits presumably have the advantage of reducing incentives to over consume, they appear to expose beneficiaries to excessively high levels of financial risk.

While many seniors have private retiree health or Medigap plans to cover Medicare's gaps in catastrophic coverage, these plans also frequently cover the first-dollar cost sharing, such as the hospitalization deductible and the 20 percent of physician fees. These plans limit the cost-consciousness of consumers and therefore increase total spending. However, neither insurers nor consumers bear the full marginal costs of the increased spending induced by these generous Medigap plans, because Medicare covers most of the increased spending.

If beneficiaries were to receive the marginal benefits of less generous coverage in a way that puts the fee-for-service component and the Medicare Advantage component on equal footing, there would be improved incentives for private plans to offer and beneficiaries to select plans with more efficient levels and forms of cost sharing. Beneficiaries, rather than Medicare administrators, should be the ones to decide the optimal mix of deductibles, coinsurance, and out-of-pocket maximums that best meets their needs and preferences under neutral incentives.

Appropriate Levels of Spending Over Time

If Medicare beneficiaries were to receive the marginal benefits of choosing a more efficient plan, the incentives to adopt costly new technologies would be improved over time. As noted earlier, costly new technologies are efficient if the value of the additional benefits from improved health exceed the additional costs of that technology. People may not be willing to spend a great deal of money on new treatments with very minor benefits. If Medicare beneficiaries were to receive the marginal benefits when selecting less technology-intensive plans that delivered higher value care at lower cost, the adoption of new technologies by health plans over time would be driven by whether new technology delivers substantial enough health benefits. As a result, consumers, rather than the government, would decide the extent to which health care spending should increase over time.

Conclusion

Medicare has significant long-term unfunded obligations. Although Social Security spending is currently much greater than Medicare spending, the unfunded obligation for Medicare is much greater than that for Social Security. Eliminating the projected 75-year actuarial deficit for Medicare Part A would require an immediate 3.51 percent increase in the HI payroll tax or a reduction in projected Medicare expenditures by 51 percent. Projected increases in Medicare Supplementary Medical Insurance (SMI) funding may appear less transparent because they are funded out of general revenues, but the economic significance of these obligations for Medicare SMI is just as great.

Policymakers face the challenge of reducing the growth of Medicare spending while preserving access to life-saving health care and the important financial protections that Medicare provides, and they cannot do so without ensuring that Medicare funds are spent more efficiently. Increases in Medicare spending over time are driven by an increasing population of aged Americans and increasing per-beneficiary spending on health care. While much of the increase in medical spending over time is driven by valuable new technologies, there also appear to be significant inefficiencies in the system. Therefore, future policies to control the growth in Medicare spending should target the sources of inefficient spending but not discourage the use medical care that is costly but delivers greater health benefits. This tension is the primary dilemma that policymakers face.

Policymakers may want to consider restructuring Medicare so that the direct spending by Medicare beneficiaries, in the form of premium contributions and out-of-pocket spending for medical care, yields a more efficient allocation of resources. Revising the Medicare fee-for-service program and the Medicare Advantage program to be more like Part D with a fixed-dollar subsidy provided toward the premium, has the potential for improving incentives for Medicare beneficiaries to consume optimal levels of care. When individuals receive the full benefits of selecting less expensive coverage, they will be more likely to select plans with optimal arrangements that balance both financial protection and technological adoption.

Catastrophe Risk Insurance

Insurance plays a vital role in America's economy by helping households and businesses manage risks. Individuals purchase insurance so they can sleep well at night; they gain comfort from the knowledge that they and their families are protected from some of the adverse effects of future events beyond their control. Businesses purchase insurance for much the same reason. It allows them to reduce the uncertainty associated with future costs and revenues, which enables them to plan for the future more effectively. Today, one can purchase insurance protection against a myriad of economic hazards, from poor health to motor vehicle accidents to legal liability to lightning strikes.

Insuring economic losses arising from large-scale natural and manmade catastrophes such as earthquakes, hurricanes, and terrorist attacks poses special challenges for the insurance industry and for Federal and State governments. This chapter examines the economics of catastrophe risk insurance. It draws the following main conclusions.

- In insurance markets, as in other markets, prices affect the way people weigh costs and benefits. Insurance prices that are artificially low can discourage people from adequately protecting against future losses. For example, subsidized property insurance prices may stimulate excessive building in high-risk areas, potentially driving up future government disaster relief spending.
- Government intervention in insurance markets can have unintended consequences such as limiting the availability of insurance offered by private firms.
- Private insurers manage catastrophe losses by being selective about which risks to insure, by designing insurance contracts to provide incentives for risk-reducing behavior, and by charging prices that are high enough to enable them to diversify risk over time or transfer risk to third parties. By adopting private sector risk management and pricing practices, government insurance programs could reduce the burden they impose on taxpayers and minimize negative effects on private insurance markets.

The Economics of Catastrophe Risk Insurance

In the United States, insurance is provided through a variety of private and public entities. Insurance companies owned by investors or policyholders sell insurance in the private sector. State-sponsored insurance pools have

characteristics of both private and public entities. They are typically owned by a group of private insurers, but they are governed under charters that grant them special rights and impose responsibilities not required of private insurers. Finally, the Federal Government operates at least 135 different programs that provide insurance-like benefits to individuals and businesses.

To understand how insurance works, imagine a large group of homeowners scattered throughout the country, each of whom faces a risk of property damage from a variety of identified hazards such as fire or severe weather. The likelihood that any particular member of the group will experience a loss is low, but the economic costs to that individual, should a loss occur, are significant. Each member of the group can reduce uncertainty about future economic losses by agreeing to pool risk with other members. One way of accomplishing this is through a mutual insurance agreement. At the beginning of the year, each member agrees to make a payment, called an *insurance premium,* into the pool. In exchange for their premiums, members are allowed to file claims with the pool should their houses incur damage from a covered hazard. Even if the insurance pool has no other resources, as long as the total value of premiums paid into the pool is at least as large as the value of insured losses over the year, all property damage will be fully covered. In this way, members of the pool gain security through diversification. Because any member's losses are paid for with premiums collected by all members, no member faces uncertainty about how much he will have to pay to cover property damage in the coming year.

The process of evaluating a risk exposure, determining whether or not to insure it, and setting terms and conditions for any insurance provided is called *underwriting*. Through underwriting, insurance providers seek to tie the premiums charged for insurance policies to the risks those policies cover. Effective underwriting serves an important social function, because when insurance prices accurately reflect underlying economic costs they can encourage a more efficient allocation of scarce resources. For example, suppose a member of a coastal community must decide where to build a new home. She may prefer to live as close to the ocean as possible, but a home located nearer the ocean may be exposed to a higher risk of damage from windstorms and flooding. If homeowners' insurance premiums are appropriately risk sensitive, then she will need to determine whether the benefits of living closer to the ocean are worth the cost of higher insurance premiums.

Underwriting is critical to the efficient functioning of insurance markets. In general, insurance markets function best under the following conditions:

1. Either all members of a pool face similar risks, or differences in risks can be observed and incorporated in insurance premiums.
2. Insurance does not dissuade those who are insured from avoiding risks.
3. The total value of insured losses for a pool can be forecast with precision.

In many insurance markets, one or both of the first two conditions may not hold. Violations of the third condition are a particular feature of catastrophe-risk insurance markets. Through effective underwriting, insurers can reduce, though perhaps not eliminate, problems that arise when these conditions fail to hold.

Effective Underwriting Reduces Information Problems

Insurance markets may fail to work effectively when differences in the risks faced by policyholders cannot be incorporated in insurance premiums. To see why, consider again the example of homeowners pooling risk. Suppose now that there are two types of homeowners: those who live in coastal areas that are at relatively high risk for windstorms and floods, and those who live in inland areas at lower risk for these hazards. If all homeowners were charged the same insurance premium, and if premiums were set equal to the average loss rate for all homes, then homeowners in inland regions would rightly feel that they were being overcharged. They face less risk from windstorms and floods than owners in coastal regions, yet they are asked to pay a premium equal to average losses for a pool that includes houses in both regions. Owners living in coastal areas would be attracted to the pool because it offers insurance at a premium that does not reflect their homes' higher risk. If the insurance policy were offered to all homeowners, a disproportionate share of those in coastal regions would accept the policy, while a disproportionate share of those living inland would seek insurance elsewhere or would choose to go without insurance. As a result, the average loss for those who chose to participate in the pool would be higher than the premium charged.

This example illustrates a general property of insurance contracts which economists call *adverse selection*. When premiums do not reflect differences in risk that are known to potential policyholders, insurance pools tend to attract members who are at greatest risk for the hazards covered. The solution to this problem is to charge policyholders with different risk exposures different premiums. In the example above, adverse selection could be avoided if homeowners in inland areas were charged lower premiums than those in coastal regions. Insurance providers generally try to set premiums commensurate with risk, but this is not always possible. In some cases it may simply be too costly for an insurance provider to identify differences in risk, but, as discussed later in this chapter, efforts by policymakers and insurance regulators to keep premiums for some high-risk policyholders low can also play a role.

Inefficiencies can also arise when insurance discourages those who are insured from taking actions to reduce potential losses. Consider the incentives faced by a homeowner thinking about how best to prepare for future windstorms. Many homeowners can reduce the damage caused by windstorms by installing storm shutters, but storm shutters are costly. If a homeowner is fully insured against the economic losses arising from future windstorms, she may

be less likely to purchase shutters. The tendency of those who are insured to work less hard to avoid losses is called *moral hazard*.

Insurance providers are well aware of the potential for moral hazard, and they attempt to address it through effective underwriting. Many insurance policies only cover losses in excess of a specified amount called a *deductible*, or they require that policyholders pay a fixed share of any losses incurred. By insuring some, but not all, economic losses, these types of policies strengthen policyholders' incentives to work to reduce the risks they face. Insurers may also require that specific action be taken as a precondition for receiving coverage, or they might provide pricing incentives for risk-reducing investments. For example, an insurer might refuse to cover windstorm risks for homes without storm shutters, or it might charge those homeowners a higher premium.

Catastrophe Losses Are Difficult to Forecast

Adverse selection and moral hazard problems are common in many insurance markets. Catastrophe risk insurers face an additional challenge, which arises from the fact that the total value of losses for a pool of insured properties or individuals is often exceptionally difficult to predict.

Forecasting annual losses from hazards like automobile accidents that only affect one or two members of a pool at a time is much easier than forecasting losses from large-scale catastrophes such as floods, hurricanes, or terrorist attacks. When the losses incurred by individual members of an insurance pool are more or less independent of one another, the average loss rate per policy is likely to be stable over time. Chart 5-1 illustrates this point by showing the annual nationwide accident rate per 100,000 registered passenger cars. While the accident rate has gradually declined over the past 15 years, it changes relatively little from year to year. It is difficult to predict whether any particular vehicle will be involved in an accident, but based on the data presented we can forecast with high confidence that about 4.5 percent of all passenger cars will be involved in some kind of accident over the next year. Because large-scale catastrophes have the potential to affect many members of an insurance pool simultaneously, spreading risk across a large number of members may not be sufficient to ensure that average losses per policy are stable over time. Compare Chart 5-1 with Chart 5-2. Chart 5-2 reports the number of loss claims filed per 100,000 homes and businesses insured for flood losses under the Federal Emergency Management Agency's National Flood Insurance Program (NFIP). Flood losses are not independent of one another; a single flood event can damage hundreds or even thousands of properties. Even though the NFIP insures a pool of millions of properties, the average loss rate per policy varies considerably from year to year.

Chart 5-1 Annual Accident Rate for U.S. Passenger Cars

Automobile accident rates have fallen over time, but change relatively little from year to year.

Number of vehicles involved in crashes per 100,000 registered vehicles
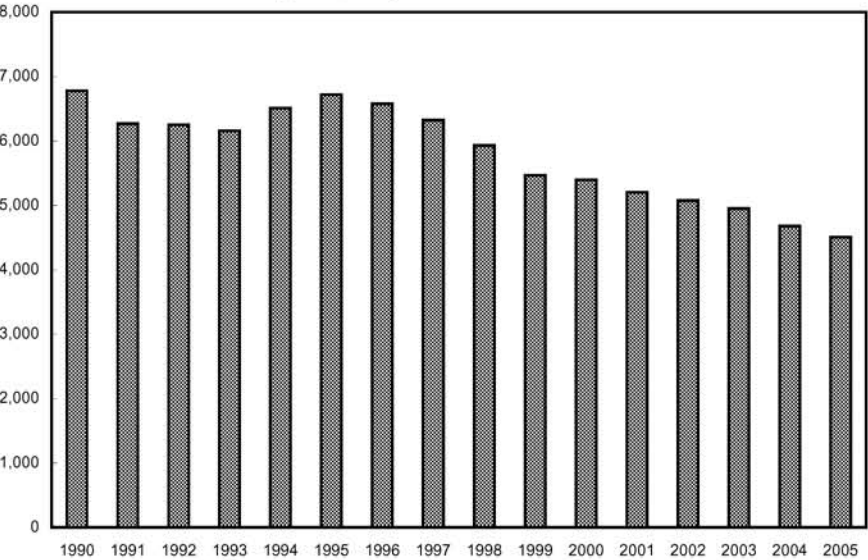
Source: National Highway Traffic Safety Administration.

Chart 5-2 Annual Claim Rate for Properties Covered by the National Flood Insurance Program

Flood-loss claim rates vary considerably from year to year.

Number of claims per 100,000 insured properties
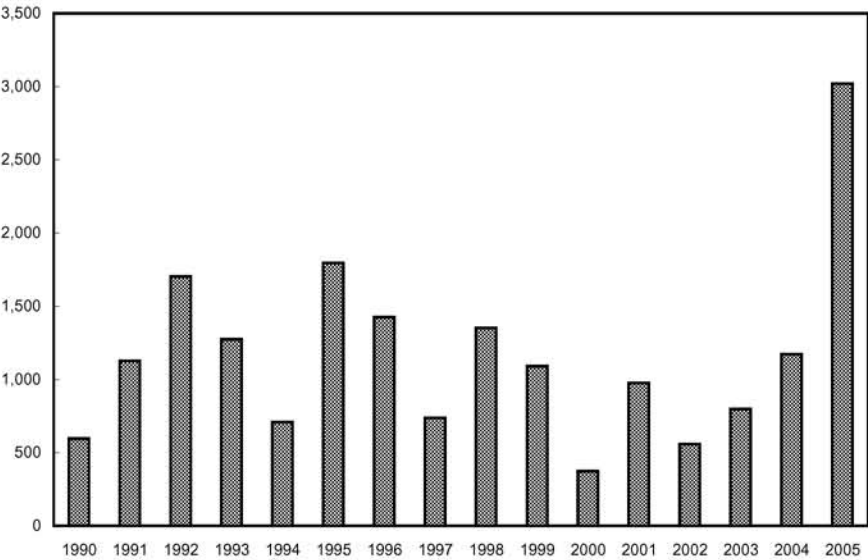
Source: Federal Emergency Management Agency.

In some catastrophe-risk insurance markets, forecast accuracy also suffers from a lack of relevant historical data and experience. This is a particular problem when catastrophes are rare, and when the character of those events is likely to change over time. For example, U.S. commercial property and casualty insurers had almost no experience forecasting losses from large-scale terrorist attacks prior to September 11, 2001. A recent report by the President's Working Group on Financial Markets on the availability and affordability of insurance for terrorism risk found that while modeling of terrorism risk has improved since 2001, insurers continue to have limited confidence in the models they use for evaluating this risk exposure.

When annual losses for a pool can be forecast with reasonably high precision, it is relatively easy for an insurance provider to manage risk. As long as its underwriting procedures ensure that the average premium paid by members of the pool is at least as large as the average loss rate per member, it is likely that in any given year total premium revenues for the pool will be sufficient to pay all claims. If, as in our automobile accident example, losses are independent across members of a pool, increasing the size of the pool actually makes it easier for an insurer to manage risk, because the more members that are included in the pool, the more stable will be the average loss rate per member.

Losses from catastrophes are not independent across exposures, and therefore they are much more difficult to manage. A severe hurricane, for example, can cause damage over tens of thousands of square miles, so even if an insurer provides windstorm coverage for properties scattered throughout a state, average losses per property are likely to be exceptionally high in hurricane years. Since catastrophes are infrequent but costly, annual premium revenues for a pool of exposures that exceed the value of claims in most years may not be sufficient to pay all claims in those rare years when a severe event occurs. Insurance providers work to address this problem by pooling risk across time or by diversifying the risk exposure more broadly by sharing it with other insurers.

Managing Catastrophe Losses

One way to manage the financial risk of insuring catastrophe hazards is to retain a portion of excess premium revenues collected in years when losses are low to pay claims in years when catastrophes generate large losses. Equity capital set aside to pay potential claims is called *surplus*. In practice, building surplus large enough to pay catastrophe losses can be difficult for private insurance companies. Owners of insurance companies expect to earn a market rate of return on their equity investments, including equity held as surplus to cover future claims. Moreover, income flowing from insurance company assets is subject to corporate income tax that effectively adds to the cost of accumulating and holding surplus.

An alternative to using surplus to cover catastrophe losses is to transfer risk to third parties. Some insurers transfer risk directly to capital market participants such as hedge funds and institutional investors (Box 5-1). More commonly, insurers negotiate risk-sharing agreements with specialized insurance companies called *reinsurers*. Reinsurers are internationally diversified companies that make a business of selling insurance to primary insurers. In a typical reinsurance arrangement, a primary insurer pays a fee to a reinsurance company that agrees to cover some of the insurer's costs in the event that claims exceed a prespecified threshold. In essence, reinsurance arrangements work much like other types of insurance. Through reinsurance a primary insurer subject to the risk of high claims caused by a catastrophe can pool its risk with other primary insurers that are exposed to different hazards. As with other types of insurance, problems of adverse selection and moral hazard can impede the efficient functioning of reinsurance markets.

Box 5-1: Catastrophe Bonds and Sidecars—Accessing Financial Markets to Better Manage Catastrophe Risks

Though reinsurance agreements between primary insurers and specialized reinsurance companies remain the most popular method for transferring and pooling risks posed by large-scale catastrophes, the capital available to reinsurers is only a tiny fraction of the total capital invested in financial markets. By one estimate, reinsurance companies worldwide had accumulated about $400 billion in shareholder funds by year-end 2005, which is only about 1 percent of the market capitalization of the world's public equity markets. To spread catastrophe risks more broadly, financial markets have developed mechanisms to allow investors who do not directly hold shares in insurance companies to assume some of the catastrophe risk exposure of primary insurers or reinsurers in exchange for an appropriate investment return. Two notable examples are catastrophe bonds and "sidecars."

Catastrophe bonds (CAT bonds), also called "acts of God" bonds, are risk-linked securities that offer a return to investors similar to that on high-yield corporate junk bonds. In a typical CAT bond transaction, a firm that wants to transfer some risk to outside investors issues a bond and invests the proceeds in safe securities. If a specified catastrophe event occurs, the proceeds from the bond issue are released to the issuer. If no event occurs during the term of the bond, the principal is returned to investors. Payouts from CAT bonds are often tied to industry-wide loss estimates or defined catastrophe events such as whether or not a hurricane makes landfall on a particular stretch of coastline. Because these types of events are presumably beyond the

---

Box 5-1 — *continued*

control of the bond issuer, investors are protected from moral hazard. A drawback of these types of CAT bonds, however, is that they do not protect the issuer against all possible catastrophe losses. For example, an insurer that issues a bond with a payout tied to a hurricane event could be exposed to large losses from a tropical storm that does not meet the definition of a hurricane. The market for CAT bonds has grown rapidly over the past decade, though the value of bonds outstanding remains small relative to the value of insured losses in recent catastrophe events. About $4.9 billion in CAT bond capital was outstanding as of year-end 2005, a 21 percent increase over the 2004 level.

Sidecars provide an increasingly popular alternative to CAT bonds. A sidecar is a special-purpose financial entity, usually designed to last 2 to 3 years. Under a sidecar arrangement, a group of investors partners with an existing reinsurance company: the investors provide the necessary funds for deployment and the reinsurance company contributes its infrastructure, business relationships, and the skills of its staff. Sidecar investors receive a portion of the reinsurance company's premium revenue from a particular reinsurance contract or line of business, and the reinsurer gains access to the investors' capital to cover potential catastrophe losses. Through sidecars, investors can decide to assume particular catastrophe risks without being exposed to all of the risks covered by a given reinsurance company. Sidecars have helped Bermuda-based reinsurance companies to expand their capacity to cover catastrophe risk exposures in the United States despite incurring significant losses in 2005. About $2.5 billion in capital was reportedly raised through sidecars organized with Bermuda reinsurers from December 2005 to June 2006.

Through CAT bonds, sidecars, and other innovative financing mechanisms, insurers and private investors are finding new ways to spread the risks posed by large-scale catastrophes. These financing mechanisms currently contribute only a relatively small share of the total capital available to cover catastrophe losses, but the volume of capital they have raised has grown rapidly in recent years. It is likely that as these markets mature, the base of investors willing to bear some catastrophe risk will continue to expand, ultimately lowering the cost of insuring catastrophe risks.

What happens if an insurance provider lacks the resources to pay claims following a catastrophe? Private-sector insurance companies that cannot afford to pay claims are usually forced into receivership. In contrast, many government-sponsored insurers can raise additional funds to pay claims after an event has occurred. Government-sponsored insurance programs often do not face the same financial constraints as private insurers because they have special rights to compel third parties such as taxpayers or private insurers to bear a portion of their financial risk. The NFIP, for example, is authorized by Congress to borrow from the U.S. Treasury, which increases taxpayer liabilities, and the Federal Government's terrorism-risk insurance program and several State-sponsored catastrophe insurance providers are empowered to levy surcharges on policies sold by private insurers.

Federal Catastrophe Insurance Programs

In 1803, Congress passed a law granting the victims of a fire in Portsmouth, New Hampshire, extra time to repay certain debts owed to the Federal Government. Though the Federal Government has assisted Americans harmed by disasters throughout the Nation's history, prior to the mid-twentieth century aid was generally provided on an ad hoc basis; a disaster would strike and Congress would then determine whether and to what extent Federal aid would be provided. Acts of Congress passed in 1947 and 1950 regularized the process by which the Federal Government extends assistance to disaster-affected communities and additional legislation enacted since then has clarified and expanded the Government's role in disaster relief.

One problem with a variety of government relief efforts is that they can make it more difficult for private insurers to sell policies for some catastrophe hazards at prices commensurate with underlying risks. People have less incentive to pay sometimes high insurance premiums if they expect to receive aid from the government when a catastrophe strikes. Policymakers have sought to address this moral hazard problem in several different ways. The Federal Government provides insurance coverage for certain catastrophe hazards, often at prices lower than those that would be charged by private insurers. In addition, in some cases the Government requires that individuals purchase insurance policies or mandates that private insurers offer policies for sale.

The National Flood Insurance Program

The National Flood Insurance Program (NFIP) was established in 1968 to make flood insurance more widely available to homeowners and businesses, to encourage local communities to prepare better for flood hazards, and to reduce reliance on direct Federal disaster relief following floods. The NFIP

currently provides flood insurance for 5.3 million policyholders nationwide, many of whom might not be able to obtain coverage without the program. Residential and commercial property owners in some 20,000 participating communities are eligible to purchase flood insurance policies under the program. Homeowners with mortgages issued by federally regulated lenders on property in communities identified to be in flood hazard areas are required to purchase flood insurance on their dwellings. Property owners can purchase policies either directly from the Federal Government or, more commonly, through local insurance companies who sell NFIP policies under their own name but pass their risk on to the Government. Whether policies are sold directly by the Federal Government or by insurance companies, the NFIP receives premium payments for the policies and bears all financial risks associated with the insurance they provide. The program is administered by the Federal Emergency Management Agency (FEMA).

FEMA relies on Flood Insurance Rate Maps (FIRMs) when underwriting flood insurance. These maps identify areas within a community that have at least a 1-percent chance per year of being inundated by high water. These areas are called 100-year floodplains. Federal flood insurance is only made available in local communities that agree to adopt zoning ordinances, building codes, and other planning measures designed to reduce future damage caused by floods. For example, communities must require that new buildings be elevated above the level that flood waters are expected to reach on average once per 100 years. According to FEMA, buildings that meet its floodplain management standards suffer 80 percent less damage from floods each year than those that do not. Not all structures insured under the NFIP meet these standards, however; structures completed prior to a community's decision to participate in the program or prior to the publication of a community's FIRM are eligible for insurance under the program even if they do not meet FEMA standards.

The NFIP charges different premiums for different properties. A structure built or substantially renovated after 1974 or after a community's FIRM was completed (whichever is later) is charged an *actuarially fair* annual premium equal to an estimate of expected annual claims under the property's flood insurance policy. Policyholders who pay actuarially fair premiums year after year should, in the long run, end up paying premiums that are just sufficient to cover their claims on average. About one-quarter of NFIP policies cover properties built prior to 1974 or prior to the publication of a community's FIRM. By law, these "pre-FIRM" properties are charged subsidized premiums. Pre-FIRM properties are much less likely to comply with modern flood risk mitigation standards since most were built before such standards were widely applied. Because of their higher risk, pre-FIRM properties are assessed higher premiums on average than newer properties, but even these

higher premiums are not adequate to cover expected losses. On average, premiums for pre-FIRM properties represent only about 40 percent of those properties' actuarially fair rates.

Not surprisingly, the NFIP pricing scheme has led to serious adverse selection and moral hazard problems. On the one hand, FEMA estimates that one-half to two-thirds of structures in floodplains do not carry flood insurance. On the other hand, some exceptionally high-risk properties continue to receive NFIP coverage at subsidized rates even though they have been damaged by floods multiple times since entering the program. Some 50,644 properties insured by the NFIP as of September 30, 2004 had incurred flood damage resulting in claims of at least $1,000 more than once during a 10-year period. While these properties only represented about 1 percent of all structures then insured under the program, repetitive-loss properties have historically accounted for 38 percent of all program claims payments. Amendments to the Flood Insurance Act passed in 2004 authorized a pilot program to remove some of the most severe repetitive-loss properties from the NFIP insurance roll by allowing FEMA to fund work to elevate or relocate some of them or, in extreme cases, to purchase and demolish them.

The NFIP illustrates how underwriting standards can either enhance or impede loss mitigation. By providing coverage only in communities that agree to adopt flood-risk mitigation measures, the NFIP may have induced some communities to take steps that FEMA credits with reducing flood damage by an average $1.2 billion annually. At the same time, by providing insurance to pre-FIRM properties at less than actuarially fair rates, the program may have discouraged some policyholders from relocating or renovating structures at high risk for flood damage. The availability of flood insurance has lowered the risk to banks of financing real-estate investment in locations vulnerable to flood losses. As a result, it is not clear whether the NFIP has reduced the size of Federal appropriations for flood disaster relief as intended. Demand for Federal disaster aid may arguably be higher than it would have been had the NFIP not facilitated development in high-risk areas.

Chart 5-3 shows that since 1986 NFIP premiums exceeded annual losses in most years, but were woefully inadequate to cover losses from Hurricanes Katrina, Rita, and Wilma in 2005. The 2005 hurricanes resulted in about $16.3 billion in NFIP program claims, some of which were not paid until 2006. Even so, claims paid in 2005 exceeded premiums collected in that year by a factor of nearly six to one. Unlike private sector insurers, who would need to accumulate surplus or purchase reinsurance to pay claims in excess of premiums, the NFIP is permitted to borrow from the Federal Government. As of August 2005, just before Hurricane Katrina struck, the NFIP had accumulated a relatively modest $300 million in debt owed to the U.S. Treasury, but the program will need to borrow an additional $21.2 billion to pay claims

filed in 2005. Though the NFIP is supposed to repay this debt using future premium revenue, it is unlikely that this will be possible. The Congressional Budget Office estimates that by 2007 the interest on NFIP debt will grow to about $1 billion annually, which is about 40 percent of the projected annual premium revenue. Even if future hurricane seasons are milder than those experienced in recent years, projected premiums are not expected to be large enough to cover both the interest on the outstanding debt and the projected future claims. The NFIP's current dire financial situation amply demonstrates that in insurance, as elsewhere, there is no free lunch. Annual premium revenue from the NFIP was able to cover losses in most of the program's recent history, but the subsidized insurance program exposed the American taxpayers to a huge potential financial liability which became an actual liability in 2005.

Terrorism and War-Risk Insurance Programs

The Federal Government provided billions of dollars in disaster assistance following the September 11, 2001 terrorist attacks on New York and Washington, DC, including about $4 billion in aid to the airline industry and about $20 billion in aid to the New York City area. To date, about $36 billion

Chart 5-3 **National Flood Insurance Program Annual Premiums and Losses**
Annual premiums were sufficient to cover annual losses in most years, but were woefully inadequate in 2005.

Billions of dollars (real)
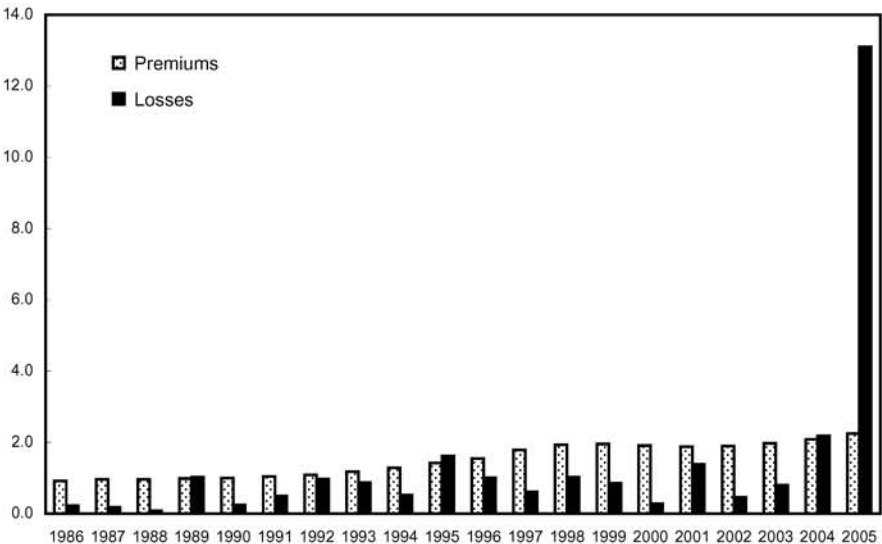
Source: Federal Emergency Management Agency.

in loss claims have been paid by private insurers. Though insured losses represented only a fraction of the total economic costs of the September 11 attacks, they were far greater than those arising from any prior terrorist event.

Following September 11, commercial property and casualty insurers reevaluated their policyholders' exposure to risk from possible future attacks. Many insurers canceled policies, began explicitly excluding coverage for terrorist attacks from new policies, or increased premiums charged to policyholders. In response to what was believed to be a temporary contraction in the supply of insurance available for terrorism risk, the Administration and Congress undertook measures to ensure that the airline and commercial real estate sectors would not be adversely affected.

Less than two weeks after the September 11 attacks, the Federal Aviation Administration (FAA) began selling insurance policies directly to U.S. airlines to cover third-party liability (e.g., harm to individuals or property on the ground) arising from acts of war or terrorism, and in November of 2002 the Homeland Security Act expanded this program to provide insurance coverage for loss of aircraft and airline passenger liability as well. The program has been reauthorized several times since its inception and it remains in effect today. As of October 1, 2006, policies under this program provided 75 airlines with insurance coverage for potential losses ranging from $100 million to $4 billion each.

The Terrorism Risk Insurance Act (TRIA) passed in November of 2002 established a second, much broader, Federal program to encourage private-sector commercial property and casualty insurers to provide terrorism risk coverage. The program was originally designed to expire after three years, but in 2005 Congress elected to extend the program with some modifications through 2007.

TRIA has two main components. First, it mandates that insurance companies that sell commercial property and casualty insurance make available to customers policies that do not explicitly exclude coverage for losses caused by acts of terrorism. Insurers may exclude losses on other grounds, however, so not all losses arising from terrorist attacks must be covered. According to the President's Working Group on Financial Markets, commercial insurance policies generally do not cover losses arising from chemical, nuclear, biological, and radiological events, whether or not these events are caused by acts of terrorism. Second, TRIA authorizes the Treasury Department to provide reinsurance to cover a portion of insurance loss claims arising from certified acts of international terrorism against U.S. targets. Under the reinsurance program, a primary insurer must cover 100 percent of its loss claims up to a specified deductible. The Federal Government then pays a fixed share of losses in excess of the deductible. For 2007 an insurance company is required to cover all losses up to 20 percent of its prior year's premiums on qualifying

lines of business and 15 percent of losses above this deductible. TRIA imposes a cap of $100 billion on total insurer losses from terrorist attacks. Under the statute, Congress would determine the procedures to govern any payments for losses beyond $100 billion in separate legislation.

Since 2001, no claims have been filed under either the FAA's aviation war-risk insurance program or the Treasury Department's terrorism-risk reinsurance program, but, like the NFIP, both of these programs expose U.S. taxpayers to large potential losses. Because they were intended to be temporary, neither program is designed to ensure that premiums will be sufficient to pay future claims. Premium revenue collected under the aviation war-risk program is subject to a cap mandated by Congress. As a result, premiums charged by the FAA are significantly lower than those that would be charged for comparable policies sold by private-sector aviation insurers. Airlines pay a total of about $160 million in premiums to the FAA each year; by one estimate, without the program these airlines would need to pay $500 million annually in premiums to private insurers. TRIA does not require property and casualty insurers to pay *any* premiums for the reinsurance protection they receive. Instead, claims under the program are expected to be paid with Federal outlays and then recouped, after the fact, through surcharges levied on future premiums for property and casualty insurance policies. Given that the program was established in part to address problems arising from high insurance premiums following the September 11, 2001 attacks, there are real questions as to whether surcharges would be set high enough to recoup expenditures following a future terrorist attack. Any surcharges would likely be spread over several years to reduce the impact on premiums, and since the Treasury Department is only required by law to recoup up to $27.5 billion, there is no guarantee that the full costs of the program would ultimately be recovered.

State Property Insurance Markets

Although the Federal Government is actively involved in insuring risks from floods and terrorist attacks, most homeowners and businesses look first to their local property insurers to obtain financial protection against a variety of hazards including potential catastrophes. State governments are responsible for regulating insurance markets. Though laws differ from state to state, all states' insurance regulators exercise some control over who is permitted to sell insurance, what terms and conditions can be attached to insurance policies, and how much insurers can charge. Insurance regulations are intended to protect consumers who may have difficulty evaluating complex insurance contracts and to ensure that insurers maintain sufficient financial resources to pay future claims. While regulation plays an important role in protecting

consumers from fraud and poor risk management practices, poorly conceived and executed regulation can create long-term problems for the operation of state catastrophe-risk insurance markets.

Every state regulates property insurance premiums charged to homeowners and small businesses. Many states require that premiums be approved in advance by regulators. Others allow insurance regulators to review existing price schedules and empower regulators to force companies to reimburse policyholders when premiums are found to be excessive. Rate regulations can make it difficult for insurance companies to set premiums that accurately reflect available information about risks, which can exacerbate moral hazard and adverse selection problems. In some states the rate review and approval process can take many months, so insurers cannot rapidly adjust premiums when new information becomes available. The rate review process may also discourage insurance companies from proposing complex pricing plans which, though difficult to explain and justify to state rate boards, more accurately reflect detailed information about the risks associated with individual insurance policies.

Efforts by regulators to keep property insurance prices artificially low can make it difficult for individuals and businesses to obtain insurance on private markets at any price. To ensure that they will be able to pay claims after a catastrophe, private insurers need to set premiums high enough to enable them to build surplus or transfer risk to reinsurers. If regulators do not allow insurers to charge rates sufficient to accomplish these tasks, the insurers will be discouraged from taking on catastrophe risks. They may choose to sell insurance only in areas at low risk for catastrophe hazards, or they may seek to exclude coverage for such hazards under the terms of the property insurance policies they offer. Regulation can also deter insurers from competing for customers, thereby reducing the range and quality of insurance options available.

Many states that face risks from hurricanes or earthquakes have established special entities to provide insurance to those who cannot obtain coverage from private insurers. In 1996, California established a quasi-public company, the California Earthquake Authority, to sell earthquake insurance policies to California residents, backed by funds contributed by a number of private insurers operating in the state. Several states maintain residual pools to cover windstorm risks. These pools operate like traditional insurance companies, but they are required to sell policies to property owners in high-risk coastal areas and they are empowered to levy surcharges on primary insurers operating in a state.

Some state-sponsored insurance programs use complicated procedures for setting premiums, and many claim to charge premiums that are actuarially fair, but they all have one thing in common: they provide insurance only to policyholders who either will not, or cannot, obtain insurance from the

Box 5-2: Gulf Coast Property Insurance Markets After Hurricanes Katrina, Rita, and Wilma

2005 was a terrible year for communities located along the U.S. Gulf Coast. Hurricane Katrina devastated a land area the size of Great Britain and displaced more than 270,000 people. The total value of property damage and business interruption caused by Hurricane Katrina has been estimated at $135 billion. Hurricane Katrina was followed a few weeks later by Hurricane Rita, which caused an estimated $15 billion in damage, and Hurricane Wilma, which caused an estimated $20 billion in damage. The President and Congress responded by appropriating about $110 billion for disaster relief and recovery aid to affected communities. Property insurers have also played an important role in recovery efforts by paying billions of dollars of loss claims, but there are concerns that rising insurance premiums for coastal properties may be a barrier to redevelopment. The response of property insurance markets to the unprecedented losses caused by the 2005 hurricane season underscores the role of effective underwriting in managing catastrophe risks.

Hurricanes Katrina, Rita, and Wilma resulted in an estimated $57 billion in insured property damages, not including claims filed with the National Flood Insurance Program. Despite bearing enormous losses, most private-sector primary insurers operating in the Gulf Coast emerged from the 2005 hurricane season in reasonably sound financial condition. At least four primary insurers failed as a result of the 2005 storms, but the share of property and casualty insurers listed as financially impaired by a major insurance company rating agency actually dropped to a 25-year low while the aggregate value of surplus available to insurers for paying future claims increased. Primary insurers fared well as a group in part because they had transferred a significant share of their catastrophe risk exposure to reinsurers. According to one industry association, reinsurance covered about 60 percent of 2005 insured hurricane losses.

Though the U.S. property and casualty insurance sector as a whole remains healthy, property insurance markets in several coastal states are under stress. Information collected during the 2004 and 2005 hurricane seasons revealed deficiencies in industry-standard catastrophe risk models used in underwriting property insurance. These models are now being adapted to reflect expectations of more violent hurricane seasons, revised analysis of the costs of repairing property damage following major catastrophes, new findings about the effects of hurricane-generated storm surges, and other factors. As a result, primary insurers and reinsurers are increasing their estimates of probable losses on windstorm policies in areas at risk for hurricanes. A leading

catastrophe-risk modeling firm reports that revised forecasts of the severity of Atlantic hurricane seasons alone will increase estimates of loss rates from future hurricanes in the Gulf Coast and southeastern U.S. by 50 percent.

As assessments of the potential costs of future hurricanes have increased, primary insurers and reinsurers have sought to limit their exposure to windstorm hazards and increase the premiums charged for insuring this hazard. Reinsurance companies, many of whom lost capital in 2005 to hurricane-related claims, have significantly increased premiums. Unlike reinsurance premiums, premiums charged by primary insurers for homeowners' and commercial property policies are regulated by state insurance commissions. Primary insurers have petitioned state regulators to allow them to raise premiums to cover rising reinsurance costs and to more closely reflect new information on the risks posed by windstorms. Where possible, some insurers have also attempted to reduce their exposure to windstorm hazards by refusing to renew existing policies in high-risk areas or by adding conditions to policies that exclude coverage of windstorm damage. In several states, government-sponsored insurance programs that are required to provide windstorm coverage to property owners who are unable to obtain insurance through the private sector have grown dramatically.

Recent developments in coastal property insurance markets have the potential to discourage some investment in areas at high risk for hurricanes, since property owners in these areas will likely have to pay higher insurance premiums or bear greater risk than in the past. For this reason, some have argued that Federal and State governments should take action to ensure that insurance for windstorm coverage in hurricane-prone regions is widely available and that the premiums charged for this insurance are relatively low. However, as discussed in the text, efforts to keep premiums for windstorm insurance artificially low may discourage property owners from taking action to lessen future windstorm losses while potentially encouraging excessive development in high-risk areas.

private market. These programs tend to attract exactly those members whose high risk makes them unattractive to private insurers. For example, in some states, residual pools are the main providers of windstorm insurance for homeowners in coastal areas exposed to high risk from hurricanes.

In recent years a number of state-sponsored insurance programs have had difficulty paying claims following major catastrophes. Different states have dealt with this problem in different ways. A few states have used government money to provide new funds for insolvent programs, thereby passing the cost

of covering losses on to taxpayers. More commonly, states have levied surcharges on premiums for policies sold by private insurers. This approach effectively forces property owners in relatively low-risk areas who can obtain insurance from private providers to pay higher premiums to cover insured losses for property owners in higher risk areas who obtain insurance through the residual pool. By effectively raising the cost of insurance in the private market, these surcharges may actually encourage more property owners to seek insurance from the residual pool so that the pool is exposed to even higher losses the next time a catastrophe strikes.

Since people consider the cost of property insurance when deciding where to live and conduct business, the use of rate regulations or state-sponsored insurance programs to keep property insurance prices in high-risk areas artificially low can have significant negative consequences. All else equal, commercial and residential development will tend to be greater in those areas where insurance prices are lower. As a result, artificially low premiums for catastrophe risk insurance can lead to excessive development in catastrophe-prone areas, putting lives and property in harm's way.

Conclusion

All insurance markets are susceptible to problems arising from adverse selection and moral hazard, but insurers of catastrophe risks must also deal with the fact that total insured losses are difficult to predict and are potentially quite large. While it may not be possible to eliminate these problems, their effects can be moderated through prudent underwriting. Adverse selection and moral hazard problems can be lessened by being selective about which risks to insure, by setting premiums to match observable differences in risk, and by requiring policyholders to bear a share of the financial risk posed by the hazards they are insured against. Insurance providers deal with uncertain losses by charging premiums that are high enough to enable them to build surplus and/or transfer excess risk to third parties such as reinsurers.

Regulations that constrain private insurers' underwriting flexibility can undermine their ability to provide insurance coverage for catastrophe risks. Government-sponsored insurance programs that can borrow from the U.S. Treasury or levy surcharges to pay claims after a catastrophe has occurred do not face the same financial constraints as private insurers. Nonetheless, government programs that do not apply prudent underwriting standards expose taxpayers to large liabilities.

Effective insurance underwriting serves an important social function by tying the premiums and terms of insurance policies to the risks covered. When insurance prices reflect underlying economic costs they can encourage a more

efficient allocation of resources. Efforts to keep premiums for insurance against catastrophe hazards artificially low, whether through regulation or through subsidized government programs, can encourage excessively risky behavior on the part of those who might be affected by future catastrophes.

CHAPTER 6

The Transportation Sector:
Energy and Infrastructure Use

Energy is a $1 trillion industry representing 8 percent of the U.S. economy. The two biggest consumers of energy from fossil and renewable fuels are electric power and transportation. While electricity can be generated from diverse sources—coal, nuclear fission, natural gas, water, petroleum, and increasingly, wind and sun—98 percent of transportation, whether by plane, train, ship or automobile, is currently powered by petroleum. The transportation sector alone accounts for two-thirds of the petroleum consumed in the United States. Thus, key to understanding the transportation sector is understanding the petroleum market, and the ways in which consumers and firms in the transportation sector respond to changes in world oil prices.

The lack of substitutes for oil means that in the short run, oil consumption in transportation is particularly unresponsive to price changes. This makes the economy vulnerable to sudden increases in oil prices. Perhaps more importantly, the world's reliance on oil creates an external cost in terms of national security.

In addition to petroleum, the transportation sector relies on infrastructure. The United States has close to 4 million miles of roads, bridges, and highways to support a wide variety of economic and social activity. Over time, however, demands on this infrastructure have outstripped its capacity. While the miles of urban roadways built have increased by nearly 60 percent since 1980, vehicle miles traveled on urban roadways increased by double that amount. The primary reason for this shortfall is that a well-functioning market that puts a price on roadway use is largely nonexistent. As a result, traffic in most metropolitan areas has become increasingly congested, costing both time and fuel. In 2003 alone, Americans were delayed about 3.7 billion hours and used 2.3 billion extra gallons of fuel (47 hours and 29 gallons per rush-hour commuter) in stop-and-go traffic. Like the costs exacted by oil use on national security and the environment, the full costs of congestion are not taken into account by individuals when they drive: each driver usually decides when and where to drive based on his or her own private needs and ignores the costs imposed on others.

This chapter discusses several developments in the use of energy and infrastructure for transportation, and reviews strategies that have been used to reduce oil use and better manage the existing infrastructure. Key points in this chapter are:

- Recent increases in the price of oil and the external costs of oil have led to renewed interest by markets and governments in the development of new alternatives. Government can play a role in ensuring that external costs are taken into account by markets, but ultimately markets are best suited to decide how to respond.
- Cars and light trucks are the largest users of petroleum. As a result, the fuel economy of the vehicles purchased and the number of miles that they are driven have a large effect on oil consumption.
- Congestion is a growing problem in American urban areas. Cities and states have shown a growing interest in and capacity for setting prices for road use during peak periods to reduce the full economic costs of congestion.

Fuel Markets and the Transportation Sector

Over the past 15 years, petroleum use in the industrial, utility, and building sectors has been relatively flat, while petroleum use by the transportation sector has grown by 27 percent. This trend is expected to continue. While new, more energy-efficient technology has reduced the energy needs of most sectors, gains in vehicle engine efficiency have been more than offset by a shift to heavier, more powerful cars and light trucks, and increases in driving.

Cars and light trucks accounted for 92 percent of U.S. roadway travel in 2006 and account for 62 percent of petroleum devoted to transport. Department of Energy projections suggest that these modes of transportation will continue to be important, and that light truck usage will show significant growth in the years to come (see Chart 6-1). Heavy trucks consume almost 17 percent of the petroleum used for transport. Air, rail, marine, and off-road vehicles currently account for the remaining 21 percent. Air travel is one of the fastest growing modes of transportation. Energy consumption for air travel is projected to increase nearly 46 percent by 2030, or about 620,000 more barrels of oil per day.

Chart 6-1 Historical and Projected U.S Oil Consumption and Production

Growth in petroleum use is projected to continue unabated in the foreseeable future. Cars and light trucks make up the bulk of U.S. oil use, while air oil use is one of the fastest growing.

Barrels/day (millions)
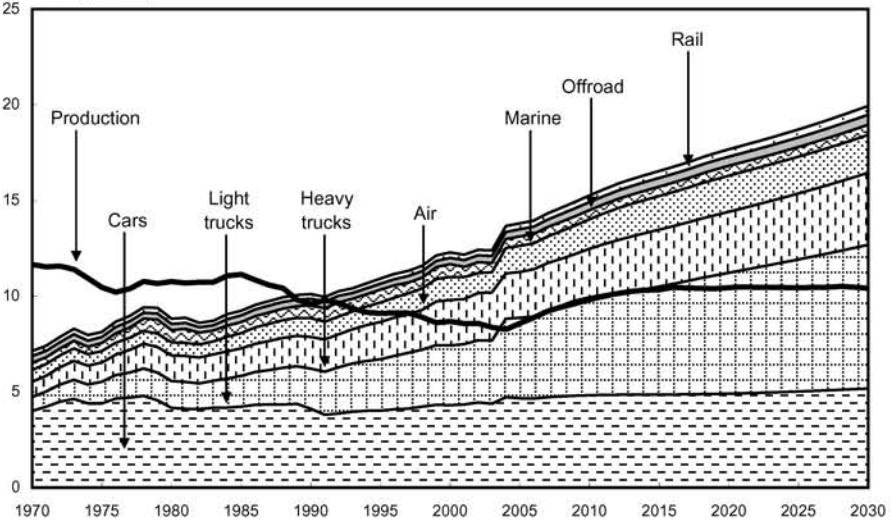
Note: The sharp increase in values between 2003 and 2004 is due to the change from historical to projected values.
Source: Department of Energy (Oak Ridge National Laboratory and Energy Information Administration).

Responding to Changes in the Price of Oil

In well-functioning markets, the price of a good or service reflects all of the associated costs and benefits—for example, the costs incurred in extracting, transporting, and refining the oil, or the benefits from using gasoline to drive. The market then uses price to achieve the most efficient level of production and consumption. Transportation has largely reacted to changes in energy markets in this way.

High demand for oil, due in part to rapid economic growth in China and India, has helped push oil prices to record levels. The real average monthly price of oil to the refiner was $26 between 1986 and 2004 (see Chart 6-2, in 2006 U.S. dollars). In 2004, the price to the refiner began to climb, approaching $70 per barrel in 2006 (other oil price measures were higher). For the transportation sector, this is a significant increase in the cost of one of its primary inputs. Normally, as the price of a good rises, consumers reduce how much they use. However, it typically takes years before the transportation sector's consumption of oil is substantially reduced, in part due to the lack of easily available substitutes. Eventually, though, consumers do react to high prices. For instance, hybrid vehicle sales have tripled since 2004, while light truck sales have fallen by 16 percent.

Chart 6-2 **The Real Price of Imported Crude Oil**
Oil prices fluctuate over time, but current prices are above the historical average.

2006 dollars per barrel

Projections
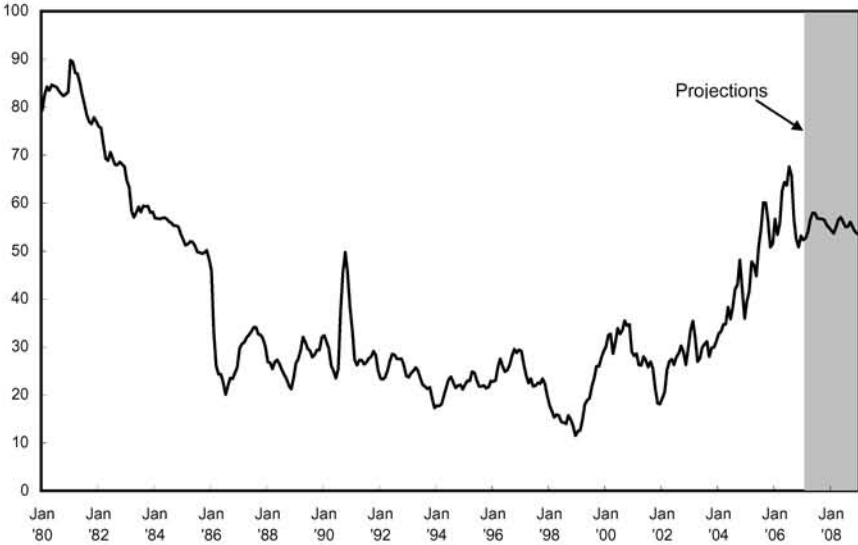
Note: Imported Refiner Acquisition Cost (RAC) Crude Oil.
Source: Department of Energy (Energy Information Administration) .

    When high oil prices are sustained, as has been the case recently, the market
shows renewed interest in investing in new technologies for developing alter-
natives to oil and improving vehicle fuel economy. Such research and
development investments tend to recede when oil prices fall. During the
period of high oil prices in the late 1970s and early 1980s, the private sector
invested billions of dollars in energy research and development before the
price of oil declined. A recent study finds that private investment in alterna-
tive fuel technologies again has increased in response to higher oil prices,
doubling between 2004 and 2006, constituting 10 percent of the total invest-
ment in energy. Because of the transportation sector's delayed response to oil
prices, these increases are likely to continue for some time.
    The lack of alternatives to oil also means that sudden major oil supply
changes—such as when oil production in an entire region is unexpectedly
shut down—can lead to large and sudden price increases in the months
following the shock. Since oil trades in a global market, the impact on the
economy from such shocks does not depend on how much we import, only
on how much we consume, and our consumption has been growing. The
market has adapted to this threat by investing in more energy-efficient modes
of production, investing in alternative energy sources, and increasing holdings
of private oil inventories.

External Costs of Oil Use

Prices determine which goods and services are produced in the marketplace. In the absence of government policy (such as taxes or regulations), the price of a good or service accounts for all private costs incurred by those who have produced or purchased the product. In the case of oil, this includes everyone from the oil company that extracts the oil, to the shipper, refiner, retailer, and driver who fuels her car. In the case of oil, the price reflects most of the costs, but there are some costs to society that remain unaccounted for.

Eighty-one percent of the world's remaining proven petroleum reserves are currently controlled by members of the Organization of Petroleum Exporting Countries (OPEC) (including Iran and Venezuela) and Russia, and nearly all of these reserves are controlled by national oil firms. Since oil trades in a world market, oil consumption anywhere in the world affects the price of oil for Americans. The importance of oil to the world economy gives the major oil-producing countries disproportionate diplomatic leverage in world affairs. Oil resources can also fuel corruption in developing countries. Air pollutants and carbon dioxide from burning gasoline also contribute to concerns about air quality, human health, and climate.

The purchase of a gallon of gasoline imposes these national security and environmental costs on everyone, not just on the buyer and seller. Though State and Federal gasoline and diesel fuel taxes and regulations help account for these other costs, many studies suggest that the total external costs of oil may be higher. Carefully crafted government policy may be a useful way to account for these additional costs. However, this objective should be balanced against additional inefficiencies that government involvement introduces into the market. Once policies are in place that ensure that individuals account for the full costs of the goods and services they consume—e.g., national security and environmental concerns—competitive markets are the most efficient means to determine how goods are produced, as well as which goods are produced in the future.

Transportation Fuel Supply

Motor gasoline and diesel fuel will continue to be the main sources of power for cars and trucks in the near future. In 2006, motor gasoline accounted for 74 percent of fuel used in highway vehicles, and diesel accounted for 24 percent (alternative fuels made up the remainder). Diesel cars and light trucks are uncommon in the United States—only 2 percent of new cars and light trucks sold use diesel engines; the majority of diesel fuel is used by commercial vehicles.

Ethanol, an alternative fuel, is currently used as an additive in gasoline to increase octane and help gasoline burn more completely, reducing emissions of carbon monoxide and other pollutants. In many states and metropolitan areas, gasoline sold at the pump contains between 2 and 10 percent ethanol, depending on State requirements. Using such alternatives to oil can reduce the environmental costs of transportation as well as the national security consequences of oil use. To further encourage alternative fuel use, a provision in the Energy Policy Act of 2005 (EPAct 2005) known as the Renewable Fuel Standard requires a certain quantity of renewable fuel to be used by gasoline producers each year. In 2006, producers were obligated to use 4 billion gallons per year; this obligation will gradually increase to 7.5 billion gallons in 2012 (Americans consumed about 140 billion gallons of motor gasoline in 2006). One of the strengths of this policy is that it does not choose which renewable fuel to promote, but allows the standard to be met with any renewable fuel that accomplishes the goal of reducing oil use. However, it does not extend to oil alternatives beyond renewable fuels, such as electric cars or hydrogen fuel cells. The Renewable Fuel Standard also allows imports to satisfy the standard, allowing U.S. consumers to take advantage of cheaper production of renewable fuels in other countries, although this is impeded by an import tariff on such fuels.

A more significant regulatory change has been applied to diesel fuel. Starting in 2006, diesel fuel sold in the United States is required to have a sulfur content of no more than 15 parts per million (ppm), down from 500 ppm in the previous standard. This reduction results in the most stringent diesel fuel standard in the world and enables U.S. consumers to purchase vehicles with engines that meet clean air requirements using clean diesel fuel. Diesel engines are between 20 and 25 percent more fuel efficient than comparable gasoline engines (even accounting for the fact that a gallon of diesel contains more energy than a gallon of gasoline). EPAct 2005 also grants tax credits to buyers of diesel cars that meet stringent emission standards.

Alternative Fuels and Advanced Technologies

To date, changes in petroleum usage have been driven primarily by the increasing price of oil and by regulatory concerns. The greatest potential for large reductions in gasoline consumption stems from new technologies that could transform how transportation is powered. Over 1 million advanced technology cars and light trucks were sold in the United States in 2006. About

two-fifths of these were flex-fuel vehicles that can use conventional gasoline or an alternative fuel called E85, which is approximately 85 percent ethanol and 15 percent gasoline. U.S. consumers also purchased 256,000 hybrid vehicles in 2006. Hybrid vehicles use an electric motor in conjunction with a gasoline engine to increase fuel economy.

Use of advanced technology vehicles in the United States is projected to grow over time (see Chart 6-3). The Department of Energy projects that over 3 million advanced technology vehicles will be sold in 2015 and that by 2030 they will make up more than 25 percent of all light-duty vehicles sold. Of these advanced technology vehicles, 71 percent are expected to be either gasoline–electric hybrids or vehicles that can be powered by ethanol and other plant-based fuels. Though alternative fuels currently power only a small fraction of our transportation needs, private-sector investments combined with government policies are expected to fundamentally change the energy landscape.

Chart 6-3 Projected Sales of Alternative Technology Cars and Light Trucks by Fuel Type
Alternative technology vehicles are projected to increasingly displace conventional oil-using vehicles in upcoming years.
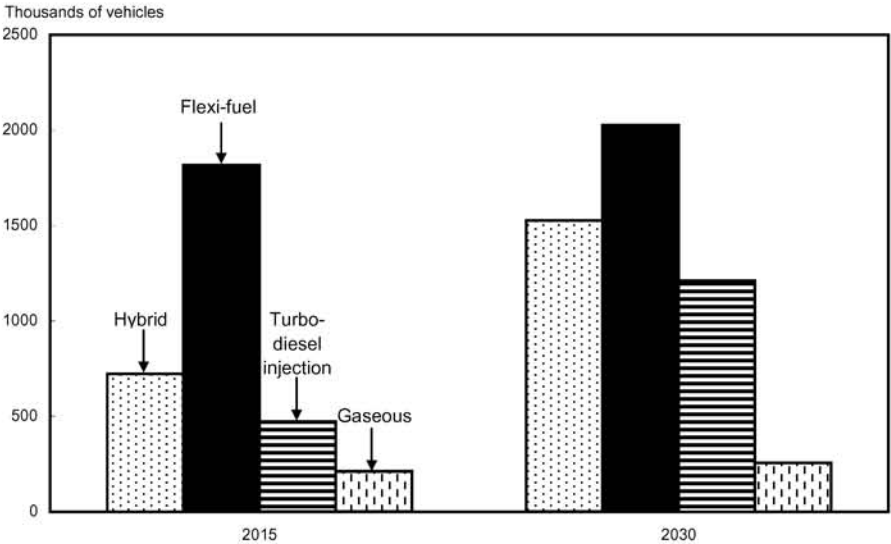
Note: Sales from fuel cell and electric cars are relatively small.
Source: Department of Energy (Energy Information Administration).

Ongoing research explores a wide variety of vehicle fuel technologies such as electricity, hydrogen fuel cells, and biofuels. Significant technological barriers exist that prevent the development of these as commercially viable alternatives. For instance, the wide-scale deployment of hydrogen fuel cells—devices that combine hydrogen with oxygen in the atmosphere to yield electricity—will depend on reductions in expense and weight as well as on the development of clean, cost-effective sources of hydrogen.

Private markets tend to underinvest in innovation of all kinds because inventors only capture a fraction of the benefits from discovery. Underinvestment is particularly likely for basic scientific research where the application to the marketplace may not be evident at early stages. Underinvestment is also likely when the results of research mainly reduce the external costs of consumption (such as national security and environmental costs associated with oil) instead of directly benefiting consumers. In response, the President's Advanced Energy Initiative proposed an increase in annual funding for alternative energy research of 22 percent for fiscal year 2007, adding to the $10 billion of government spending devoted to such research since 2001.

Several studies find that Federal research and development (R&D) investment in energy has yielded sizeable societal benefits, not only in economic terms, but also in terms of knowledge creation and pollution reduction. Still, the government's ability to predict which technologies will best meet a given goal is questionable, so the most effective government policies allow the market to choose the path of innovation.

Demand for Transportation Fuel

The United States is a vehicle-dependent society. More than 9 out of 10 American households own at least one vehicle, and most households own two. In 2004, vehicles in the United States traveled close to 3 trillion miles, up more than 20 percent from 1995. Commuting and other business-related activities account for about 35 percent of vehicle miles traveled (see Chart 6-4). Americans also use their cars and trucks to go shopping (15 percent of miles driven), attend to personal and family business such as medical appointments and dropping children off at school (25 percent of miles driven), and for social and recreational activities, including vacations (22 percent of miles driven).

Chart 6-4 Share of Vehicle Miles Traveled
Americans use their vehicles 35 percent of the time for commuting or other business-related activities.

Work-related
business
8%

Other
3%

Shopping
15%

To/from work
27%

Other
family/personal
business
25%

Social/recreational
22%
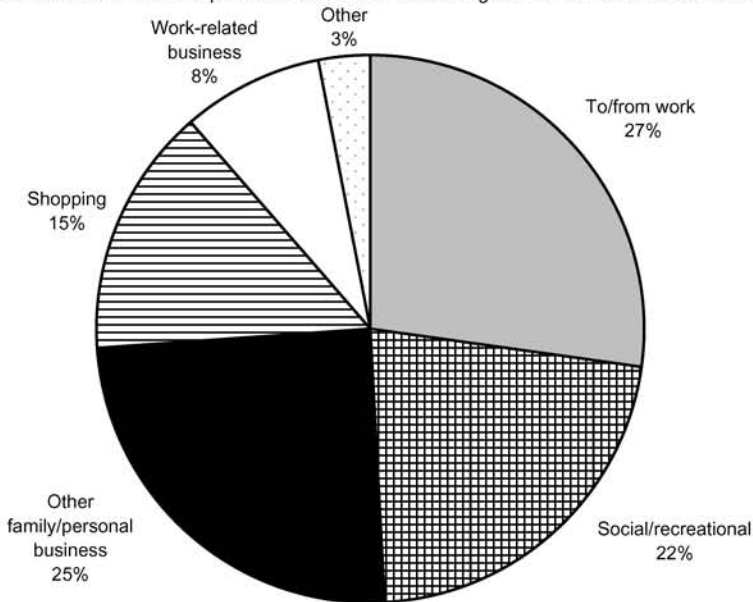
Source: Department of Transportation.

In spite of widespread vehicle use, the proportion of the American household budget spent on transport fuel is small (less than 4 percent). That said, Chart 6-4 shows that a significant share of vehicle miles traveled are related to nonwork activities, indicating that households may have some flexibility to quickly adjust when the costs of travel are high. In response to higher prices, drivers make two adjustments: they drive less and they purchase more fuel-efficient vehicles. Several studies have found that these two effects combined imply that a 10 percent increase in the price of gasoline will result in about a 4 percent decrease in gasoline consumption in the long run. Compared to other commodities, households' gasoline consumption may take several years to respond to price changes.

State and local initiatives that encourage use of mass transit and carpooling focus on encouraging people to drive less. In New York City, the most densely populated of all cities in the United States, mass transit accounts for 45 percent of all commutes into the central city. New York, however, is

unique. Many U.S. cities, such as Phoenix and Los Angeles, are spread out over a large area, making it difficult to design mass transit corridors that effectively meet the commuting needs of travelers. Public transportation also has difficulty competing with the flexibility and convenience of car travel in these types of cities. In the entire United States, 5 percent of commuters rely on public transportation.

One way many urban areas try to encourage carpooling is through the designation of high-occupancy vehicle (HOV) lanes. This method rewards carpooling by allowing vehicles with two or more passengers to travel in lanes not open to vehicles with only one person in them. In this way, HOV drivers can reduce travel time when roads are congested. Unfortunately, HOV lanes are often underutilized and the popularity of carpooling is not increasing. In 2000, 90 percent of American commuters drove to work each day, but of these drivers only about 13 percent carpooled, down from almost 20 percent in 1980. This trend makes it unlikely that initiatives focused on carpooling will make large strides in reducing vehicle fuel use.

Improving Fuel Economy

Evidence shows that drivers switch to more fuel-efficient vehicles in response to higher gasoline prices. One study finds that higher gasoline prices accelerate the retirement of older, less fuel-efficient vehicles, and shift new purchases toward more fuel-efficient vehicles. Government policies have also been used to influence vehicle fuel economy. The Corporate Average Fuel Economy (CAFE) standard, passed in 1975, mandates a minimum mile per gallon (mpg) requirement for each manufacturer's fleet of new cars and a minimum requirement for each manufacturer's fleet of new light trucks. If a given vehicle is less fuel efficient than the requirement, the manufacturer must offset it by producing a vehicle that is more fuel efficient, so that the average fuel economy for all cars (or for all trucks) the manufacturer sells is above the required miles per gallon level. One rationale used to justify increasing the stringency of the CAFE standard is to further induce improvements in the fuel economy of vehicles sold to consumers, reducing the demand for transport fuel and the external costs associated with oil use.

It is important to note that while improvements in fuel economy translate into gasoline savings, it is not a one-to-one relationship. Higher CAFE standards encourage increased driving. Since higher fuel economy vehicles can go the same distance using less gasoline, the cost of driving a mile is reduced. As the per-mile cost of driving declines, the quantity of miles driven by individuals tends to increase. This "rebound effect" reduces potential fuel savings from improvements in fuel economy by 10 to 30 percent. Recent estimates suggest that as incomes grow, driving decisions will depend less on the cost of driving, and therefore, the rebound effect is expected to shrink in the future.

In 1978, CAFE mandated 18 mpg for cars and 17.2 mpg for light trucks. The CAFE standard became increasingly stringent until 1990, after which it remained virtually unchanged. It only recently became more stringent for light trucks. Currently, the CAFE standards are 27.5 mpg for cars and 22.2 mpg for light trucks (including SUVs). The Federal government has increased the CAFE standard for light trucks through two separate regulations, raising it in increments each year beginning in 2005. By 2011, new light trucks will meet a 24 mpg standard, reflecting a 16-percent increase. Also by 2011, the largest SUVs—those weighing between 8,500 and 10,000 pounds—will be subject to the CAFE standard for the first time. The Department of Transportation based the new standard for light trucks on vehicle footprint, a measure of size, in line with a recommendation by a National Academy of Sciences panel as a way to mitigate safety concerns. The footprint-based CAFE standard for light trucks is also an improvement over its previous configuration because it ensures that all manufacturers make fuel economy improvements instead of only those producing a wide mix of vehicles. The Department of Transportation is seeking similar authority to reexamine CAFE for new passenger cars (see Box 6-1).

The fuel economy of new vehicles rapidly increased over the first 8 years of CAFE. In part, this was a market response to the dramatic increase in gasoline prices between 1973 and 1981. By the late 1980s, however, overall fuel economy had stagnated. While the fuel economy of cars has continued to slowly increase over time and has been above the CAFE standard since 1986, consumers have bought an increasing number of SUVs and light trucks whose fuel economy has remained close to the mandated level of the light truck standard. Half of all vehicles sold in 2005 were light trucks, including SUVs, compared to 20 percent when CAFE was first put in place. This shift in consumer preferences is a rational response to more than a decade of low real gasoline prices, rising household incomes, and incentives created by CAFE requirements. Manufacturers also responded to changing consumer preferences and CAFE requirements. For instance, while station wagons and minivans have similar fuel economies, the former are counted as cars, and the latter are counted as light trucks. In the late 1980s, many manufacturers took advantage of the difference in the stringency of CAFE standards across cars and light trucks to phase out the station wagon—a relatively fuel-inefficient car—and replace it with the minivan—a relatively fuel-efficient light truck. This shift improved the individual fuel economy of both the car and light truck fleets but did little to change overall fuel economy. While the CAFE standard has contributed to improved fuel economy since its inception, understanding its precise impacts and its interaction with gasoline prices is a matter of some debate. A recent National Academy of Sciences study also finds that CAFE may have led manufacturers to produce smaller and lighter cars, posing a risk to safety.

Box 6-1: The President's New Energy Initiatives

The President has announced several energy initiatives designed to increase the country's energy security by reducing projected gasoline consumption in the light-duty vehicle transportation sector by 20 percent within a decade.

About three-fourths of this goal will be met by greatly increasing and expanding the Renewable Fuel Standard. The new standard will mandate that 15 percent of transportation fuels come from alternative fuels. In 2006 about 3 percent of fuels used in light-duty vehicles were not petroleum-based. Under the revised standard 35 billion gallons will be alternative fuels in 2017. This initiative reflects the belief that technological change is the key ingredient to diversifying America's energy portfolio. Energy security will increase as the dominance of oil use in the transportation sector diminishes.

The standard will continue to allow refiners, importers and blenders to use renewable fuels to meet the standard but will expand to allow for current or future viable alternatives to petroleum to compete. Expanding the alternatives that meet the standard makes it easier for blenders and refiners to comply and affords the market broad flexibility to find the most cost-effective non-petroleum-based fuel options. In the event that production of alternative fuels proves more costly than expected, the President has built in two safety valves to protect consumers. First, the Administrator of the Environmental Protection Agency, and the Secretaries of the Department of Energy and the Department of Agriculture will have the authority to waive or modify the standard if refiners and blenders have difficulty finding alternative fuels for purchase. Second, an automatic mechanism will be in place to prevent the price of gasoline from rising above a threshold due to this policy. These two provisions ensure a degree of market stability as use of alternative fuels expands in the marketplace.

The 20 percent goal will also be met through increasing the fuel efficiency of automobiles. This will occur through reforming and modernizing CAFE standards for cars and further increasing light truck and SUV standards. These changes are predicted to reduce consumption of gasoline by an estimated 5 percent, based on the assumption that increases in the standard of 4 percent each year starting in 2010 for cars and 2012 for light trucks prove warranted. Three reforms are key to the President's proposal of increased stringency of CAFE. First, paralleling recent changes for light trucks, the law for cars should be changed to allow the standard to be based on a vehicle attribute (such as footprint) to address safety concerns. Second, CAFE for both cars and light trucks should allow manufacturers the option of increased flexibility in how they meet the standard, by allowing them to trade

credits. Any manufacturer that increases fuel economy by more than what is mandated could generate credits that other manufacturers could purchase to reduce their costs of meeting the standard. The benefit of trading credits is that it allows the same overall goal of improved fuel economy to be met at a lower cost. Third, the rate of increase of the CAFE standard as well as how fuel economy improvements will be divided between cars and light trucks should be at the discretion of the Secretary of Transportation, as is currently done for light trucks. The Department of Transportation will employ the regulatory process to determine these increases based on sound science and an assessment that balances the costs and benefits.

The President has also proposed a new $175 million initiative to give State and local governments the opportunity to explore innovative ways—such as roadway pricing and increased use of real-time traffic information—to reduce traffic congestion and save fuel.

In addition to improving the nation's energy security profile, these initiatives will also produce significant benefits by reducing air toxics associated with petroleum-based fuel. They will also help confront the challenge of climate change by potentially stopping the projected growth of carbon dioxide emissions from this sector.

Transportation Infrastructure and Management of Existing Traffic Flow

In addition to its reliance on oil, the transportation sector also relies heavily on the existing infrastructure of roads and highways. Under the Intermodal Surface Transportation Efficiency Act of 1991, the Federal government plays an important role as overseer of the National Highway System to ensure that the highway system is "economically efficient and environmentally sound, provides the foundation for the Nation to compete in the global economy, and will move people and goods in an energy-efficient manner." In recent years, however, the road and highway infrastructure has not kept pace with the number of miles driven in the United States. When more people use a roadway than the capacity for which it is built, traffic slows. Commercial trucking—the most common method of moving freight across the United States—is increasingly reliant on urban interstate highways, many of which are congested. Between 1982 and 2003 the share of roads in U.S. urban areas that are congested rose from 34 percent to 59 percent. Changes in

commuting patterns have also spread congestion to more roads. The traditional suburb-to-city commute has diminished in importance: As of 2000, half of all commuters drove to jobs in the suburbs, while only 20 percent drove to jobs in central cities.

Congestion is defined as the marked slowing of traffic as a roadway reaches capacity. Congestion in the United States manifests itself primarily as a bottleneck on a roadway (see Chart 6-5). A bottleneck is a hindrance to vehicle movement because it involves delays at key intersections, backed-up traffic, or narrow or obstructed sections of a roadway. Unexpected events such as accidents or other traffic incidents also cause congestion on crowded roadways. Together, they are responsible for 65 percent of all congestion.

Chart 6-5 Main Sources of Congestion
Bottlenecks and traffic incidents are responsible for causing 65 percent of congestion on U.S. roads.

Traffic Incidents 25%
Bad Weather 15%
Work Zones 10%
Special Events/Other 5%
Poor Signal Timing 5%
Bottlenecks 40%
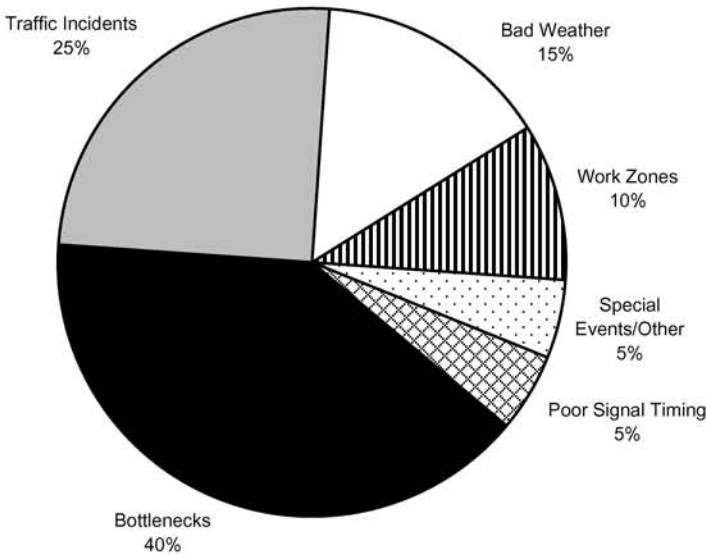
Source: Department of Transportation (Federal Highway Administration).

It is important to note that roadways are not congested at all hours of the day. For instance, on one particular roadway in the Seattle area, a trip that occurs prior to 6 a.m. or after 10 p.m. takes about 10 minutes (see Chart 6-6). That same trip takes about 30 percent longer at 8 a.m. and almost twice as long at 6 p.m. due to slowing traffic. This general trend appears in many U.S. cities and suggests that it is the timing of vehicle miles traveled more than their growth that is at the root of the congestion problem.

Chart 6-6 **Travel Time and Hour of Day on I-90 from Bellevue to Seattle**
Roadways are not congested at all hours of the day.

Note: Travel times are for general purpose (non-HOV) lanes.
Source: Washington State Department of Transportation.
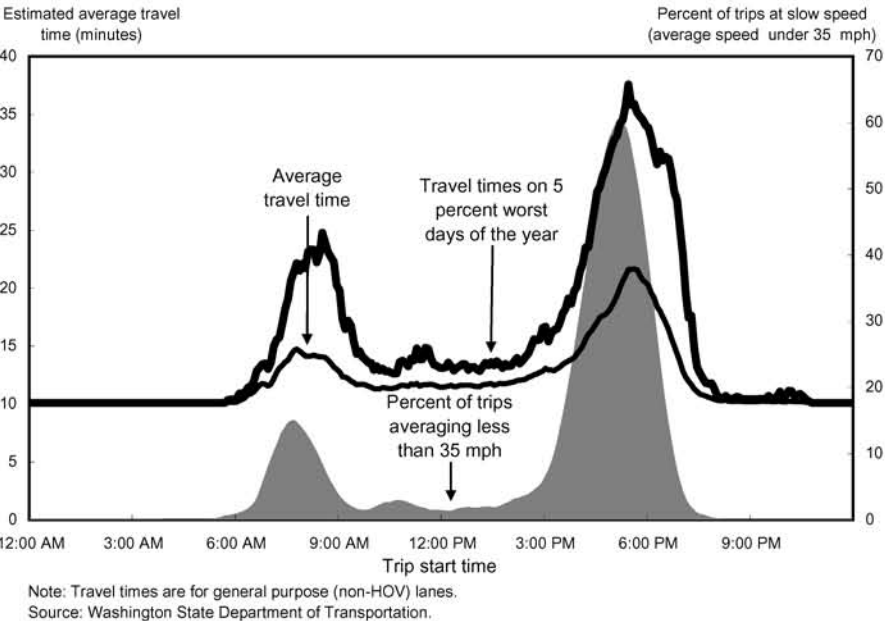
One underlying reason why congestion exists on U.S. roadways is the lack of a private market to price roadway use. Most roads in the United States are provided by the government, are open to all, and are free of charge. Economists generally believe that a good may be better provided by the government when it is difficult for private markets to charge for its use. Because one motorist's use of a congested road reduces the road's value for other drivers and drivers can be selectively prevented from entering the roadway through the use of gates or technologies that monitor use, it is increasingly appropriate to charge drivers for some roadway use in the same way the private market charges for other goods and services.

A driver decides which road to use based on private needs: for instance, the shortest distance or fastest route between destinations, or the closest, most accessible highway. The fact that each driver decides on a route independently of other drivers is not a problem when the number of drivers is well below the roadway's capacity. However, when drivers have free access to roads, crowding occurs at times of high demand, decreasing vehicle speed and flow. Each additional driver slows down other drivers on the roadway, causing them to lose time and to burn extra gasoline. However, drivers typically do not consider the added costs they impose on others. This is a "get in line" or "queuing" approach to allocating road space. When there is a shortage of something—

for instance, space on a ski lift, or attendants at the Department of Motor Vehicles—those willing to get in line and wait eventually receive what they want. This approach to road-use management is inefficient because it allocates road space to those with the time to wait in traffic, not necessarily to those who value its use most highly.

If a roadway is priced—that is, if drivers have to pay a fee to access a particular road—then congestion can be avoided by adjusting the price up or down at different times of day to reflect changes in demand for its use. Road space is allocated to drivers who most highly value a reliable and unimpaired commute. This arrangement encourages drivers to consider the tradeoff between the price of using the road and the additional time and inconvenience of using a nonpriced, alternate route, or driving at a noncongested time. Drivers who place a high value on the predictability and reduced time of commuting, for instance, a doctor who has been called to the hospital for an emergency, have the option to pay for access to noncongested roads. Drivers with more time flexibility, for instance a person doing his or her grocery shopping, can avoid the road and the fee. They can use alternative but more congested roads, shift when they drive to nonpeak hours, or use mass transit when it provides a cheaper alternative to driving. The average cost to each driver falls because drivers have a choice in how they pay for roadway use, in time or in money.

The Cost of Congestion

Over time, slowing traffic exacts heavy costs on drivers. On average, congestion caused 47 hours of delay for U.S. commuters and commercial truck drivers in 85 urban areas during peak hours in 2003. For America's 13 largest cities, this number is much higher: 61 hours. Extra fuel is consumed on congested roads because of the effect that waiting in stop-and-go traffic has on fuel economy. In 2003, sitting in traffic wasted about 2.3 billion gallons of fuel, or almost 1.4 percent of all fuel consumed by light-duty and commercial vehicles that year. Waiting in traffic can also increase the cumulative amount of pollution emitted from a vehicle's tailpipe, which contributes to poor air quality and more greenhouse gas emissions.

Aggregating over the 85 most congested U.S. cities, the cost of time wasted in traffic and extra fuel consumed by commuters and commercial truck drivers due to congestion is estimated to have exceeded $63 billion in 2003 (see Table 6-1). In Los Angeles, the city with the worst congestion, the fuel and time cost of waiting in traffic was calculated to be almost $1,600 per traveler in 2003. In Philadelphia, congestion is noticeably less than in Los Angeles, but the estimated cost to travelers is still high: $641 per traveler per year. In addition, businesses that rely on regular and on-time delivery of supplies have begun to maintain larger inventories to safeguard against

TABLE 6-1.— *Cost of Congestion in Wasted Time and Fuel in the largest Urban Areas*

Metro area	Annual delay per traveler (in hours)	Total cost ($ in millions)	Cost per peak traveler
Los Angeles–Long Beach–Santa Ana CA	93	$10,686	$1,598
San Francisco–Oakland CA	72	$2,605	$1,224
Washington DC–VA–MD	69	$2,465	$1,169
Atlanta GA	67	$1,754	$1,127
Houston TX	63	$2,283	$1,061
Dallas–Fort Worth–Arlington TX	60	$2,545	$1,012
Chicago IL–IN	58	$4,274	$976
Detroit MI	57	$2,019	$955
Miami FL	51	$2,486	$869
Boston MA–NH–RI	51	$1,692	$853
Phoenix AZ	49	$1,294	$831
New York–Newark NY–NJ–CT	49	$6,780	$824
Philadelphia PA–NJ–DE–MD	38	$1,884	$641

Source: Texas Transportation Institute, 2005 Urban Mobility Report.

unanticipated delays caused by congestion. A recent study conducted by the Department of Transportation confirms that congestion has resulted in higher transportation prices and less reliable pickup and delivery times for freight.

Building More Roads

Expanding road capacity may be an important component of any long-term strategy to accommodate traffic growth in urban areas. However, there are a number of reasons why a construction-only strategy to alleviate congestion is likely not the best solution. First, increasing capacity can take years to complete and is expensive—one study found that a lane costs between $1 million and $8.5 million per mile to build. Second, new lanes are often needed in densely populated areas, but these are often also the areas where it is most difficult to find unoccupied space for expansion, making new lanes politically controversial. Third, a body of evidence suggests that the addition of a nonpriced lane to an already congested roadway may do little to alleviate congestion. This happens for two reasons: new roads generate additional traffic as drivers take trips to destinations that previously took too long to reach. And since traffic flow improves initially, drivers who were previously using alternative, often less congested routes now find the highway with the added lane more attractive. Drivers continue to redistribute themselves across the various routes until the costs of using the new route and the costs of using the existing route are about equal. At this point, no driver can be made better off by changing routes. Ultimately, the reason why building more roads is insufficient is because it does not address the underlying problem: roads are not priced and are therefore subject to overuse.

Pricing Road Space

There is reason to believe that reductions in traffic congestion would be relatively easy to attain. Small changes in the number of cars using a particular roadway at a given time can result in large improvements in the flow of traffic. For instance, the addition of just a few school buses makes traffic flow noticeably worse on the first day of school, while traffic flow is noticeably better on some State holidays when only a small number of residents stay home from work.

Congestion pricing dampens demand for roads during peak hours and spreads usage over a longer time period. Differentiating the price of a good by the time of day effectively allocates limited space during periods of higher demand. This approach is used by many providers of goods and services: movie theaters charge more in the evening than they do midday; ski runs charge more during weekends than they do on weekdays; airlines raise prices on tickets during peak seasons; taxi cabs charge more during rush hour; and railroads often charge lower prices for offpeak traveling.

In addition to improved allocation of road space, charging a fee also provides urban planners with useful information about when and where to invest in the expansion of existing road capacity. Expansion should be focused on roads where drivers demonstrate a willingness to pay that is higher than the costs of construction. Revenues from roadway pricing may also prove a viable alternative to taxes as a way to fund the building of new roads in urban areas. As is the case in other markets, those who use the roadway would pay for its maintenance and expansion.

In general, there are two ways to price road space to address congestion: cordon pricing and roadway pricing. *Cordon pricing* charges a toll to vehicles for access to a congested area regardless of which roads in the area are used. It is typically in effect during the work week and varies by time of day. Cordon pricing has been implemented in a number of cities including London, Stockholm, and Singapore. While cordon pricing has been considered for several cities in the United States, it has not yet been implemented here. It is likely to be less effective in cities that are less dense, do not have adequate public transportation systems, and have multiple areas of centralized economic activity (such as Phoenix or Los Angeles).

Evidence suggests that cordon pricing fees have been effective in reducing congestion where they have been tried. After the first year that cordon pricing was imposed in London, for instance, congestion fell by 30 percent, average vehicle speed increased by 20 percent, and bus travel became more reliable (see Box 6-2). One important mechanism for reducing congestion appears to be the ability to substitute some form of public transportation for driving.

Box 6-2: Cordon Pricing Experiences in London and Stockholm

In London, drivers pay an 8-pound fee for daily access to a portion of downtown between the hours of 7:00 a.m. and 6:30 p.m. on weekdays. There are no toll booths around the perimeter of this area. Instead, cameras record the license plates of vehicles and check them against a list of prepaid vehicles. Drivers have a variety of choices in how they pay: they can pay at designated service stations, through the Internet, by text message or phone, or by mail. Weekly and monthly charges also are available for regular commuters. If drivers have not prepaid, they have until midnight of the next day to do so. Anyone who drives within the zone without paying during this time period is fined 100 pounds through an automated system.

Stockholm also recently implemented cordon pricing, but it differs from the London system in two ways. First, it charges vehicles via a card mounted on the windshield that is read electronically by roadside beacons when cars drive past them. Second, Stockholm uses a variable pricing system, which means that the fee is higher during rush hour periods.

A recent report on the London policy indicates that cordon pricing has led to a 30 percent reduction in delay time for city commuters. Initial reports from Stockholm's 6-month test period indicate that there were decreases in traffic of about 22 percent due to cordon pricing. Large reductions in London and Stockholm traffic were due in part to increased use of bus transit. In spite of early criticism from drivers and businesses within the central city, cordon pricing has grown in popularity in London. In Stockholm, this has also been the case: a majority of residents voted to retain cordon pricing after the test period ended.

Roadway pricing aims to limit congestion on certain routes by charging variable fees (tolls) to access a particular lane or road, regardless of the final destination. Ideally, road tolls should be responsive to the actual level of congestion at each moment. By increasing the fee during periods of high demand and reducing it during periods of low demand, the variable tolls reduce congestion by encouraging offpeak driving and the use of alternative routes.

Variable tolls are rare in the United States. Most of the over 5,000 miles of toll roads in the United States have flat tolls designed to generate revenue, rather than variable tolls to relieve congestion. Where they do occur, they are typically limited to a single road or freeway. On the congested bridges and tunnels connecting New York and New Jersey, tolls are discounted by

20 percent ($1.00) during nonpeak hours. Results of a small survey indicate that about 7 percent of drivers changed their behavior as a result of these variable tolls. The most common changes were to switch to mass transit, carpool, or to increase offpeak driving.

Recently, the Department of Transportation helped fund a small pilot project in Seattle to examine how drivers would respond if the entire road system in the city were subject to a variable tolling system. Where and when participants drove was automatically tracked and transmitted by a device installed in their car. Participants received prepaid accounts between $600 and $3,000 to pay the tolls. At the end of the pilot, they were allowed to keep whatever they did not spend. Tolls ranged from 5 to 50 cents per mile and varied by road and time of day. Preliminary results show that nearly 80 percent of participants decreased the amount they drove or changed when they drove. On average, participants took 5 percent fewer trips by automobile and drove 2.5 percent fewer miles each weekday due to tolls. Participants took 10 percent fewer trips and drove 4 percent fewer miles during the morning commute.

Currently, there are about six U.S. highways that use high-occupancy toll (HOT) lanes, many of which incorporate variable pricing and were piloted using Federal funds. HOT lanes are variations of the high-occupancy vehicle (HOV) lanes discussed earlier in the chapter, but they have greater potential to reduce congestion since they are less likely to be underutilized. Similar to HOV lanes, they allow carpoolers to use the road for free or at a discount but charge a toll to single occupancy drivers for access. The toll frequently varies by time of day. Some tolls set variable prices based on historical highway use and adjust rates monthly or quarterly. Other tolls use real-time information on congestion conditions to adjust tolls dynamically over the course of the day. In locations where HOV lanes are underutilized, conversion to HOT lanes is suggested as a way to increase use and to provide more choice to drivers. For instance, in San Diego, conversion of HOV lanes to HOT lanes on a portion of Interstate 15 increased usage by 64 percent over a 3-year period. Several studies confirm that there are substantial gains in societal welfare from allowing solo drivers to pay for access to existing HOV lanes. Others caution, however, that when only one HOV lane is converted to a variable toll and other lanes are free of charge, any temporary decrease in congestion on the remaining free lanes may be offset by the redistribution of traffic.

The use of real-time or historically based variable tolling on HOT lanes may have a significant effect on traffic flow. For instance, San Diego's variable toll uses real-time pricing, which changes every 6 minutes to reflect the amount of traffic on the road. Computerized electronic signs make information on the toll amount and the speed and flow of traffic available to drivers before they have to decide between the free and priced lanes. Results show that travel times

vary little on San Diego's variable toll lanes because free-flow conditions are almost always maintained. In Orange County, the tolls vary by hour and day of the week, but are based on historical information. While they are adjusted several times each year, the toll does not convey actual conditions to drivers, only average conditions. Thus, unexpected events such as accidents can cause major delays on the variable toll lanes and because drivers do not have up-to-date information on road conditions, travel time is less predictable.

Despite their potential benefits, toll lanes are sometimes portrayed as "Lexus Lanes." The contention is that tolled roadways supply faster routes only to high-income drivers who can afford to pay the tolls, while lower income drivers continue to be stuck in traffic. One study finds that drivers with higher incomes tend to use HOT lanes more often than lower income drivers, but that lower income drivers rely on toll lanes when on-time arrival at their destination is important. For instance, you can imagine a case where a parent is running late, but needs to be at the daycare to pick up his or her child by a certain time. If the parent is late, and the daycare fines him or her $10, then paying a $4 toll to arrive on time saves $6. A recent survey also finds that support for or opposition to HOT lanes is unrelated to income. Another study finds that lower income, bus commuters were some of the largest beneficiaries of cordon pricing in London. Bus riders are exempt from paying the cordon fee, but their commute times greatly improved. Not surprisingly, the number of bus passengers during morning hours increased.

Experts note that implementation of congestion pricing faces less resistance where motorists are unaccustomed to free and unrestricted roadway access. For instance, it may be more feasible to implement congestion pricing on a new road than on an existing road. Likewise, it may be easier to convert HOV lanes to HOT lanes. The advent of new technologies that electronically charge the toll by sensing a microchip placed on the windshield of the vehicle eliminates the need for a driver to stop and physically pay the toll. These are increasingly used to charge drivers tolls on existing roadways, making congestion pricing systems easier and less costly to implement.

Historically, one of the largest hurdles to variable price tolling on roadways in the United States has been the Federal-aid highway program, which has prohibited states from collecting tolls on interstates or other roads that receive Federal funding. Federally funded pilot projects that explored variable price tolling brought the advantages of congestion pricing to the attention of policymakers. Policymakers also began to explore the use of pricing mechanisms to reduce congestion in other contexts, such as for allocation of runway access at airports (see Box 6-3). A transportation bill signed into law in 2005 (The Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users) provides states with increased flexibility to use tolling to manage congestion and finance infrastructure improvements, and provides ways to

participate in pilot demonstrations of variable tolling. States such as Texas and Colorado have passed laws allowing the formation of toll authorities at local levels that can then construct and operate toll roads. States such as Washington, California, Florida, and Minnesota have identified candidate freeways for variable tolling.

Box 6-3: Airport Pricing to Decrease Congestion

Though traffic jams are easily observable manifestations of congestion, flight delays and runway bottlenecks also waste time and fuel. Landing fees at most U.S. airports are directly related to the weight of the plane, even though lighter and heavier planes tend to consume approximately the same runway time. This contributes to airport congestion because it encourages smaller, lighter planes (which can use smaller satellite airports) to overuse the airport, displacing larger, heavier passenger planes and reducing the number of passengers that an airport can serve at a time.

A short-lived experiment at Boston's Logan airport in 1988 demonstrates how a change in the landing fee structure can effectively reduce airport congestion. Boston changed its runway use fee from one based only on aircraft weight to one that combined a non-weight-based fee and a smaller weight-based component. The fee for a small single-engine plane increased from $25 to about $100, while the fee for a large jumbo 747 jet decreased from $800 to less than $500. By flattening the landing fee, Logan made it relatively more costly to land small planes, decreasing their volume. This allowed it to more easily accommodate the larger planes that carry more passengers. The result was that Logan airport reduced delayed landings from 30 percent to 14 percent in less than 4 months. Despite a reduction in congestion, the new landing fee structure abruptly ended when the program was deemed to be in violation of the Federal Aviation Act.

The auctioning of runway access for planes may prove to be an even more effective way to reduce congestion at airports. An auction would award landing rights to the carrier that values the slot the most. Such auctions have been successful in other contexts such as to allocate radio waves while still accommodating smaller local and public radio stations.

Conclusion

The transportation industry relies overwhelmingly on petroleum for fuel. In spite of its reliance, the market largely functions as it should; while transportation is particularly unresponsive to changes in oil prices in the short run due to the lack of readily available substitutes, it does eventually respond. Also, the price reflects the costs to the firm of producing the oil and the benefits to drivers from consuming the oil. That said, the use of oil by the transportation and other sectors generates costs to national security and the environment that users typically do not take into account. Likewise, the full costs of congestion are not taken into account by individual users when they drive, since roadway use is not priced by the market. Carefully crafted policies could help address these costs but care should be taken as government action itself imposes inefficiencies.

Currency Markets and Exchange Rates

In the modern economy, firms buy and sell products from more than just local or national markets. Often a firm's supplier is located in a different country. To make purchases and sell their own goods internationally, firms need to change units of one currency for units of another currency. For instance, when a British firm trades with a U.S. firm, the U.S. firm may pay in U.S. dollars. However, the British firm needs to pay many of its costs in British pounds. When the U.S. firm pays the British firm, then, one of two things has to occur: the U.S. firm must convert its dollars to pounds and then pay the British firm in pounds, or the British firm must accept dollars from the U.S. firm and then convert the dollars into pounds to pay its workers. And, to be sure that the sum in pounds is equivalent to the sum in dollars, all parties to the transaction must know the value of dollars in terms of pounds. Now multiply this single transaction by the number of countries and firms involved in all aspects of the production of all internationally traded goods and services and one can see that multiple currencies make international trade far more complex and difficult than domestic trade.

The desire to transact internationally provides the impetus for a huge, well-functioning market that facilitates such currency conversions and allows global economic integration and trade to take place smoothly and quickly at low cost. Both by volume of trade and ease of making transactions, currency markets today are the world's deepest, most liquid markets in the world. Currency markets range from simple markets where parties simply exchange one currency for another, to sophisticated markets where parties buy and sell currency far into the future.

In 2005 the United States imported and exported over $3 trillion worth of goods and services. In addition, gross sales and purchases of long-term U.S. securities, such as corporate and Treasury bonds, to residents of foreign countries amounted to around $41 trillion. Most of these transactions either directly or indirectly required a *foreign-exchange transaction*. A foreign-exchange transaction is a trade of any two currencies. For example, a purchase of Japanese yen with U.S. dollars is a foreign-exchange transaction.

As cross-border transactions have become larger and more frequent, foreign-exchange markets have become increasingly important to the global economy and have grown in relative size: whereas U.S. cross-border trade in goods and services and long-term securities are measured in trillions of dollars per month or year, turnover in foreign-exchange markets is measured in trillions of dollars *per day*. Daily average turnover in global foreign-exchange

markets averaged $1.9 trillion in April 2004. (Note: Unless otherwise noted, all foreign-exchange transactions data in this chapter are from April 2004, the latest date for which global turnover data are available.)

Foreign-exchange transactions vary in size and complexity. A foreign-exchange transaction is simply a trade of one country's currency for that of another, whether the amount traded is a few dollars or a few billion dollars; whether the entity making the exchange is a tourist changing money at the border for a short holiday or a foreign company building a new factory needing to exchange millions in domestic currency to pay for materials and labor; or whether the form of money being acquired is foreign currency notes, foreign currency bank deposits, or assets such as stocks or bonds denominated in foreign currency. Key points of this chapter are:

- Foreign-exchange markets not only allow firms to trade goods and services across borders but also allow firms to manage the risks they face from fluctuations in the price of their domestic currency.
- As with any other good, the exchange value of a currency is determined by its supply, as well as the demand for the country's assets, goods, and services.
- Over much of the 20th century, countries tended to favor fixed exchange rates. In recent decades, there has been a shift away from fixed regimes toward freely floating exchange rates.
- Monetary and exchange-rate policies are tightly linked. A nation's government must decide between controlling its exchange rate and controlling its domestic inflation rate.

Currency Markets Are Large

On an average day in April 2004, an amount equivalent to $1.9 trillion was traded in the foreign-exchange market. These trades occurred between different agents (individuals, firms, banks, governments) and for different reasons, varying from tourist demand for currency to firms needing payment for goods in local currency. To put this number in perspective, on average in 2004, every 7 trading days a sum greater than the entire value of the U.S. annual GDP changed hands in the foreign-exchange market. Not surprisingly, turnover in the foreign-exchange market is larger than turnover in most other financial markets. For example, the dollar value of average daily trading on the New York Stock Exchange, the largest exchange in the world, was around $46 billion in 2004, roughly 2 percent of the turnover in all world foreign-exchange markets.

When currencies are traded in the foreign-exchange market, participants need to know the value of their currency relative to other currencies, just as participants in a traditional stock market need to know the value of the stocks they wish to buy or sell. In foreign-exchange markets, this price is known as

the *exchange rate*, the number of units of one nation's currency that must be traded to acquire one unit of another nation's currency. For example, on October 11, 2006, a person wanting to acquire one British pound would have had to pay $1.86 in U.S. dollars. By November 30, 2006, a person wanting to make the same trade would have had to pay almost $1.97 for one British pound. In this case, the dollar is said to have *depreciated*. After the *depreciation*, more dollars are required to buy the same number of pounds. If the transactions are viewed from the perspective of the pound, the pound is said to have *appreciated*; fewer pounds are required to purchase each dollar.

In principle, an exchange rate exists between each possible pairing of the individual currencies in the world. Among the 52 nations (out of a world total of 193 nations) that reported formal exchange-market transactions in 2004, there are 820 possible bilateral exchange rates. If the 12 European nations that share the euro as their national currency had separate currencies, this number would be even higher.

In reality, a substantial portion of foreign-exchange trading occurs through an *intermediate* or a *vehicle currency*, that is, a currency that is widely used throughout the world. For example, the U.S. dollar serves as a global vehicle currency and the euro is becoming an important vehicle currency in Europe. A Turkish bank that wishes to exchange Turkish lira for Swedish krona may first exchange lira for euros and then exchange the euros for krona. Vehicle currencies reduce transaction costs in foreign-exchange markets because a bank wishing to provide foreign exchange for its customers need not keep stores of large numbers of currencies on hand. Instead, it need only maintain stores of its own domestic currency and one or two other vehicle currencies.

The U.S. dollar is the most important vehicle currency in the world. The dollar has served as an important vehicle currency in part because it has remained remarkably stable over time. This stability is in part a result of the United States' long history of flexible exchange markets and its commitment to improving capital market and trade access to the United States. As of 2004, the U.S. dollar was used in almost 89 percent of world currency transactions; its average turnover was over $1.5 trillion per day, more than twice as much as the next most-used currency, the euro. Most of this trading occurs outside of the United States.

Just as a few vehicle currencies dominate the transactions, two trading locations dominate foreign-exchange market transactions. In 2004, over half of the world exchange-market transactions occurred either in London (31.3 percent) or New York (19.2 percent). The next-largest location in terms of trading share was Japan, with 8.3 percent of transactions. Foreign-exchange market transactions are also concentrated among a few large banks. In the United States, 75 percent of transactions were conducted by only 11 banks in 2004. In the United Kingdom, 16 banks captured 75 percent of foreign-exchange market transactions.

Innovations in technology, such as computers and international communications networks, and breakthroughs in economic theory that have improved our understanding of the value of currencies, have made foreign-exchange markets among the most sophisticated markets in the world. Investors can easily take advantage of small differences in exchange values across the different global markets, buying a currency for a lower amount in one location and selling it for a higher amount in another, making the global currency market one global exchange.

The sophistication of modern currency markets also helps multinational firms protect themselves, or *hedge*, against *currency risk*. Because costs and revenues of multinational firms are often denominated in different currencies, currency risk is a fundamental part of international trade, and changes in the exchange rate affect the cash flow of the firm. For example, a Mexican manufacturer may enter into a contract with a U.S. firm, agreeing to sell its product at a fixed dollar price for a set period of time, for example, 1 year. The Mexican manufacturer must pay its employees in Mexican pesos but will receive a fixed dollar stream of revenue. If the peso appreciates over the year (that is, if the peso becomes more valuable so that it takes fewer pesos to buy one U.S. dollar), the manufacturer's dollar-denominated revenue will fall in value relative to his peso-denominated costs. If the peso appreciates sufficiently, the manufacturer may not be able to cover his costs. To see this dilemma more clearly, suppose that when the Mexican firm enters into the contract with its U.S. counterpart, the exchange rate is 10 pesos per dollar. If the firm has costs of 1,000 pesos, and it receives $110, then the firm is able to cover its costs and has 100 pesos of profit after the transaction. However, if the peso appreciates over the year from 10 pesos per dollar to 8 pesos per dollar, after the firm receives payment of $110, it will only hold 880 pesos. The firm would not be able to cover the costs from the revenue it receives. If the firm has no way to hedge this risk, its owner may be unable or unwilling to enter into the contract and thus the opportunity for Mexico and the United States to realize gains from this trade may not be realized. Advances in economic theory that have helped companies learn how to price risk appropriately have enabled financial markets to develop contracts that allow firms to sell their currency risk. Boxes 7-1 and 7-2 provide more detail.

Box 7-1: Types of Currency Market Transactions

A *spot transaction* is an immediate exchange of one currency for another. A tourist exchanging currency upon arrival at an airport is an example of a person making a spot transaction. Spot transactions between professional currency traders specify a *clearing date* that

---

requires the actual exchange of currency within 2 business days; the 2 days gives each side of the transaction ample time to move funds. As a share of total foreign-exchange market turnover, spot transactions have declined from 54 percent in 1989 to 33 percent in 2004. The share of spot market transactions has not shrunk because the spot market is smaller—the volume of spot transactions almost doubled between 1989 and 2004—but because the growth rate of other types of foreign-exchange transactions has grown at a much faster rate. For example, over the same time period, transactions with clearing dates in the future have increased almost eightfold.

A *forward transaction* is similar to a spot transaction except that the clearing date (also called the *settlement date*) is in the future. The price at which the parties agree to exchange currency on the settlement date is known as the forward exchange rate and it almost always differs from the spot rate at the time the contract is entered into. In a forward transaction, no currency changes hands until the settlement date. The primary purpose of a forward transaction is to allow multinational firms to hedge their currency market risk. A foreign-exchange *futures transaction* is virtually identical to a forward transaction. The main differences between a forward and a future transaction lie in the institutional details of the transaction. For example, futures contracts tend to be much more standardized than forward contracts and are sold on organized, centralized exchanges.

Foreign-exchange swaps combine a spot and a forward transaction into one transaction. Foreign-exchange swaps are typically used by banks and other dealers when they wish to temporarily reallocate their portfolio into or out of a currency without incurring any exchange-rate risk. In the swap, one currency is swapped for another for a prespecified period of time. In about two-thirds of foreign currency swaps, the swap period is less than 1 week. In 2004, foreign-exchange swaps accounted for about 50 percent of the foreign-exchange market turnover. A foreign-exchange swap is particularly useful for a firm that has payments and expenses payable in the same currency but payable at different dates. For example, a U.S. firm may receive a euro-denominated payment from its German affiliate. The firm plans to use the payment to purchase euro-denominated goods in 1-month's time. However, over the month, the firm would like to invest the money in the United States. This firm could use a foreign-exchange swap in which it trades the euros for dollars today and trades the dollars for euros at the end of the month.

A *foreign-exchange* or *currency option* gives the buyer the right, but not the obligation, to purchase a prespecified amount of currency at a prespecified price. Depending on the type of option, the contract can either specify a date on which the option may be exercised (European option) or may specify an expiration date, where the buyer may exercise the option anytime prior to the expiration date (American option).

Box 7-2: Hedging Against Foreign-Exchange Rate Fluctuations

In 2005 Volkswagen, a German automobile company, announced to the world that it was going to increase its hedging of foreign-exchange risk. Volkswagen was exposed to foreign-exchange risk because the majority of its operating costs, in particular a portion of its labor costs were denominated in euros, while a substantial share of its revenues were denominated in U.S. dollars. In other words, Volkswagen paid its workers in euros and received U.S. dollars for the cars it sold in the United States.

Between 2002 and 2004, the euro appreciated considerably relative to the dollar. That is, more dollars were required in order to purchase each euro. Since Volkswagen was unable or unwilling to change the price of cars sold in the United States enough to offset this swing in the exchange rate, the company's dollar revenues from sales in the United States lost substantial value in terms of euros. With costs holding steady and revenues falling, Volkswagen's profits on U.S. operations were reduced by an unfavorable change in the euro/dollar exchange rate.

To avoid similar losses in the future, the company chose to combat the appreciating euro by increasing its hedging of foreign-exchange risk. Between 2004 and 2005, Volkswagen more than doubled its use of a variety of currency market contracts. In essence, this hedging strategy involved buying forward contracts for euros at a predetermined rate so that if the euro were to appreciate relative to the dollar and cause an unexpected reduction in dollar revenue, the company would receive an offsetting profit from its forward contract. If the euro were to depreciate and cause an unexpected increase in dollar revenue, the company would incur an offsetting loss from its foreign currency position. In this way, Volkswagen was able to shield its revenue flow from foreign-exchange volatility for the duration of its futures contracts.

Volkswagen's strategy highlights the benefits of hedging against the currency risk posed by short-term fluctuations in exchange rates. When faced with a *permanent* shift in the exchange rate, however, companies operating in multiple currencies are forced to either change their prices, which are in one currency, or change their costs, which are in another. Volkswagen has therefore shifted some of its euro costs into dollar costs by expanding production facilities in the United States. This strategy, known as natural hedging, permanently eliminates the currency mismatch between revenues and costs.

What Determines Currency Values?

The exchange rate is a market price, and like other market prices it is determined by the interaction of buyers and sellers in the market. In the foreign-exchange market, the demand for a country's currency arises from two sources: demand for a country's assets and demand for a country's goods and services. When analyzing foreign-exchange markets, the supply of a country's currency is usually taken as given and fixed at an amount determined by the country's central bank. The role of the central bank and the supply of money will be revisited when exchange-rate policies are examined later in this chapter.

The concept of *parity* is central to any analysis of how exchange rates are determined in the foreign-exchange market. Two types of parity are particularly important: *interest rate parity* and *purchasing power parity (PPP)*. Exchange rates and prices that move too far from either concept of parity will tend to move back toward the level implied by interest rate parity and purchasing power parity as economic agents try to exploit pricing differences across countries. In this way, the prices for currencies in the foreign-exchange market adjust. Just as prices across markets within a country tend to move toward each other as buyers tend to go to the lower priced market and sellers tend to go to the higher priced market until prices are equalized. The absolute volume and speed of asset trading tends to make interest rate parity a short-term condition while purchasing power parity tends to hold over a somewhat longer time horizon.

Interest Rate Parity

For the United States, the volume of international trade in assets is many times larger than the volume of international trade in goods and services. As a result, day-to-day fluctuations in the exchange rate tend to be driven much more by the value and desirability of a nation's assets than by the value and the desirability of goods and services the nation is selling. That is, demand for assets tends to determine the value of a nation's currency in the very short run because asset trade drives such a large part of the day-to-day transactions in the foreign-exchange market.

Goods and services are purchased for use today while assets are bought in order to purchase goods in the future. A financial asset is at its most basic a contract that offers a payment at some future date. For some assets, the contract is explicit: U.S. Treasury notes contain a promise to pay the face value of the bond at a certain date in the future as well as a fixed sequence of interest payments over the life of the note. For other assets, the contract is implicit: buying a stock in a company gives the holder the right to sell the stock at a future date but not at any explicit price. Because assets involve a future payment, the *return* on an asset—the return is the future payment

divided by the purchase price—is typically uncertain. Assets differ in the amount of risk they offer. For example, a Treasury bond is considered to be less risky than a stock. For any given level of risk, assets with higher future payments are more desirable and tend to have higher prices.

Further, because the payment of an asset may vary depending on the conditions at the time the payment is due—the stock may have a high price or a low price when the holder sells the stock—information about the likely amount of the future payment also affects how much of the asset people want to hold today. For example, when a firm announces an increase in future dividend payments, the price of the firm's stock often increases. This increase in price reflects an increase in the desire to hold the stock. Every time new information is released, investors reevaluate their holdings of assets.

The foreign-exchange market plays an important role in determining the value and return to foreign currency assets. When buying assets that are denominated in a foreign currency, investors must take into consideration both the future payment in terms of the foreign currency and any change in the relative values of the two currencies, the exchange rate. For example, in August 2006 the interest rate paid on 10-year U.S. Treasury bonds was 4.9 percent and the interest rate paid on 9- to 10- year German Treasury bonds was 3.9 percent, a difference of 1 percentage point (see Chart 7-1). Does this difference imply that investors should have preferred U.S. Treasury bonds to German Treasury bonds?

Chart 7-1 **Interest Rates on U.S. and German Long-Term Bonds**
U.S. bond yields were higher than German bond yields in 2006.

Interest rate (percent)
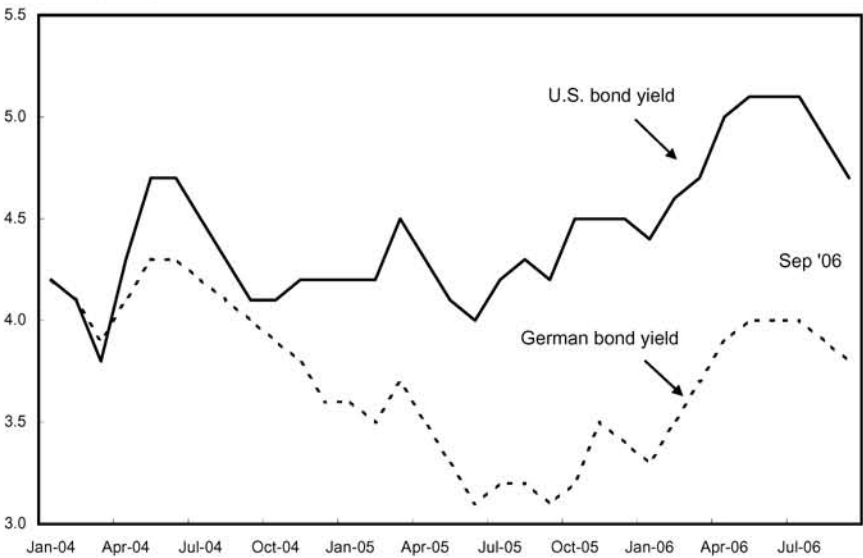
Source: U.S. Federal Reserve Board and Deutsche Bundesbank.

Not necessarily. The expected return for a U.S. resident who purchases a German bond includes both the interest paid on the bond, in euros, and the expected change in the exchange rate over the period during which the bond is held. In other words, the return on a German bond, from a U.S. investor's perspective, includes both the explicitly defined interest rate and the value of this return once converted back to U.S. dollars, an effect that can increase or decrease the return to the bond.

An example will clarify the concept. A German investor wishes to calculate the expected return on investing €100 in a savings deposit at a bank in the United States for 1 year. She needs three pieces of information to calculate the expected return: the current exchange rate between the dollar and the euro, the interest rate paid on the savings deposit, and the exchange rate that will prevail 1 year in the future. The investor knows the first two variables (today's exchange rate and the interest rate) with certainty. The one element of the calculation that is not readily available is the future exchange rate. For this example, let's first assume the investor knows all three variables: today's exchange rate is $1 per euro, the interest rate to be paid on the savings deposit is 5 percent, and the future exchange rate is $0.99 per euro (the euro *depreciates* relative to the dollar). The calculation of the investor's return is straightforward: after exchanging her €100 for dollars, she has $100 in hand and deposits it in the U.S. bank account. At the end of one year, she withdraws $105 from the bank account and takes it to the foreign-exchange market to trade the $105 dollars for ($105 / .99 =) €106.06. The effective return on the savings deposit was 6.06 percent: the $5.00 in interest earned by the $100 at 5 percent plus the €1.06 gained because the euro depreciated by 1 percent.

In the example, the future exchange rate was taken to be 0.99, a 1-percent depreciation of the euro relative to the dollar. Holding the U.S. interest rate fixed, changes in the future exchange rate have large implications for the rate of return. For example, if the euro had appreciated by 1 percent, the return would have been just under 4 percent: the $5.00 in interest earned by the $100 at 5 percent minus the €1.04 lost because the euro appreciated by 1 percent. Had the future exchange rate been the same as the initial exchange rate, the return would have been the 5 percent paid on the deposit. Investors must take into consideration future changes in the exchange rate in order to decide which asset has a higher expected return.

Now, what happens if investors all decide that the expected return—the return considering both the exchange rate and the interest rate—is higher on German bonds than on U.S. bonds? In this case, U.S. investors will sell U.S. dollars and purchase euros and then use the euros to purchase German bonds. The investors will keep doing this until they no longer perceive German bonds as having a higher return than U.S. bonds. That is, investors keep buying German bonds until prices adjust. In this example, there are three

prices: the two interest rates and the exchange rate. Here, the euro would appreciate because the demand for euros is rising; the yields on German bonds would fall; and the yields on U.S. bonds would tend to rise.

Interest rate parity is one of the key equilibrium relationships in international economics: The foreign-exchange market is in equilibrium when deposits of all currencies offer the same expected risk-adjusted rate of return. Interest parity is expected to hold except when countries prevent the free flow of assets. If, in the example above, German and U.S. residents could only buy their own domestic bonds, interest parity would not necessarily hold. The return on the bonds would be determined independently in each country. This issue is revisited in Box 7-4 later in the chapter.

Purchasing Power Parity

The last section focused on the influence the return on a country's assets tends to have on the country's exchange rates. Purchasing power parity is a second equilibrium concept that also helps determine exchange rate. PPP also relies on the concept that prices (and returns) must be consistent internationally. At a weekend farmer's market, the price of corn cannot vary too much between any two vendors. If there is a large difference in price for the same corn, most of the corn sales will be at the cheaper booth. In other words, people at the market, perceiving the corn to be the same quality, will tend to buy from the cheaper vendor until either that vendor's supply is exhausted or the prices at the two booths adjust so that they are closer together. Purchasing power parity is an extension of this simple concept on a global scale. That is, prices of goods sold in any two countries should exhibit about the same price once those prices are converted to a common currency. If goods are sold for different prices in different countries, then either the prices of those goods or the exchange rate would be expected to change until the exchange-adjusted prices in the two countries were similar.

An example may clarify how this process works. Imagine a farmer's market with three booths. One booth sells corn in U.S. dollars, the second booth sells identical corn in euros, and a third booth (the foreign-exchange market) sells and buys euros at a posted price. A buyer arrives at the market with a single U.S. dollar in his pocket and wishes to buy corn. The prices are as follows: 1 bushel of corn from the U.S. dollar booth sells for $1, the same bushel of corn sells for 1 at the euro booth, and at the foreign-exchange booth $1 can buy 1.1. Therefore, the buyer finds that euro-corn is cheaper; he exchanges his U.S. dollar for 1.1 and is able to buy 1.1 bushels of corn instead of the 1 bushel he could buy at the dollar market. If nothing changes, all buyers who show up to the market will prefer euro corn. In response to this preference, two things are likely to occur. The price of corn at the dollar booth will begin to fall as the dollar booth sees less demand for its corn, and the

foreign-exchange booth will raise the price of euros relative to dollars as it perceives an increased demand for euros.

The above example is quite stylized; however, the economic forces in the global marketplace work in exactly the same way. Buyers and sellers search for the best location to sell their goods. However, unlike asset markets in which the adjustments can happen on a large scale very quickly, purchasing power parity depends in part on the adjustment of goods markets, which tend to take place over a relatively long period of time. Therefore, purchasing power parity tends to hold over a very long time horizon—months and years rather than day-to-day. In addition, because there are real costs to shipping goods internationally, very small differences in purchasing power parity will not necessarily disappear. Nevertheless, purchasing power parity is a powerful concept.

Fixed versus Floating Exchange Rates

The previous sections assumed that currency values could immediately adjust as the demand for either goods or assets changes. In reality, some countries do not allow the value of their currencies to fluctuate. Instead, by systematically changing the supply of their own currency through changes in monetary policy, they control the changes in the value of their currencies and limit exchange-rate movements. The choice of exchange-rate policy is often called the exchange-rate regime of a country. This section discusses the two most basic categories of exchange-rate regimes, fixed and floating. Defining a country's exchange-rate regime is, in practice, not an easy task. For example, in 2004 the International Monetary Fund (IMF) identified eight distinct exchange-rate regimes (see Box 7-3). Using the simplification of fixed versus floating allows a simpler discussion of the links between the exchange rate and monetary policy, a topic discussed in the next section.

Floating Exchange-Rate Regimes

Floating exchange-rate regimes are regimes in which the government takes no action to influence the exchange rate. Under this regime, the exchange rate is completely determined by the general market forces discussed above. One advantage of a floating exchange rate is that the government does not have to have any knowledge over what the correct or true exchange rate should be. Market forces drive the exchange rate toward its true value.

Over the past 25 years, there has been a general trend away from fixed exchange rates and toward floating exchange rates. Chart 7-2 illustrates this general trend. The chart shows that the number of countries using floating exchange rates has risen gradually over time. In 1980, over 75 percent of the

Box 7-3: A Description of the IMF Classification of Exchange-Rate Regimes

Exchange arrangements with no separate legal tender: A country gives up its own currency and allows the currency of another country to circulate as the sole legal tender. This exchange regime is often referred to as *dollarization*. This classification includes countries, such as members of the euro area, that form *currency unions:* arrangements by which the same legal tender is shared by the member countries.

Currency board arrangements: An exchange-rate regime in which a country commits to exchange domestic currency for a foreign currency at a preannounced price. Currency board arrangements feature restrictions on the nation to ensure that it will abide by its legal obligation.

Conventional fixed peg arrangements: A regime in which a nation announces that it will buy or sell its currency in exchange for a foreign currency at a preannounced price. This regime differs from a currency board arrangement only in the legal structure of the regime.

Pegged exchange rates within horizontal bands: A regime in which a country allows only limited movements in the exchange rates. The nation announces a high and a low value for the currency and only agrees to sell the domestic currency at the high price and to buy the domestic currency at the low price.

Crawling pegs: A crawling peg is essentially the same as a pegged exchange rate except that the price at which the currency is traded changes over time. For example, a nation that wishes to allow a long-term appreciation of its currency may choose to do so by adopting a crawling peg that allows the currency to appreciate on average.

Exchange rates within crawling bands: This regime is a combination of a crawling peg and a pegged exchange rate with horizontal bands.

Independently floating: The exchange rate is driven by the market. The country does not attempt to influence the value of the exchange rate. For example, the United States has an independently floating exchange rate.

Managed floating: The exchange rate is driven by the market part of the time but on occasion the government seeks to systematically influence the exchange rate through purchases or sales of the currency.

countries listed in the IMF exchange classification maintained a specific target for their exchange rate. By 2005, this number had dropped to 55 percent.

Even among countries that are considered to be freely floating, the government may occasionally or even periodically intervene in the exchange market. For example, Turkey, listed as freely floating in the IMF classification system, does not have a fixed exchange rate but reserves the right to intervene in the exchange market to limit volatility in its exchange rate (and has done so many times over the past few years).

Fixed Exchange-Rate Regimes

A fixed exchange-rate regime is a regime in which a nation's government announces the price at which its currency will trade for another currency. To maintain the exchange rate, the government must stand ready to buy or sell unlimited quantities of currency at the preannounced price. To keep the exchange rate from appreciating, the government sells its domestic currency in exchange for foreign currency. The increased supply of the currency lowers the value of the currency. To keep the exchange rate from depreciating, the government buys its domestic currency using foreign currency. To make these transactions, the government must have sufficient supplies of both domestic and foreign currency. Maintaining a supply of domestic currency is simple, as

Chart 7-2 Exchange Rate Policy Has Moved Away from Fixed Rates
Countries as a percentage of total
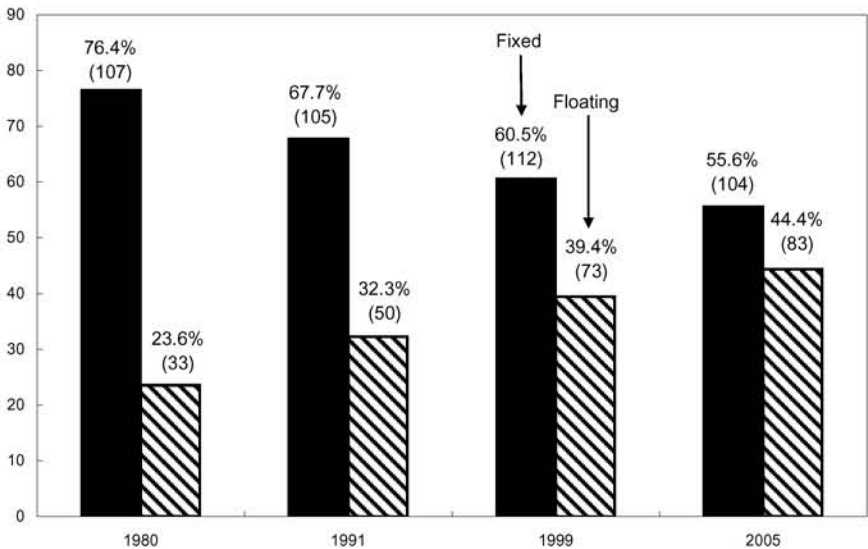
Note: The number of countries is listed in parenthesis.
Source: International Monetary Fund, *International Financial Statistics, Annual Report on Exchange Arrangements and Exchange Restrictions.*

the government has the right to print unlimited quantities of its own currency. However, supplies of foreign currency must be held in reserve and the government does not have the option of increasing its supply. The possibility of running out of foreign currency and being unable to keep the currency from depreciating is one of the reasons that many nations have given up fixed exchange-rate regimes.

Fixed exchange rates have been used by a large number of countries and for a large portion of modern economic history. Following World War II, the major industrialized countries agreed to fix the value of their currencies with respect to each other. This agreement was known as the Bretton-Woods agreement, and the IMF was established in 1949 to monitor this system of exchange rates. To a greater or lesser degree, this system remained in place until the early 1970s, when countries began to allow their exchange rates to drift.

Following the breakdown of the Bretton-Woods agreement, the Western European nations joined together in a fixed exchange-rate regime. After suffering several major exchange-rate crises, 12 of the European nations preferred so strongly to maintain a fixed exchange rate that they agreed to give up their national currencies and the euro area was established. By giving up their national currencies and forming a monetary union, the member nations hope to avoid future crises. While the euro area is still relatively young (it was formally established in 1999), the currency union has not yet suffered a major crisis.

The Links Between Monetary and Exchange-Rate Policies

A nation's choice of exchange-rate policy is tightly linked to a nation's choice of monetary policy. They are tightly linked because exchange-rate policy is a form of monetary policy. Monetary policy, broadly defined, is the policy that controls the growth rate of the money supply. In order to fix the exchange rate, a government must use its ability to control the money supply to sustain a fixed level of the exchange rate. If the supply of money is dedicated to controlling the level of the exchange rate, it cannot simultaneously be dedicated to controlling inflation. Given the earlier discussion of interest rate parity, the choice of monetary target is essentially a choice between stabilizing domestic prices and stabilizing the exchange rate. If the exchange rate is fixed, then domestic prices, both asset prices and goods prices, must do all of the adjusting.

The increase in the number of economies preferring floating exchange rates and the rise of independent central banks with mandates to maintain price stability is not a coincidence. An increasing number of countries have come to desire central banks charged with maintaining low and stable inflation. To achieve this goal, central banks need a nominal target to automatically stabilize the money supply. Most modern central banks have chosen a domestic short-term interest rate for the nominal target. The short-term

policy rate allows the central bank complete autonomy over choosing the rate of domestic inflation.

The short-term policy rate is not the only nominal anchor available to the central bank, however. The central bank could choose to fix the domestic price of gold or any other commodity. The use of the gold standard has a long and reputable history. A nation's exchange rate with another country can also be used as the nominal anchor for monetary policy. By fixing the value of the domestic currency against another currency, a country essentially adopts the monetary policy of the foreign country; one of the problems of using a strict fixed exchange rate is that the monetary policy of the foreign country may differ from what the central bank would have chosen given complete autonomy. That is, the bank could be forced to print either more or less currency than it would have otherwise chosen.

Thinking through a specific example will help clarify the relationship between exchange-rate policy and overall monetary policy. For a long time, China had a fixed exchange rate with the United States. To maintain its fixed exchange rate, the Chinese government had to stand ready to buy or sell yuan, China's domestic currency, for U.S. dollars at a fixed price. From 2000 to July 2005, this price was set at approximately 8.28 yuan per dollar. Over this time period, Chinese productivity growth was much higher than U.S. productivity growth and Chinese prices on average grew much more slowly than U.S. prices. High productivity growth implies a high return to investment in China relative to the United States. The slow growth of Chinese prices implies that, holding the exchange rate constant, Chinese goods were becoming cheaper relative to goods in the United States. Therefore, both in terms of maintaining interest rate parity and in terms of maintaining PPP, there was pressure for the yuan to appreciate relative to the U.S. dollar. How did the Chinese authorities prevent the appreciation?

The Chinese authorities prevented the appreciation by buying U.S. dollars and exchanging these dollars for yuan. The pressures for appreciation of the yuan implied that the yuan was facing higher demand—that more goods could be purchased for dollars converted to yuan, and investments in China delivered, on average, a higher return. To offset the increase in demand, the Chinese government effectively increased the supply of Chinese assets and decreased the supply of U.S. assets. Chinese foreign-exchange reserves increased from around $150 billion in early 2000 to almost $1 trillion by September 2006, a truly remarkable increase. In other words, the Chinese prevented an appreciation of the exchange rate by effectively printing yuan and using those yuan to accumulate U.S. dollar assets.

By fixing the exchange rate, the Chinese monetary authority is unable to use monetary policy for any other goal. By printing yuan, the Chinese raise the amount of currency in the country, which in turn, holding all else equal, raises the domestic price level, thus raising the economy's inflation rate.

But if they are just printing enough to buy and hold U.S. assets, from where does the domestic price pressure arise? The price pressure arises as the yuan, which are used to purchase the dollar assets, flow back into the Chinese economy. In other words, the prices increase because of foreign demand for Chinese goods. On the surface, this foreign demand appears to arise as a result of the Chinese exchange-rate regime; however, this demand is the same demand which was originally putting pressure on the Chinese exchange rate. At the old prices, there was not enough supply of Chinese goods to meet all of the demand. Because the exchange rate was unable to adjust, the price of Chinese goods had to adjust.

Could the Chinese conduct a monetary operation to lower inflation? To lower inflation, the Chinese would need to remove yuan from circulation, perhaps by selling domestic bonds. This transaction is sometimes referred to as sterilization. The action, however, will tend to raise the value of the currency: the currency would become scarcer as a result of the reduction in supply. As the currency becomes more valuable the foreign-exchange value of the currency would tend to appreciate. Any monetary action the Chinese undertake to reduce domestic inflation tends to undo their exchange-rate intervention (see Box 7-4).

This example also illustrates why the Chinese intervention does not systematically change the relative real prices between the United States and China. Had the Chinese government not intervened, Chinese domestic prices would have remained the same in terms of yuan and become more expensive in terms of dollars through a change in the exchange rate. With the intervention, Chinese domestic prices rose in terms of yuan and became more expensive in terms of dollars even though the value of the nominal exchange rate was unchanged. This outcome occurs any time a country takes actions to fix its exchange rate: fixing the nominal exchange rate does not necessarily have any impact on the relative prices between two countries. In other words, fixing the nominal exchange rate does not tend to move countries away from purchasing power parity. The only effect is that domestic goods prices have to do all of the adjustment since the exchange rate is fixed.

In the end, central banks that choose to fix the value of their exchange rate relative to another currency and central banks that choose to set a short-term interest rate are each choosing a different tool to conduct monetary policy. Economic theory does not dictate a clear preference between the two tools; however, by 2006 no central bank from any major industrialized nation has opted to use a fixed exchange rate, while maintaining their own domestic currency, as a monetary policy instrument. These central banks understandably believe that interest rate targeting, in practice, is a preferred tool in the conduct of monetary policy.

Box 7-4: The Impossible Trinity

A fixed exchange-rate regime forces a country to choose between allowing free flows of assets in and out of the country or restricting the flows in order to preserve independent monetary policy. This choice is forced on countries because only two of the following three policies—free asset flows, a fixed exchange rate, and an independent monetary policy—can be maintained at any point in time.

The underlying reason for this restriction is that free asset flows and monetary policy operations may yield a foreign-exchange value of the currency which is inconsistent with the fixed rate that the government is trying to maintain. The United States, for example, allows free asset flows and maintains an independent monetary policy. As a result, the U.S. central bank, the Federal Reserve Board, can influence domestic interest rates relative to foreign rates. If the Federal Reserve elects to raise domestic rates, however, then the United States becomes a more attractive investment environment relative to other countries, and assets flow into the U.S. economy. Because this shift in asset flows raises demand for the U.S. dollar, the exchange rate appreciates. Since the U.S. government lets the market determine the dollar's foreign-exchange value, the dollar's appreciation can occur without any active intervention by the Federal Reserve.

In this example, the only way to break the direct link between the exchange rate and the interest rate would be for the United States to restrict asset flows. If assets cannot flow into the United States, demand for the dollar does not rise with the increase in interest rates, and the exchange rate does not necessarily appreciate. In other words, one of the key assumptions of interest rate parity—that assets can flow to the location with the highest return—is broken.

Denmark, on the other hand, effectively pegs its domestic currency to the euro and allows free flows of assets, as evidenced by the nearly 632 billion kroner of foreign direct investment in Denmark in 2005 (over 40 percent of Denmark's GDP). By pegging its currency and allowing free asset flows, Denmark essentially loses the ability to independently determine its domestic inflation rate. If Denmark were to alter interest rates so that they deviated from world rates, assets would flow in or out of the Danish economy and lead to a shift in the exchange rate. To correct this shift and maintain its fixed exchange rate with the euro, Denmark would then have to buy or sell kroner, thus negating the interest rate changes it achieved through its monetary policy. In this sense, free asset flows and a fixed exchange rate make an independent monetary policy virtually impossible.

Conclusion

Currency markets facilitate global trade and investment by making it easy for firms and investors to buy or sell the currencies they need to do business globally. In the absence of global currency markets, the benefits of international openness would be nearly impossible to realize—international trade would effectively be reduced to barter arrangements. The growing importance of international trade and investment has been accompanied by an increasing number of transactions in the foreign-exchange markets.

The value of a nation's currency is determined like any other good, service, or asset. The more people demand the currency and the scarcer the supply of the currency, the higher the currency's value. The value of a currency is measured by its purchasing power relative to other currencies. In other words, the value of a currency is measured by its exchange rate with other currencies.

Exchange-rate policy is a form of monetary policy. When a country fixes its exchange rate relative to another country, that country must use its monetary policy to maintain the exchange rate. A country with a fixed exchange rate does not have the ability to use monetary policy for any other purpose, just as a nation which sets a short-term interest rate must devote its monetary policy to achieving that goal.

In addition, the value of a country's currency is in large part determined by the value of that country's goods, services, and assets and the ability of people and firms to freely trade these items across national borders. Any policy that restricts the free flow of these items will lower the value of the currency, in addition to lowering the value of the restricted asset. The value of a nation's currency is tied to people's ability to move assets and goods. Small changes in a nation's openness to trade and investment will likely have a small impact on the value of the currency; however, every movement towards more protectionist policies is likely to be associated with a lower value of a nation's currency than would have been true otherwise.

International Trade and Investment

The United States derives substantial benefits from open trade and investment flows. Over many decades, increased trade and investment liberalization has been an important catalyst for greater productivity growth and rising average living standards in the United States.

Trade liberalization and globalization remain controversial subjects because competition invariably raises both anxieties and opportunities. Reducing obstacles to trade can help economies grow more rapidly in the long run and create better, higher paying jobs. Increased competition, however, can lead to hardships for others in the short run. Constructive policies that help displaced workers train for and find new work and increase the portability of pension and health benefits can help to ease adjustment.

The key points in this chapter are:

- Engagement in the global economy through increased trade has contributed to rising average living standards in the United States. Firms engaged in international trade are more productive, have higher employment growth, and are higher wage firms than domestically oriented firms. Looking ahead, international trade liberalization in services presents significant opportunities for U.S. workers, firms, and consumers.
- Foreign direct investment (FDI) flows into the United States benefit the U.S. economy by stimulating growth, creating jobs, and financing the current account deficit. FDI flows into the United States also stimulate investment in research and development in high-technology areas that promote innovation and competitiveness.
- U.S. direct investment abroad is an important channel of global market access for U.S. firms. U.S. multinational companies have contributed to productivity growth, job creation, and rising average living standards in the United States.

Trade Liberalization: A Key Contributor to the Strength of the U.S. Economy

Increased international trade has raised real incomes, restrained prices, introduced greater product variety, spurred technological advances and innovation, and raised living standards in the United States. Studies have estimated that the annual payoff from U.S. trade and investment

liberalization to date, including from the Kennedy Round, the Tokyo Round, the Uruguay Round, the North American Free Trade Agreement and other free-trade agreements, is up to $1.5 trillion. These gains arise through many channels: higher long-term levels of commerce in goods and services that come from trade and investment liberalization; increased product variety; more efficient allocation of resources; and better transportation and communication technology. Some economists have conjectured that trade liberalization alone has accounted for about half of these gains, which implies that the annual income gain from trade liberalization to date is over $2,500 per capita, or $10,000 for an average American family of four. Existing studies suggest that U.S. incomes could rise further by approximately $590 billion per year by moving all the way to global free trade in goods and services.

International trade in goods and services exposes firms to foreign competition and reduces their ability to charge high markups above production costs. International trade also increases the variety of goods available such as silk sweaters from China, wine from Australia, and winter blueberries from Chile. Consumers value variety and one study estimated that the U.S. economic value of increased varieties through imports over the past three decades is equivalent to $350 billion per year, or 2.8 percent of gross domestic product (GDP).

Engagement in the global economy through increased trade has contributed to rising average living standards in the United States. Research shows that firms engaged in the international marketplace tend to exhibit higher rates of productivity growth and pay higher wages and benefits to their workers than domestically oriented firms. Economists agree that the most important determinant of living standards in a country is the average level of productivity, or output per worker.

A free and open international trade regime is vital for a stable and growing economy, both here at home and throughout the world. The United States will continue to work aggressively toward multilateral trade liberalization through the World Trade Organization's Doha Development Agenda negotiations. The prospects for these negotiations to produce significant benefits for this country and our trading partners, particularly developing countries, demand that we promptly reach a balanced and ambitious outcome.

Firms That Engage in International Trade Are Strong Performers

At the microeconomic level, firms engaged in international trade outperform domestically oriented firms on many dimensions. Research has shown

that firms engaged in international trade have higher productivity than their counterparts engaged solely in domestic activity. One study found that value added per employee, one simple measure of productivity, was 15 percent higher in manufacturing exporting firms than in firms that did not export (controlling for industry effects, plant size, and geographic location). And these productivity effects are reflected in higher wages: the wages paid by manufacturing plants that export are 9 percent higher on average than wages paid by non-exporting plants of the same size. Wages in service-oriented firms that export are, on average, 13 percent higher than their purely domestic counterparts of the same size.

One recent study that examined the dynamics of globally engaged firms between 1993 and 2000 found that firms engaged in international trade had a higher survival rate (65 percent) than the average for all firms in the country (53 percent). In addition, a firm that began to trade during this time period increased employment by nearly 100 percent on average, while a firm that quit trading experienced a decline in employment.

An increasing number of American workers are employed by firms engaged in international trade. Between 1993 and 2000, firms that trade increased employment by 9.8 million workers, and the share of the American workforce employed by a firm engaged in trade increased from 40 percent to approximately 42 percent. Applied to today's workforce, this result implies that over 57 million American workers are currently employed by a firm that engages in international trade.

The Effects of Nontariff Barriers on International Trade

While trade can generate many economic benefits, governments at times set up barriers to international trade. One of the more common and harmful barriers is a *nontariff barrier*, a barrier behind the border that is a policy (other than a tariff or tax) or official practice that can unfairly inhibit competition. Unjustified nontariff barriers can distort the prices and quantities of goods and services traded internationally, restrict international investment, and reduce economic welfare in exporting and importing countries. As tariffs have fallen both in the United States and in many other countries, nontariff barriers have increased in importance and are often cited as more trade-restricting than tariffs. Nontariff barriers can arise as a result of government policies aimed explicitly at protecting domestic firms from international competition, or from rules or laws within a country that effectively hinder trade (see Box 8-1).

Box 8-1: Nontariff Barriers Restrict Trade

Unjustified nontariff barriers (NTBs) make it more difficult for international goods and services to compete freely and fairly with those produced domestically. Common examples of NTBs are burdensome or nontransparent product standards or regulations. For example, in Korea, pharmaceutical imports must be tested on Korean nationals, and each individual batch produced must undergo testing. In China, the process of standards certification for telecommunications and IT products can be burdensome and unpredictable, as two separate Chinese regulatory agencies each check for conformity to the same set of standards. Other often-cited NTBs include investment restrictions, government procurement laws, and lax enforcement of intellectual property rights.

Measuring the effects of NTBs on trade is more difficult than assessing the effects of tariffs, but some attempts have been made. A growing body of evidence consistently shows that the economic welfare gains from eliminating NTBs are at least as large as those obtained from further tariff liberalization. One study shows that the U.S. payoff from eliminating NTBs with just seven of our trading partners (Australia, Canada, Germany, Italy, Japan, the Netherlands, and Great Britain) would generate annual income gains of $90 billion for the United States (0.72 percent of GDP), compared with $37 billion from tariff liberalization (0.30 percent of GDP). These benefits arise largely from the pro-competitive effects of increased international trade and more efficient allocation of resources.

Tariff negotiations are fairly straightforward, and forums such as the World Trade Organization (WTO) exist for this purpose. Members are required to report their tariff schedule to the WTO each year, so members know the tariff rate for each product in every country. However, countries do not always agree on what constitutes a NTB and there is no formal, consistent notification process, thereby making negotiations aimed at addressing such barriers more complicated. Part of the policy problem is making distinctions as to whether NTBs are warranted for nontrade reasons (e.g., product safety standards) or whether they are simply covert barriers to imports (nontransparent licensing requirements for foreign firms). For instance, customary regulatory and legal procedures within one country might be seen as complex and overly burdensome to would-be exporters.

Apart from the challenges of identifying NTBs, policymakers face difficulties in knowing which NTBs they should seek to dismantle first. The U.S. Department of Commerce has surveyed its industry and trade experts and country desk officers in an effort to identify the most prevalent NTBs faced by U.S. exporters and to identify which export products

are most likely affected. The survey results suggest that, on average, at least one NTB affects U.S. exporters for each major product category in which they export to our main trading partners. For instance, a problematic regulatory environment was cited as a problem in 43 of the 49 countries covered by the survey, and was cited as the top problem in 14 of those countries. The industries facing the most NTBs included entertainment, pharmaceuticals, and information technology.

International Trade in Services

Liberalizing trade in services is important for economic growth here and abroad. As an economy grows and matures, services tend to increase as a share of GDP and as a share of trade. The United States has a global competitive advantage in services, yet services remain highly protected abroad.

Services such as financial, insurance, transportation and storage, telecommunications, express delivery, and business services generate 68 percent of world GDP but account for just under 20 percent of global trade. While global advances in information and communications technology are making services increasingly tradable, existing trade barriers to services are significant. These barriers are currently subject to negotiation in a host of bilateral, regional, and multilateral trade talks.

U.S. Competitive Advantage in Services

A large and growing part of the U.S. economy and workforce is employed in services. In 1800, 9 out of 10 American workers were employed in agriculture; today that number is less than 1 in 10 (Chart 8-1). In contrast, nearly 8 in 10 American workers are employed today in the service sector.

The vast economic benefits from trade liberalization for services stem in part from our competitive advantage in services. That is, the United States can produce many services at a lower cost than our trading partners, and our trading partners can produce some other set of goods and services at a lower cost than the United States. When we trade our lower cost services for their lower cost goods, we and our trading partners gain from trade. Chart 8-2 shows the changing structure of U.S. trade, which in part mirrors the changing structure of the U.S. economy. Since the 1970s, the United States has consistently run a surplus in services trade, with a $66 billion surplus in 2005.

Chart 8-1 Percent of Private U.S. Workforce by Sector, 1800–2005

A large and growing share of the private U.S. workforce is employed in services.

Percent of U.S. workforce
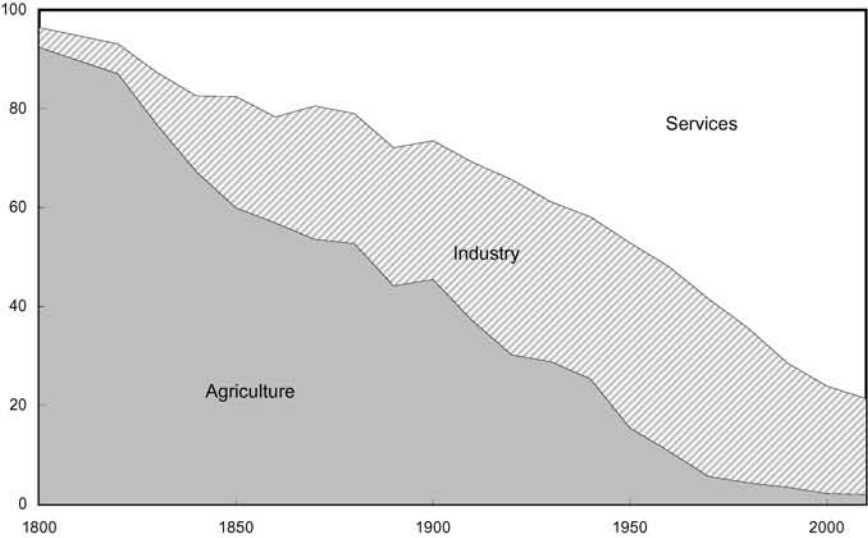
Note: Industry includes manufacturing, construction, and mining.
Source: Federal Reserve Bank of Dallas.

Chart 8-2 Trade Balance by Sector, 1970–2005

The U.S. trade deficit in goods and surplus in services have increased.

Billions of dollars
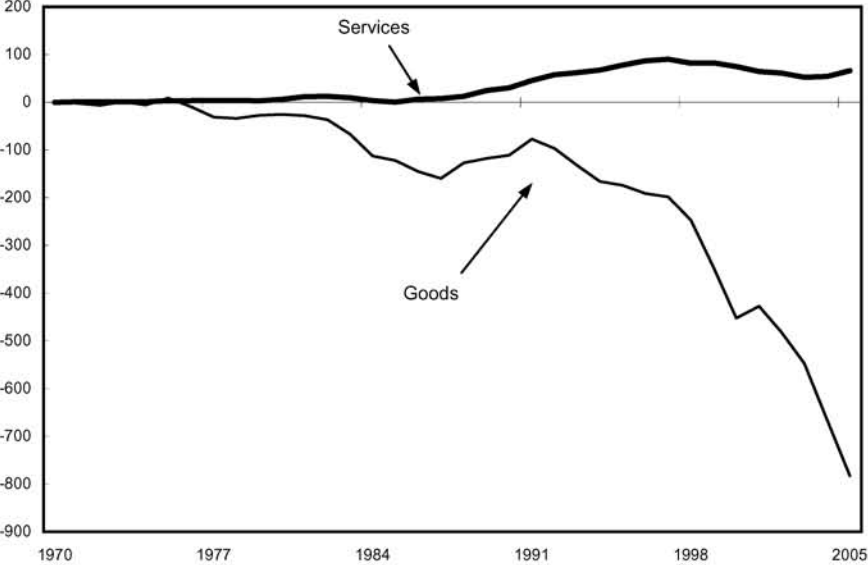
Source: Department of Commerce (Bureau of Economic Analysis).

Technological Change Is Fostering International Trade in Services

Services have become increasingly tradable, particularly knowledge-based or information technology-enabled services that are beyond the traditional notion of internationally traded services such as transportation, travel, and tourism. For many of these services, a physical commercial presence is necessary. For example, a financial institution is able to offer a host of financial products to international clients, but the multinational firm must still set up intermediary branches to serve their clients overseas. Other services can be delivered with virtually no physical presence. An increasingly wide range of commercial transactions ranging from stock trades, to manufacturing orders, to airline reservations, can occur almost entirely over networked digital media located in many countries around the world.

Trade in services previously involved high transaction costs between businesses and customers. Technological innovations and changes in global technology such as the Internet, information technology (IT) hardware such as personal computers, and IT networks have greatly reduced communication and transaction costs for trade in services.

Table 8-1 reports U.S. trade in private services. The largest subcategories in "other private services" trade, which captures many of the IT-enabled services, include financial and insurance services; computer, management, and consulting services; and other business, professional, and technical services.

TABLE 8-1.— *U.S. International Trade in Private Services, 2005*
(billions of dollars)

Total private services traded	Exports	Imports	Balance
Total private services traded	**$360.5**	**$280.6**	**$79.9**
Travel	81.7	69.2	12.5
Passenger fares	20.9	26.1	-5.1
Other transportation	42.2	62.1	-19.9
Royalties and license fees	57.4	24.5	32.9
Other private services	158.2	98.7	59.5
Education	14.1	4.0	10.1
Financial services	34.1	12.3	21.7
Insurance services	6.8	28.5	-21.7
Telecommunications	4.7	4.7	0.1
Business, professional, and technical services	80.8	47.7	33.1
Computer and information services	8.2	9.0	-0.7
Management and consulting services	6.4	5.9	0.5
Research and development and testing services	10.1	6.7	3.4
Operational leasing	9.4	1.3	8.1
Other business, professional, and technical services	46.6	24.8	21.8
Other services	17.7	1.5	16.2
Film and television tape rentals	10.4	0.9	9.5
Other	7.3	0.6	6.7

Source: Department of Commerce (Bureau of Economic Analysis).

Trade growth in "other private services" has far outpaced growth in the rest of services. From 1995 to 2005, U.S. exports of "other private services" grew 143 percent, compared with 44 percent growth in all other services. The bulk of the overall trade surplus in services comes from the "other private services" category, which accounted for 90 percent of the *overall* U.S. services trade surplus in 2005, up from 38 percent in 1995. In contrast, the surplus in more traditional services (e.g., travel and transportation) has fallen. The surplus in "other private services" has grown from $30 billion in 1995 to $60 billion in 2005, and the surplus in the rest of services has fallen from $48 billion to $7 billion. Many of these trends are consistent with the global IT advancements that have fostered international trade in services over the past decade.

High Barriers Restrict International Trade in Services

Barriers to trade in services are mostly regulatory and investment restrictions and tend to be higher than trade barriers in merchandise. For instance, U.S. banks that wish to offer retail banking services abroad face a host of barriers that limit their ability to compete in foreign markets. Examples of such barriers might be investment restrictions that limit the number of bank licenses the country will issue to a U.S. bank; requirements for U.S. banks to enter the banking market through a joint venture with a domestic bank; or limits on the degree of control that a U.S. bank can exercise over its foreign affiliate. Foreign firms wishing to enter the U.S. airline industry face ownership restrictions that limit their ability to compete with domestic firms.

Despite such barriers, services trade is expected to continue to grow. Research suggests that as countries' incomes grow, their demand for services and their trade in services will each grow more than one-for-one with income. U.S. producers are well-positioned to continue to engage in increased services trade, as many have already incorporated the technology in their operations to facilitate trade.

Looking Ahead to Larger Gains from Trade Liberalization

Despite decades of trade liberalization, the world economy is still far from a global marketplace of unfettered trade. Many of the remaining barriers lie in services, and the prospective gains for the United States from further trade reform are substantial. While global tariff liberalization in manufacturing and agriculture could generate over $16 billion in income for the United States each year, the prospective gains from services liberalization are immense: an estimated $575 billion in annual U.S. income (4.3 percent of GDP). Summing up, this is an additional $591 billion in annual income that will be foregone in the absence of further trade reform.

The magnitude of the payoff to the United States from services trade liberalization reflects a number of factors: the U.S. competitive advantage in many services, the large share of services in the global economy compared to the relatively small share of services in global trade, and the high barriers to services trade. These barriers are often regulatory in nature or involve restrictions on the form of investment, such as foreign equity restrictions that limit foreign investors' holdings and control in a company, transfer limitations on capital flows, and the repatriation of profits. Removing these barriers would free up capital to move across borders to the location with the highest rate of return.

Developing countries also stand to benefit greatly from global liberalization of services trade. The service sector share of GDP exceeds the manufacturing share in most developing countries. The increased availability and quality of services enhances the competitiveness of manufactured goods, agricultural products, and existing services. For instance, India stands to gain an estimated $12 billion in national income each year (1.7 percent of GDP) from removing barriers to trade in services, and China stands to gain an estimated $105 billion (4.0 percent of GDP) each year.

Foreign Direct Investment

International trade in goods and services is an important channel of international commerce, but it is not the largest channel. For many U.S. firms, foreign direct investment (FDI) is a more significant path to accessing foreign markets than are exports.

FDI is investment of foreign assets into domestic structures, equipment, and organizations (e.g., a manufacturing plant, an R&D facility, an office or a warehouse), whether in the form of acquisition or "greenfield" establishment. FDI is distinguished from passive portfolio investment (FDI does not include foreign investment in the stock market). Only the former can confer managerial or operational control. The two types of foreign direct investment are inward FDI and outward FDI. *Inward foreign direct investment* is generally understood to imply ownership by a foreign person or corporation of at least a 10-percent stake in a U.S. business enterprise. Similarly, *outward foreign direct investment* is ownership by a U.S. person or corporation of at least a 10-percent stake in a foreign business' operation abroad. A foreign automaker building or buying a production plant in the United States is an example of inward FDI, while a U.S. automaker building or buying a production plant in China is an example of outward FDI.

Before we examine each type of FDI and its importance to the U.S. economy, it is useful to define some of the terms that are commonly encountered when discussing FDI. A *multinational corporation* is a business enterprise

(i.e., the parent) headquartered in one country that has at least a 10-percent ownership stake in a foreign business enterprise (i.e., the affiliate) in another country. That 10-percent ownership stake is the minimum stake used by many statistical agencies around the world, including those in the United States, for identifying meaningful managerial influence over the affiliate.

A *majority-owned U.S. affiliate* is an affiliate of a foreign-owned company that is located in the United States and has at least 50 percent foreign ownership (we focus on majority-owned U.S. affiliates here but use the term "U.S. affiliates"). Similarly, a *majority-owned foreign affiliate* is a foreign affiliate with at least 50 percent U.S. ownership.

U.S. firms are more reliant on FDI for the international delivery of services than they are for the international delivery of goods. While services are becoming increasingly tradable, their actual delivery often requires some physical presence, for example, distribution and express delivery services. Even with widespread use of ATMs and electronic banking, financial or retail banking often requires physical presence in the country in which services are being offered. Based on data from the Bureau of Economic Analysis for 2004, the ratio of sales by U.S.-owned services affiliates abroad to total U.S. services exports was 5.5, compared to 2.5 for goods. That is, U.S. firms deliver over five times the value of services through their foreign affiliates as they do through cross-border trade. Similarly, U.S. firms deliver 2.5 times the value of goods through their foreign affiliates as they do through cross-border trade.

Contributions of Inward FDI to the U.S. Economy

The United States receives inward FDI from firms and individuals located in countries from all over the world. Countries with the largest FDI positions in the United States include Great Britain, Japan, Germany, and Canada. These funds support firms across the U.S. economic landscape, from food, mining, and manufacturing firms to service sectors such as finance, telecommunications, and wholesale and retail trade. Every state in the United States is a recipient of foreign direct investment.

Presence of U.S. Affiliates

Decades of trade and investment liberalization both here and abroad have encouraged the growth of multinationals and global supply chains. Today, U.S. affiliates of foreign multinationals account for an important part of the U.S. economy. In 2004, the latest year for which data are available, U.S. affiliates owned $5.5 trillion in assets and had $2.3 trillion in sales. They produced $515 billion of goods and services inside the United States and accounted for 5.7 percent of total U.S. private output—up from 3.8 percent in 1988. U.S. affiliates employed 5.1 million workers or 4.7 percent of the

U.S. workforce in 2004—up from 3.6 percent in 1988. While historical data show upward trends in the presence of U.S. affiliates, since 2000 U.S. affiliate investment, output, and employment have leveled off or decreased slightly.

Microeconomic Benefits to the U.S. Economy

Inward FDI provides a number of benefits to the U.S. economy at the microeconomic level. Research has shown that multinationals are more productive than firms focused primarily on domestic markets. The relatively high productivity of U.S. affiliates of foreign-owned firms is attributable, in part, to their relatively high levels of investment in physical capital, R&D, and exporting and importing. Specifically, while U.S. affiliates account for 5.7 percent of output and 4.7 percent of employment, they account for a disproportionately high share of U.S. exports (19 percent), imports (26 percent), physical capital expenditures (10 percent), and R&D expenditures (13 percent) (see Chart 8-3). Studies show that all of these activities are correlated with strong productivity performance. (Chapter 2 discusses productivity growth and long-run effects on the standard of living.)

Chart 8-3 **Economic Activities of U.S. Affiliates of Foreign Companies (2004)**
U.S. affiliates account for a high share of U.S. trade, capital expenditures, and R&D expenditures relative to output and employment.
Percent of total for U.S. economy
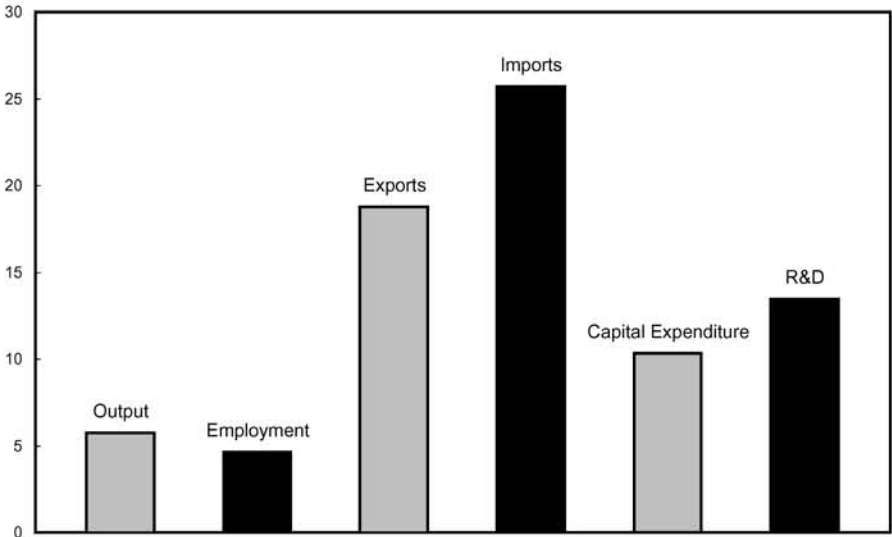
Sources: Department of Commerce (Bureau of Economic Analysis), Department of Labor (Bureau of Labor Statistics), National Science Foundation.

At the firm level, U.S. affiliates pay higher compensation (wages and benefits) on average than their counterparts in the rest of the U.S. economy. In 2004, an average U.S. worker employed by a U.S. affiliate of a foreign-owned firm received $63,400 in annual compensation compared to $48,200 for workers in the rest of the economy. Research suggests that this difference is largely attributable to above-average labor productivity at U.S. affiliates. Part of this productivity advantage reflects these firms' ability to integrate production processes across borders and their organizational efficiency. Another part reflects differences in plant size, capital intensity (that is, higher use of capital relative to other factors, such as labor, in the production process), and employee skill level. The data also suggest that these firms have higher levels of efficiency (how well labor and capital inputs are used), the gains of which are passed on, in part, to workers. In other words, firms can break up their production process across borders to lower average costs and realize increased productivity and revenues, which can be shared with workers through higher compensation and/or captured by firm owners as higher profits (see Box 8-2).

## *Macroeconomic Benefits to the U.S. Economy*

Inward FDI provides a number of benefits to the U.S. economy at the macroeconomic level. For instance, inward FDI is an additional source of investment that helps to modernize the U.S. capital stock. Another benefit is that it provides a source of financing for the U.S. current account deficit, which measures net flows of goods and services between the United States and the rest of the world. As the United States continues to run a current account deficit, foreigners continue to accumulate U.S. assets, and inward FDI is one of the main ways in which they do so.

The accumulation of FDI flows over a period of time results in a stock of assets, or the gross foreign investment position. In 2005, the inward FDI position at market value totaled $2.8 trillion and was the largest component of foreign holdings of U.S. assets. Other components were U.S. Treasury securities ($2 trillion); corporate stocks ($2.1 trillion); and corporate and other private bonds, excluding official holdings ($2.3 trillion) (see Chart 8-4).

The share of foreign holdings is not concentrated in any particular class of assets, which implies a general broad-based confidence in the U.S. economy. Inward FDI is generally considered to be the most stable among the four types of assets shown in Chart 8-4—that is, the least subject to sudden withdrawal. FDI flows are generated by long-term risk–return considerations and are far less liquid and less reversible than portfolio investments. Therefore, FDI flows provide stability to U.S. capital flows because they are not easily reversed for short-term considerations.

Chart 8-4 Foreign Investment Position in the U.S. by Asset Type (2005)
Inward FDI (at market value) was the largest component of foreign holdings of U.S. assets in 2005.

Trillions of dollars
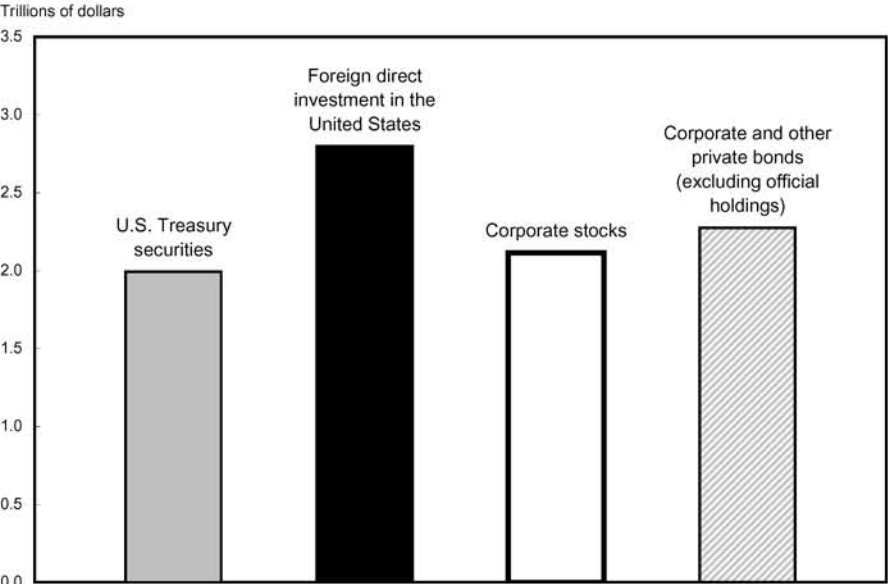
Source: Department of Commerce (Bureau of Economic Analysis).

Box 8-2: Multinationals Bring New Products and Processes to the Host Country

The benefits to the U.S. economy from inward FDI mirror those of many other countries. A growing body of evidence across countries and industries demonstrates that globally engaged firms tend to be strong performers—such firms are more productive, pay higher wages, and generate beneficial productivity side effects that accrue to domestic competitors. The three case studies that follow provide a snapshot of the benefits of inward FDI.

Increasing Living Standards in the United States
Infineon Technologies of Munich, Germany, built a state-of-the-art manufacturing plant in Richmond, Virginia, using leading-edge technology to produce dynamic random access memory products that are used in computers. The Richmond company's annual payroll exceeds $100 million, with average wages that are nearly double average Virginia salaries. Over 3,000 North American workers are employed by

continued on the next page

Box 8-2 — *continued*

this German-headquartered multinational, with over 1,750 workers in Richmond alone. The firm has built extensive ties with its customers and suppliers worldwide, and many advanced technology suppliers have emerged in Virginia to support Infineon and other semiconductor firms. Semiconductors are now Virginia's second largest export.

Enhancing Productivity for Mexican Producers and Retailers

One case study documents impressive efficiency gains for Mexico's domestic soap producers once Wal-Mart entered its retail sector. Wal-Mart helped improve Mexico's retail sector by improving the way Mexican retailers interacted with their suppliers. These changes brought about efficiency improvements such as modernization of warehousing, distribution, and inventory management; triggered greater use of information technology in supply management; and required delivery trucks to have appointments and drivers to carry standard identification cards. These innovations have been adopted by other retailers and producers outside of Mexico's soap industry. Mexican soap producers improved their productivity and have gained market share in key export markets, including in the United States.

Improving Banking and Telecommunication Services for Czech Manufacturers

The change toward a freer and more open investment climate in the Czech Republic was followed by the entrance of foreign-owned banks and telecommunication firms. These foreign-owned service providers helped to improve the availability, range, and quality of services. These improved services contributed to better performance of Czech manufacturing firms that rely on services as inputs. For instance, foreign banks accelerated the processing of loan applications, offering decisions to small and medium Czech enterprises within 2 days, compared to a previous waiting period of several weeks. Foreign banks were among the first to offer Internet and remote banking services, including ATMs, which save individual customers and business clients days and sometimes weeks in transaction times. The time needed to send a fax went from hours (or sometimes days for rural areas) to just minutes following the liberalization of the telecommunication sector.

Is Inward FDI on the Decline?

The increase of inward FDI since the late 1980s has coincided with the generally solid performance of the U.S. economy, along with a surge in U.S. worker productivity that has occurred since 1995. Recently, however, some trends have developed with respect to FDI in the United States that may be cause for concern. First, while the U.S. affiliate share of U.S. output has grown over the past two decades, it has stagnated and even declined in recent years. Second, the U.S. affiliate share of employment has declined, from 5.1 percent in 2000 to 4.7 percent in 2004. Third, the share of inward FDI in the U.S. capital account—that is, FDI in the United States as a share of all the assets owned by foreign interests—has declined since 1999. It is not yet clear whether these are benign and temporary trends or whether this development is symptomatic of deeper issues with respect to the attractiveness of the United States as a country in which to make direct investment. To ensure that inward FDI remains a strong, positive force in the U.S. economy, foreign investors in the United States must continue to receive fair and equitable treatment as a matter of both law and practice.

Historically, the United States has opposed the use of government actions that distort, restrict, or place unreasonable burdens on foreign investment. No property can be expropriated pursuant to U.S. law unless it is done for a public use with payment of just compensation. The United States has historically provided a domestic environment conducive to investment by providing foreign investors fair and equitable treatment based on the national treatment principle: foreign investors should be treated no less favorably than domestic investors in like circumstances. Moreover, while taking every necessary step to ensure that foreign investments do not jeopardize national security, the Administration recognizes that our economic vitality depends on our openness.

The Contributions of Outward FDI to the U.S. Economy

A U.S. multinational company is headquartered in the United States and, through outward FDI, has affiliates (often production or marketing facilities) in other countries. Activities of U.S.-headquartered multinationals have contributed strongly to productivity growth in the United States, and thus to rising U.S. living standards.

Because multinationals are engaged in cross-border investment and production networks, they are better able to enhance their organizational efficiency. Studies have shown that multinationals are more productive than firms that are focused primarily on domestic markets. By combining domestic production with foreign production, multinationals can produce at lower costs, earn

higher profits, and pay higher wages and benefits. Domestic firms can benefit from outward FDI as multinationals are exposed to the world's best business practices that can be adopted by other U.S. firms.

Basic Facts About U.S. Multinational Companies

U.S. multinationals are relatively small in number but have a disproportionately large economic footprint. Less than 1 percent of U.S. firms are multinationals, but these multinationals account for 20 percent of total U.S. employment and 25 percent of total U.S. output. In 2004, there were 2,369 U.S. multinationals with 22,279 foreign affiliates, with 21.4 million employees in the United States and 9 million workers abroad. The operations of U.S. multinationals are concentrated in the United States. In 2004, the combined value-added output of U.S. multinationals was $3.04 trillion. U.S. parents accounted for over 70 percent of this output and foreign affiliates for less than 30 percent.

While U.S. multinationals have increased employment and output in an absolute sense, their share of the workforce has decreased slightly over the years while their share of output has remained fairly constant. U.S. multinationals employed 18.7 million American workers, or 25 percent of the workforce, in 1982 (the first year for which annual employment data are available). In 2004, those figures stood at 21.4 million workers and 20 percent, respectively. The value of output by U.S. parents was $1.3 trillion or 24 percent of the total private U.S. output in 1994 (the first year for which annual output data are available). In 2004, those figures were $2.2 trillion and 25 percent, respectively. In terms of recent trends, both employment and output by U.S. parents peaked in 2000 and then began to decline. Output rebounded in 2003 and employment rebounded in 2004, largely reflecting economy-wide trends.

Why Do U.S. Firms Become Multinational?

There are three conditions required for a firm to be willing to invest abroad: (1) the firm has specific assets that can be transported to foreign affiliates; (2) the host country has certain characteristics that make it attractive for the firm; and (3) the firm wishes to maintain control over its intellectual assets.

Multinationals often face large costs and barriers to doing business abroad compared with domestic firms in the host country that are familiar with the local business climate. Physical and human capital are needed to establish an affiliate, and additional resources are needed to understand the local business environment (for example, regulations and tax laws, supply networks, cultural differences, and property rights). Thus, a multinational firm must have certain advantages to compensate for these costs. Three types of compensating advantages are commonly cited. One advantage is firm-specific resources or knowledge-based assets and services (such as technology, patents, trademarks,

and managerial or engineering expertise) that can be used by the foreign affiliate. Another advantage is the location and characteristics of the host country such as market size, trade costs, and differences in the prices for key inputs such as land, labor, or capital. The existence of a large market or the high costs of trading with a certain country or region can motivate multinationals to produce and sell in foreign countries. Price differences in land, capital, or labor; transportation and telecommunications infrastructure; or good business practices can also motivate a multinational to invest and produce abroad.

The third type of advantage is known as *internalization advantage*. A firm may choose outward FDI over giving a foreign company a license to produce its goods so that it can retain control of its intellectual assets. For example, a firm may be reluctant to reveal the details of its product's construction or its production process to a prospective licensee. There is also the danger that a licensee may produce a lower quality product and consequently reduce the value of the multinational's trademark. The difficulty of guaranteeing quality control, monitoring and managing employees, achieving a satisfactory licensing agreement, and enforcing patent or trademark rights all tend to favor outward FDI.

The Organization of Multinational Production

There are two main organizational strategies for multinational production. One strategy is *vertical FDI,* whereby the multinational geographically fragments the production process and carries out different stages of production at different locations. In contrast, *horizontal FDI* occurs when the multinational conducts the entire production process in the host country to sell locally through its affiliates.

Vertical FDI establishes cross-border production networks. A multinational firm may perform many activities—for example, R&D, assembly, marketing, and sales—that require different mixes of capital, more- or less- skilled labor, land, and other inputs. Separating these activities across borders (and across the parent company and affiliate companies) enables the firm to locate each activity in countries with relatively low costs for each activity's intensively used inputs. Because each stage of the production process is carried out in the optimal location in terms of the input mix, vertical FDI production networks can allow firms to take advantage of differences in comparative advantages across countries and produce at an overall lower unit cost. Trade between U.S. parents and their affiliates ("intra-firm" trade) has risen over time, accounting for 20 percent of total U.S. goods exports in 2004, and 14 percent of total goods imports.

Horizontal FDI can allow U.S. multinationals better access to foreign markets. Ninety-five percent of the world's consumers live outside U.S. borders. Companies can reach foreign markets through FDI or exporting. But for U.S. multinationals, the predominant mode of serving foreign markets is

through FDI and affiliate sales (producing and selling locally), not exporting. In 2004, U.S. multinationals sold $2.3 trillion of goods abroad through affiliate sales compared to $400 billion through exports (see Chart 8-5). In other words, for every $1 of exports in goods, U.S. multinational firms sold $5.84 through their foreign affiliates, up from $3.40 ten years earlier.

A common allegation is that U.S. multinationals set up production plants to serve as export platforms back to the United States. However, the data do not support this claim. In 2004, sales by foreign affiliates of U.S. multinationals totaled $3.2 trillion. Most of these sales were to customers outside of the United States; 89.6 percent of total sales were to foreign customers and 10.4 percent were to U.S. customers.

Outward FDI Complements Domestic Economic Activity

Studies show that economic activity abroad by U.S. multinationals complements domestic economic activity. One dollar of additional foreign capital spending is associated with $3.50 of additional domestic capital spending. Firms combine home and foreign production to generate final output at a lower cost than would be possible in just one country, resulting in increased output and profits. Further, when multinationals hire abroad, they also expand employment here at home, making multinationals an important force behind job creation in the United States (see Box 8-3).

Chart 8-5 **U.S. Multinational Goods Sales through Foreign Affiliates and Exports (2004)**
U.S. multinationals serve foreign markets primarily through their foreign affiliates.

Billions of dollars
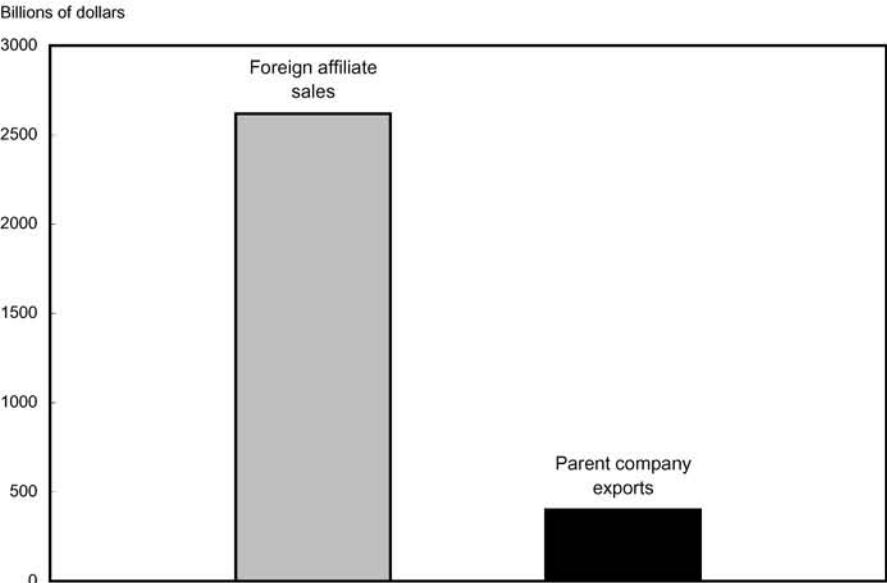
Source: Department of Commerce (Bureau of Economic Analysis).

From a broader perspective, U.S. multinationals enhance U.S. competitiveness by engaging in the same activities and possessing the same characteristics that make the U.S. economy competitive in world export markets. Research has shown that the competitiveness of U.S. multinationals tends to be driven by relatively high levels of R&D and highly skilled labor. Studies have also shown that U.S. firms tend to control larger shares of world markets in industries with high levels of R&D and highly skilled labor. Because their competitive interests largely coincide with broader U.S. economic interests, U.S. multinationals make the economy as a whole more competitive.

Box 8-3: U.S. Multinational Companies and U.S. Jobs

In recent years, many observers have expressed dismay that U.S. companies have expanded their operations overseas, claiming that when U.S. firms hire workers in foreign countries, they reduce the number of jobs available to U.S. workers. The idea that U.S. multinationals hiring abroad are "exporting jobs" relies on at least two assumptions: first, that jobs abroad at foreign affiliates are substitutes for domestic jobs at U.S. parent companies; and second, that when U.S. parent companies expand overseas, they do not change the overall scale or scope of their domestic activities. However, in looking at historical data regarding the activity of U.S. multinationals, we see exactly the opposite: when U.S. companies expand their employment abroad, they also tend to expand domestically.

When U.S. Multinationals Hire Abroad They Also Expand Domestic Employment

Over the last two decades (1984–2004), U.S. multinationals expanded employment at their foreign affiliates by 3.8 million and at their parents by 3.2 million (see chart). In other words, the long-run data show that when U.S. multinationals hire abroad they also expand domestic employment. There have been short-run anomalies to this historical trend that largely reflect economic business cycles both here and abroad. For instance, between 1990 and 2000, for each job U.S. multinationals created abroad they created nearly two at home. Between 2000 and 2003, U.S. multinationals continued to expand employment abroad, albeit at a slower pace, while decreasing their U.S. payrolls. Since 2003, both U.S. parent company and affiliate employment have risen.

One study found that as U.S. companies expand employment abroad, increase their compensation of foreign workers, and invest in their overseas operations, they also increase their hiring, employee compensation, and investment in the United States. Thus, rather than being

Box 8-3 — *continued*

Employment by U.S. Parent Companies and their Majority-Owned Foreign Affiliates
Employment by U.S. parent companies and their foreign affiliates have both grown since the early 1980s.

Millions of employees

Sources: Deparment of Commerce (Bureau of Economic Analysis), Department of Labor (Bureau of Labor Statistics).

substitutes for one another, the domestic and foreign operations of U.S. multinationals have tended to be complements. Consider the operations of General Electric. According to its latest annual report, since 2001 this multinational has expanded foreign employment by 3,000 while also expanding domestic employment by the same amount.

One reason for the complementary relationship between domestic and foreign activity is that a firm may change the overall size of its operations and expand both at home and abroad. Alternatively, a firm may change the scope of its operations and change the mix of its activities (for example, manufacturing, services, or R&D). In fact, it is common for parent companies in one industry to own foreign affiliates in another industry. In 2004, U.S. parent companies primarily engaged in manufacturing owned over 15,000 foreign affiliates, but over 6,500 of these affiliates specialized in areas outside of manufacturing.

In sum, the decision of a firm to expand abroad is based on many factors, and it may be part of a larger overall expansion strategy or a change in the scope of its operations. It is difficult to predict beforehand what such an expansion means for U.S. workers and the U.S. economy. The only way to tell the effect is to examine the data, and thus far the data show that, over the long run, when U.S. multinational firms hire abroad, they also hire at home.

## Good Performance Features of U.S. Multinationals

U.S. multinationals differ from the average U.S. firm in a number of ways. For example, while U.S. multinationals account for 25 percent of total U.S. output and 20 percent of employment, they account for a disproportionately high share of U.S. goods exports (49 percent), goods imports (31 percent), physical capital expenditures (29 percent), and research and development (68 percent) (see Chart 8-6). In fact, U.S. affiliates and multinationals combined conduct over 80 percent of all private sector R&D in the United States. Also, the plants operated by these companies tend to be larger in size than the U.S. average. These differences are important because each of them—international trade, capital expenditure, research and development, and plant size—is associated with high labor productivity. And because of the strong link between labor productivity and employee compensation (see Chapter 2), this higher productivity is a potential benefit to U.S. workers.

U.S. multinationals pay higher average compensation than firms in the rest of the economy. In 2004, U.S. workers employed by U.S. parent companies received an average of $57,800 in annual compensation, compared to about $46,800 for workers in the rest of the economy. The relatively high productivity of U.S. multinationals may be one of the causes for the difference in compensation.

Chart 8-6 **Economic Activities of U.S. Multinational Companies (2004)**
U.S. multinationals account for a high share of trade and R&D expenditures relative to output and employment in the United States.
Percent of total for U.S. economy
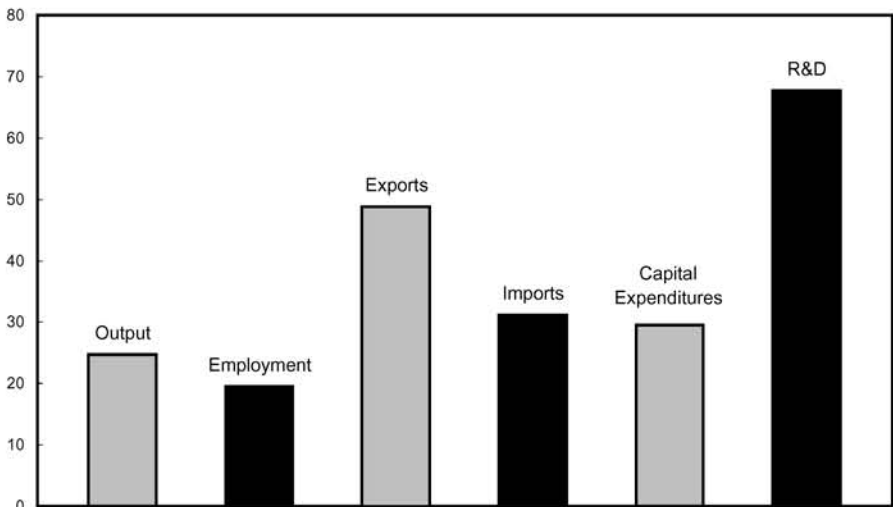
Note: R&D shown is for 2003.
Sources: Department of Commerce (Bureau of Economic Analysis), Department of Labor (Bureau of Labor Statistics), National Science Foundation.

U.S. multinationals have had high productivity growth over at least the last three decades, and because they make up a sizeable part of the overall U.S. economy, they have been one of the main drivers of overall U.S. productivity growth during this period. U.S. multinationals accounted for over half of U.S. productivity growth between 1977 and 2000, and for half of the *increase* in U.S. productivity growth between 1995 and 2000. During this 5-year period, productivity at U.S. multinationals surged, growing 6.0 percent annually.

Conclusion

Engagement in the global economy through increased trade and investment has contributed to rising average living standards in the United States. Further trade liberalization, particularly in services, could bring even larger gains to American consumers, firms, and workers. Advancing free and fair trade in multilateral, regional, and bilateral negotiations will help to ensure that America continues to derive benefits from international trade. This includes renewal of the Trade Promotion Authority and a successful outcome of current global trade talks, the World Trade Organization's Doha Development Agenda negotiations.

Both inward and outward FDI have contributed to higher levels of productivity in the United States. Inward FDI contributes to productivity growth, provides a source of financing for the current account deficit, and generates high-paying jobs for American workers. Outward FDI is an important channel of market access for U.S. multinational companies. U.S. multinationals are an important force behind job creation in the United States and have contributed to productivity growth and rising average living standards in the U.S. economy.

In order to continue to derive important economic benefits from global economic engagement, the United States must continue to break down barriers to trade and investment abroad, and keep our markets open to international trade and secure protections for foreign investors.

Immigration

Immigrants play a vital role in the dynamic U.S. economy. Understanding the forces that drive immigration can help us design more effective immigration policies. This chapter discusses the economics of immigration; the incentive effects of immigration policies on migrants, native workers, and employers; and the benefits of comprehensive immigration policy reform.

The United States is a nation of immigrants and a nation of laws, and we value both historical legacies. Although immigrants continue to make positive contributions to our nation and our economy, our current immigration laws have proven difficult to enforce and are not fully serving the needs of the American economy. It is unofficially estimated that between 11 and 12 million foreign-born persons reside in the United States illegally, almost one-third of the total foreign-born population and about four percent of the total U.S. population.

Effective immigration policy can curtail illegal immigration and at the same time promote America's national and economic interests. Comprehensive immigration policy reform, which combines more effective enforcement capabilities and a temporary worker program, is the most promising route to an immigration system that is legally functional, security conscious, economically beneficial, and humane. In this comprehensive approach, the various elements of policy reform reinforce and enhance one another. In contrast, any given partial reform, standing alone and without the reinforcing measures that characterize the comprehensive approach, cannot fully address the problems and engage the opportunities that accompany immigration.

The key points of this chapter are:

- International differences in economic opportunities and standards of living create strong incentives for labor migration. Once established, migration flows from a certain region tend to be self-perpetuating because past migrants facilitate the movement of new migrants, employers become familiar with the migrant group, and U.S. immigration policy favors family reunification. A large supply of potential migrants will exist for decades to come.
- Foreign-born workers make significant contributions to the American economy, but not all Americans gain economically from immigration. Understanding the labor-market effects of immigration requires consideration of the migrants' skill mix and the capital-accumulation response to labor force growth. Foreign-born workers tend to be concentrated at

the low end and the high end of the educational spectrum relative to native-born workers.

- Immigration policy plays a key role in determining the volume and composition of the foreign-born workforce. Comprehensive immigration reform can help ensure an orderly, lawful flow of foreign-born workers whose presence benefits the American economy.

The Economics of Immigration

International migration patterns are strongly influenced by the interaction of economic forces and public policy. In this sense migration is similar to other aspects of international economic integration and exchange, such as trade in goods and services and investment flows. The fundamental motivation for such movement—whether of goods, capital, or workers—is that people perceive more profitable economic opportunities abroad. The ultimate results are that the world's economy functions more efficiently, entrepreneurship is rewarded, and many Americans reap economic gains.

Compared to barriers to the movement of goods, policy restrictions on the international movement of labor are tight. Immigration policy determines the volume and composition of both permanent immigrants and temporary workers legally admitted to the United States. But many more people would like to come to the United States than are legally permitted to do so, and millions manage to reside and work here illegally. There is broad agreement among U.S. citizens that immigration policy needs to be reformed. To this end, the reform of U.S. immigration policy should be based on an understanding of the forces that drive migration, relevant lessons from American immigration history, and the ways in which immigration affects the economy. This chapter highlights some facts and principles that can help guide the design of a better immigration policy.

The Migration Decision and the Volume of International Migration

Economic analyses of migration typically start by imagining an individual who has many choices about where to live and work at various times in his life. If this person perceives that job opportunities and living conditions are approximately the same everywhere, then he will not have an economic motive to choose one place over another. More realistically, because migration costs time and money and often requires leaving behind one's friends and family and adjusting to a new culture and language, our imaginary individual will be strongly inclined to live and work near his original home. On the other

hand, if the same person perceives that incomes and living conditions differ significantly across places for workers with similar skills, then he might find it worthwhile to incur the costs of migration to secure a higher standard of living. In this sense, migration is like an investment decision—a cost is borne today in return for an increased flow of income and well-being in the future. Essentially, the potential migrant must decide whether the expected benefits from migration outweigh the expected costs.

From the perspective of workers in many countries today, the potential income gains from migration are large. One study measured average wages for Mexican-born men who had recently moved to the United States and compared them to the wages of similar men who were still working in Mexico. The real wage ratios (that is, wages adjusted for international differences in prices) ranged from about 6-to-1 to 2-to-1 in favor of the U.S.-based workers, depending on the age and education group. For example, in 2000 those who were 18 to 22 years old with 5 to 8 years of education earned $7.60 per hour in the United States compared to the equivalent of $1.56 per hour in Mexico. Another study compared the earnings of fast-food restaurant workers who performed nearly identical jobs but in different countries. Again, the real wages in the United States were much higher than in several less advanced economies.

Facing such large international wage differences, a worker might hope to move abroad permanently or with the expectation of returning home after accumulating a nest egg. Indeed, migrants often work intensively at relatively high wages (compared to home) and save or send back home a portion of their earnings. In this scenario the opportunity to work abroad temporarily can help finance large purchases or investments (like a house, car, or new business) in home countries where credit markets are underdeveloped and where wealth accumulation is difficult due to low wages. Migration might also allow households to expand and diversify their income sources, thereby serving as a lifeline to a higher and more stable income level for family members who remain based in a less-developed economy. The large volume of international remittances of migrants' earnings testifies to the strength of the links that migrants maintain with their home country. A recent study estimated that U.S.-based workers from Latin America sent home $45 billion in remittances in 2006, about 10 percent of their total earnings. Nearly three quarters of the migrants in the survey remitted some portion of their earnings.

The decision framework described thus far emphasizes a potential migrant's expectations regarding the future stream of income at home compared to that available abroad, after accounting for broadly defined migration costs (including transportation costs, time spent out of work, difficulties adjusting to a new culture and labor market, and perhaps fees paid to "coyotes" or other smugglers who facilitate illegal migration). But these are not the only

determinants of the migration decision. A potential migrant might consider the risk of unemployment, uncertainties associated with illegal status, and other sources of income variability in different locations. The migrant might also consider factors that are not narrowly economic but that certainly would count as "benefits from migration," such as family reunification or safety from religious or political persecution.

Even if the incentives to migrate are strong, however, the economic costs of migration might be impossible for poor workers to meet by saving or borrowing. Moreover, immigration policies often make it difficult for workers to relocate to high-wage countries, especially if they are not highly skilled or closely related to someone in the high-wage country who can sponsor their application for admission. In this sense, immigration policy acts as a filter that selectively allows some workers to migrate but also deters many potential migrants.

This simplified model of an individual's migration decision is a useful starting point for understanding the economic pressures for labor to move internationally. To make sense of the overall volume and composition of immigration, we must expand our scope to consider the sum of many individuals' migration decisions and the role of immigration policy. Within any given country, some inhabitants might perceive promising economic opportunities abroad whereas others do not; some might have sufficient means to finance the move whereas others do not; and some might have family connections or skills that make it easy for them to relocate legally whereas others do not. Against this backdrop, events (such as economic or political crises) that widen international gaps in expected well-being or that lower the costs of international movement will tend to amplify the volume of international migration because a higher proportion of any given population will find it optimal or feasible to relocate. Working in the other direction, events that narrow gaps in expected well-being and policies that make it more difficult for people to relocate will tend to dampen the volume of international mobility.

The immigration pressures felt by virtually all high-income countries today reflect the ongoing tension between declining costs of migration and persistent international differences in material standards of living, on one hand, and policy responses that seek to manage the inflow of foreign-born persons, on the other. In this context, the flow of legal migration is determined by selective immigration policies. In the United States, these policies facilitate permanent immigration for family reunification and, to a lesser degree, for those with high levels of skill. For other workers, legal channels for migration are narrow while the economic incentives, underpinned by labor demand from U.S. employers and consumers, remain strong. Consequently, many seek employment through illegal channels.

Lessons from American Immigration History

The surge of immigration in recent decades is not unprecedented, and we can better understand the economics of immigration by examining the current situation in light of historical experience. In the decades after the Revolutionary War, migration to the United States was hindered by the high costs of international transport, the relative immobility and poverty of agrarian populations in potential emigrating regions, and political disruptions to international economic integration. By the 1840s, however, economic, technological, and political conditions had combined to launch the first era of voluntary mass migration. The first big waves of U.S.-bound migration originated in northwest Europe, but by the end of the nineteenth century migrants from eastern and southern Europe dominated the immigration flow. The foreign-born proportion of the U.S. population increased from 9.7 percent in 1850 (the earliest census to record place of birth) to 14.4 percent in 1870, and it hovered around 14 percent until 1910 when it began to decline steadily. In recent decades it has risen again, and in 2005 the foreign-born proportion of the population reached approximately 12.4 percent.

The mass migration of labor between 1840 and 1914, along with extensive trade in goods and capital mobility, contributed to a high degree of global economic integration that in many ways was a precursor to our more recent and familiar era of globalization. World War I abruptly curtailed the earlier era of globalization, and the political and economic turbulence of subsequent decades further disintegrated the international economy. Since World War II, policymakers have worked toward re-integrating the global flow of goods, services, and capital. However, in comparison with the pre-1914 era, significant policy restrictions on the international movement of labor remain in place.

Four historical lessons are especially relevant for contemporary thinking about American immigration and the policies that manage the inflow of foreign-born workers. First, migration to the United States has always reflected the relatively high level of labor productivity here. In the previous section, we cited the wage gap between the United States and Mexico. Similarly, estimates of real wage gaps in the late nineteenth century suggest that U.S. wages were often 1.5 to 4 times higher than those available in Europe. Thus, immigration is a sign of our economy's ongoing success and the relatively high rewards that it has long offered its workers. While immigration policy reform is surely necessary, we should be glad that after more than 200 years the United States is still a magnet for ambitious foreign workers.

Second, immigration flows are often self-propagating. From the perspective of a potential migrant, the cost of migration drops sharply when one has a number of friends and family abroad who can help locate employment and housing opportunities and who can provide a sense of community. One

consequence of this self-propagating mechanism is that macroeconomic and political shocks can have long-lasting ramifications for American immigration patterns. The Irish famine in the late 1840s is a salient example of how a dire economic situation abroad accelerated a process of mass migration that continued long after famine conditions had passed. Macroeconomic shocks in Mexico in recent decades, though far less severe than the Irish famine, may have had a similar effect. Durable networks of family, friends, and employers have always facilitated migration, especially given current policy preferences for family reunification.

The third historical lesson is that regions of emigration that are in the process of economic modernization and development often send out an increasing number of workers. Migration has always been a costly enterprise that the very poor cannot easily finance. As the process of modern economic development unfolds, a larger number of workers surpass the necessary threshold of wealth and education for long-distance migration; employment declines in the agricultural sector and young workers seek employment in urban areas at home and abroad; and stronger migrant networks and financial systems develop to facilitate long-distance movement. Along these lines, it has been argued that the spread of economic modernization in Mexico has promoted emigration even as it has raised gross domestic product (GDP) per worker. The ongoing process of economic development in many parts of the world may lead to a growing pool of potential international migrants for decades to come.

Eventually, at advanced stages of economic development when domestic wages rise to levels that are comparable to those that are available elsewhere, the rate of emigration from a particular place tends to decline. The long-run experience of parts of Europe that were massive exporters of labor in the late nineteenth and early twentieth centuries exemplifies this pattern of rising and then falling emigration rates. Thus, a secondary point is that the pool of potential migrants may change substantially as some countries enter into the process of economic modernization and as others reach comparatively high levels of economic development.

Fourth, the demographic structure of regions of emigration is relevant to the volume of international migration. Migrants to the United States have generally been drawn from the pool of relatively young workers. In 2005, for example, foreign-born persons who reported being in the United States for only one year (recent migrants) had a median age of 25, whereas the median age of native-born persons was 35. The young have the most to gain from migration, and they also have fewer ties binding them to a specific location in the home country. Relatively large groups of workers came of age in Mexico in the 1980s and 1990s, and emigration surged when the Mexican macroeconomy stumbled. Reinforcing the point made above, the sheer number of

young people in less advanced economies ensures that many foreign workers will be interested in migration opportunities in the future.

In sum, past experience and current economic and demographic realities suggest that the forces that attract migrants to the United States will continue to be strong in the twenty-first century. Managing the inflow of migrants is an important and complex challenge for policymakers. It demands a comprehensive immigration strategy that views the process for what it is and has always been for the United States—a significant contributor to labor force growth and vitality.

Foreign-Born Workers in the U.S. Labor Force

Foreign-born workers (the sum of both legal and illegal migrants) make up 15 percent of the total U.S. labor force, and since 1996 they have accounted for about half of the total growth in the labor force, thereby fueling macroeconomic growth. In 2005, foreign-born men had higher labor force participation rates than natives (81 percent compared to 72 percent), whereas foreign-born women worked somewhat less than their American counterparts (54 percent compared to 60 percent). Among those in the labor force, foreign-born men had lower unemployment rates than natives (4.1 percent compared to 5.3 percent), whereas foreign-born women had slightly higher unemployment rates than native women (5.4 percent compared to 5.0 percent).

At the high end of the skill spectrum, foreign-born workers were more likely than natives to work in computer, mathematics, architecture, engineering, and science occupations (6.5 percent of foreign born compared to 5.0 percent of natives). Lower in the skill spectrum, the foreign born were two to four times as likely as the native born to work in building and grounds cleaning and maintenance; farming, fishing and forestry; and construction and extraction occupations.

Tables 9-1 and 9-2 report more detailed occupational information for the foreign born. Table 9-1 lists the ten occupations that the foreign born are most likely to fill. For comparison, it also reports the proportion of native-born workers in the same set of occupations. Construction laborers, maids and housekeepers, janitors, and cooks are at the top of the foreign-born occupation list. Together these four occupational categories account for 11 percent of all foreign-born workers compared to about 4 percent of native-born workers. Table 9-2 lists the occupations that have the highest proportion of workers who are foreign born. Tailors and dressmakers, graders and sorters of agricultural products, miscellaneous personal appearance workers (such as manicurists), and plasterers and stucco masons are the occupations with the highest proportions of foreign-born workers, all with over 50 percent. The foreign born are also strongly represented among medical scientists (46 percent).

TABLE 9-1.— *Ten Most Common Occupations for Foreign-Born Workers, 2005*

Occupation	Proportion of Foreign Born (%)	Proportion of Native Born (%)
Construction labor	2.8	0.9
Maids and housekeepers	2.8	0.6
Janitors	2.7	1.4
Cooks	2.7	1.1
Cashiers	2.2	2.1
Drivers/sales workers and truck drivers	2.1	2.3
Grounds maintenance	2.1	0.6
Carpenters	2.0	1.0
Retail salesperson	1.8	2.5
Supervisors, retail sales	1.8	2.3

Note: The sample includes all employed individuals over the age of 15. The "Drivers/sales workers and truck drivers" category includes both truck drivers and those delivering goods in smaller vehicles.

Source: American Community Survey.

TABLE 9-2.— *Ten Occupations with the Highest Proportion of Foreign-Born Workers, 2005*

Occupation	Foreign-Born Proportion of All Workers (%)
Tailors, dressmakers, sewers	53
Graders and sorters (agriculture)	53
Miscellaneous personal appearance workers	52
Plasterers and stucco masons	52
Pressers, textile, garment, and related materials	49
Miscellaneous agriculture workers	49
Drywall, ceiling-tile installers and tapers	48
Sewing machine operators	48
Medical scientists	46
Maids and housekeepers	45

Note: The sample includes all employed individuals over the age of 15.

Source: American Community Survey.

In recent decades, a handful of states have absorbed the majority of foreign-born persons. In 2005, California, New York, Texas, and Florida together accounted for 57 percent of all the foreign born in the United States. The same states accounted for only 29 percent of the native-born U.S. population. These states still attract a large share of the foreign born, as one would expect given the importance of family and information networks in facilitating migration, but there is also evidence of significant gains in many other parts of the country. Georgia, for instance, gained more than 200,000 foreign-born persons between 2000 and 2005, raising its total foreign-born population by 38 percent. Several other states had comparable percentage increases, though smaller gains in absolute numbers. The largest percentage changes were in New Hampshire (51 percent) and South Carolina (50 percent). These geographic shifts reflect foreign-born workers' responsiveness to changes in labor demand across regions within the United States.

The Foreign-Born Skill Mix and the Labor Market Impact

The inflow of foreign-born labor has complex effects on the productivity and earnings of American factors of production—capital, land, and labor. To understand how immigration affects the labor market, it helps to consider the determinants of the skill mix among the foreign born and the nature of substitutability among different factors of production.

American immigration policy acts as a filter that strongly favors potential migrants with family connections to U.S. citizens and lawful permanent residents. In 2004, 946,142 persons were granted lawful permanent resident status. Forty-three percent were admitted as immediate relatives of U.S. citizens and an additional 23 percent were admitted under other family-based sponsorship. Only 16 percent were admitted under the employment-based preference category.

To some extent, this policy structure helps explain observed differences in the economic performance of immigrants from different countries. Most permanently admitted Mexican immigrants, for example, were selected on the basis of family connections rather than skills. Therefore, it is not surprising that as a group they do not fare as well economically as groups of migrants who were selected largely on the basis of their skills, such as those from India.

Out of the employment-based permanent admissions category, only 10,000 lawful permanent resident slots are reserved for less-skilled workers. For less-skilled seasonal workers, H-2A visas (for agriculture) and H-2B visas (for other sectors) admit workers for short durations and specific jobs. These visas help alleviate peak seasonal demands, but there is still demand for less-skilled workers to work for longer durations. In an environment in which unauthorized migrants can find employers without great difficulty, the mismatch between labor market forces and immigration policy has resulted in a large number of unauthorized migrant workers.

Standard surveys, such as the Current Population Survey, do not specifically identify the legal status of the foreign born. Therefore, it is difficult to measure and characterize the unauthorized population with precision. With this caveat in mind, Box 9-1 discusses current estimates of the illegal population's size and economic characteristics.

H-1B visas permit temporary employment for skilled professionals who are sponsored by a U.S. employer, typically in occupations in science, computers, or engineering. The worker can remain in H-1B status for up to six years. Current law permits only 65,000 new H-1B issuances per year, with some exceptions for those with advanced degrees from U.S. universities and those going to work for institutions of higher education or government research organizations. For fiscal year 2007, the H-1B application cap was reached in May 2006.

Box 9-1: The Number and Characteristics of Unauthorized Migrants

Due to the clandestine nature of illegal migration, the unauthorized foreign-born population cannot be precisely enumerated. Nonetheless, reasonable estimates have been made using data from the Census Bureau's Current Population Survey (CPS). The CPS data do not explicitly identify unauthorized individuals, but they do record a great deal of relevant information. Using an estimate of the number of *legal* foreign-born residents that is based on official U.S. immigration data, the total number of *illegal* migrants can then be estimated as the difference between the total foreign-born population and the number of foreign-born estimated to be present legally. In 2006, a study estimated that there were between 11 and 12 million unauthorized migrants residing in the United States, accounting for approximately 30 percent of the total foreign-born population. A related study estimated that between one third and one half of the unauthorized migrants entered the country legally but then overstayed their visas.

To provide more detailed characterizations, the study used statistical techniques to select a certain number of potentially unauthorized foreign-born residents from the March 2005 CPS. Keep in mind that the following conclusions are unofficial estimates. They are subject to error, but are also the best current characterization of the illegal population.

It appears that the labor force participation and occupational choices of unauthorized migrants differ substantially from that of the general U.S. population. Unauthorized adult males (ages 18 to 64) were more likely to participate in the labor force than their native counterparts (94 percent participation rate compared to 83 percent for natives). Unauthorized adult females were less likely than natives to participate in the labor force (54 percent participation rate compared to 72 percent for natives). In this case, the difference partly reflects the migrant women's higher likelihood of having young children in the household.

In general, unauthorized migrants were concentrated in jobs that require comparatively little formal education. Thus, they are under-represented relative to natives in "white collar" jobs in management, business, and professional occupations, and in sales and administrative support occupations. Relative to native-born workers, unauthorized migrants were highly concentrated in other service jobs (31 percent compared to 16); construction and extraction (19 percent compared to 6); production, installation, and repair (15 percent compared to 10); and farming (4 percent compared to 0.5). Although unauthorized migrants represented just 4.9 percent of the total U.S. labor force in 2005, they represented large proportions of the workforce in several specific occupations: 24 percent in

farming occupations, 17 percent in cleaning occupations, 14 percent in construction, and 12 percent in food preparation.

Approximately 40 percent of the unauthorized migrants had been in the country for five years or less. The vast majority of unauthorized migrants had come from Mexico (56 percent, or 6.2 million) and elsewhere in Latin America (22 percent, or 2.5 million).

Nearly half (5.4 million) of the unauthorized migrants were adult males, with a little less than half (2.4 million) of the adult males residing without a spouse or children. Adult females accounted for 35 percent (3.9 million) of the unauthorized migrants, and less than one-fifth of the women were residing without a spouse or children. Approximately 1.8 million children accounted for the remainder of the unauthorized population. In addition, approximately 3.1 million U.S.-born citizen children were living in households where the head or the head's spouse was an unauthorized migrant.

The interaction of migrant supply, labor demand, and policy structure results in a foreign-born skill mix that is described in Chart 9-1 (for all foreign born, age 25 and above). Educational attainment is only one component of productive capability and it does not fully capture ambition, reliability, or knowledge of a specific trade or language. Nonetheless, many jobs have strict educational requirements, and economists frequently study the labor market in terms of educational categories. The height of each bar in Chart 9-1 represents the number of foreign born from each region (age 25 and above). Clearly, Latin America supplies more migrants than any other region, and many from Latin America have less than a high school degree.

Foreign-born workers are found disproportionately at the extremes of the educational spectrum. The educational mix of foreign-born workers relative to native-born workers is shown in Chart 9-2. It differs from Chart 9-1 in that it pertains to all employed workers over age 15, it groups all foreign-born workers together, and it has more detailed information about the top end of the educational scale. The first bar indicates that 15 percent of all workers in the United States in 2005 were foreign born. The foreign born were heavily over-represented in the group of workers with less than a high school degree; they were slightly under-represented among workers with only a high school degree, those with some college, and those with only BA degrees; and they were over-represented among workers with advanced degrees, especially among those with Ph.D. degrees who worked in scientific and technological fields. All together, and remarkably, over 40 percent of Ph.D. workers in computer, mathematical, architectural, engineering, and science occupations were born outside the United States.

Chart 9-1 Educational Attainment Among Foreign-Born U.S. Residents, 2005
Latin America is the most common source of foreign-born U.S. residents.

Number of foreign-born from region
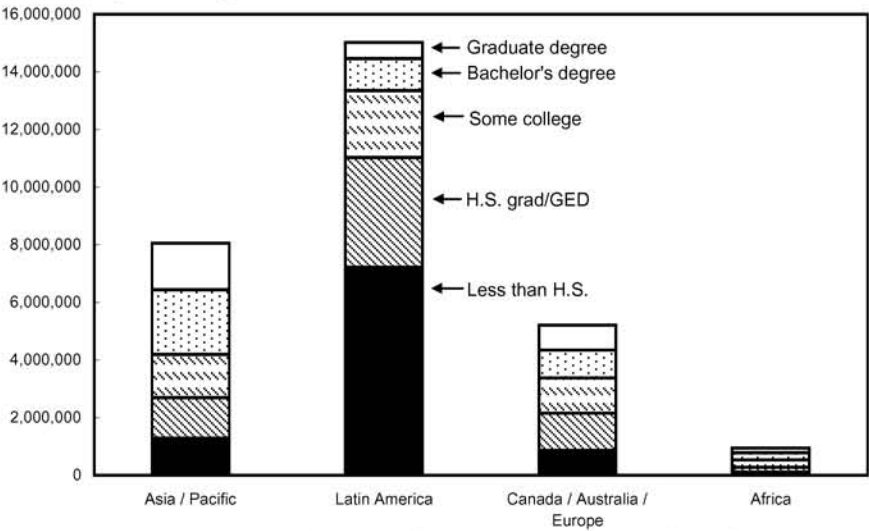
Note: The sample includes all foreign-born U.S. residents ages 25 and older. "H.S. grad/GED" includes high school graduates and those who have passed the general educational development test (GED).
Source: American Community Survey.

Chart 9-2 Foreign-Born Proportion of U.S. Workers by Education Level, 2005
Foreign-born workers are concentrated at the top and bottom of the education distribution relative to native-born workers.

Percent of employed workers
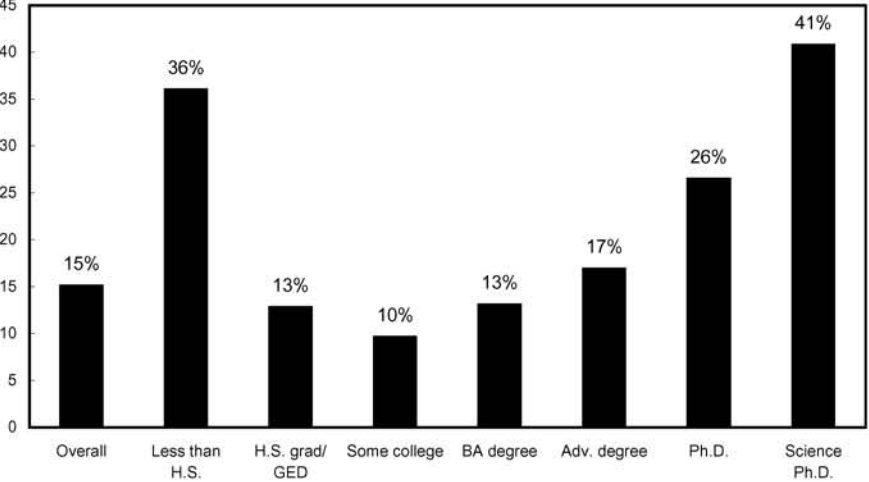
Note: The sample includes all employed persons over the age of 15. "Adv. Degree", "Ph.D.", and "Science Ph.D." are not mutually exclusive categories. The "Science Ph.D." group includes workers in computer, mathematical, architectural, engineering, and life, physical and social science occupations.

Source: American Community Survey.

Highly skilled migrants make many economic contributions to the United States, and a strong case can be made that policy should accommodate more of them. Skilled migrants, whether permanent or temporary, enrich our scientific and academic communities, boost the technical capabilities of U.S. firms (and the native-born workers employed there), augment the supply of healthcare providers, and pay far more in taxes than they absorb in government services. Many of these workers were educated at American universities, and nearly all adjust easily to life in the United States in terms of language skills and employment. They make major innovative contributions in science, medicine, and engineering, and help keep the United States at the forefront of technological capability. For example, between 1901 and 2005 approximately one third of U.S. Nobel Prize winners in medicine and physiology were born abroad.

Because the foreign born, as a group, do not have the same mix of skills as U.S. natives, they alter the relative supply of different types of labor in the economy. The extent to which this alteration of labor supply influences natives' wages depends in large part on whether the foreign born are complements or substitutes for natives in the labor market. When two inputs closely resemble one another, they are likely to be substitutes, and an increase in the supply of one will lower the earnings of the other. In some cases, however, inputs are likely to be complements, and an increase in the supply of one will raise the productivity and, therefore, the earnings of the other. For example, construction laborers may be complements to skilled craftsmen because additional laborers may raise craftsmen's productivity. Conversely, new construction laborers may be close substitutes for other construction laborers and for similar less-skilled workers, and so additional construction laborers would tend to make the services of less-skilled laborers less valuable on the labor market.

The impact of immigration on the labor market also depends on how other factors of production, such as capital, respond to the change in labor supply associated with immigration. In particular, in the short run an increase in the supply of labor puts downward pressure on wages, allows more hiring, and raises the productivity of capital. This increase in capital productivity, in turn, induces firms to invest in more physical capital which ultimately makes labor more productive. Thus, over time the capital accumulation response to immigration tends to offset the downward pressure on wages caused by an increase in the labor supply. The key point is that in trying to understand the effect of immigration on labor markets it does not make sense to suppose that all the other factors that influence labor markets remain the same over a long period of time; rather, these other factors adjust to immigration in important ways.

Economists have produced many data-intensive analyses of the response of native-born workers' wages to immigration, and the debate is still ongoing. To

some extent the estimates depend on the methodological approach that is used to isolate the effects of immigration. One recent study concluded that immigration between 1990 and 2004 slightly raised the wages of most native-born workers but slightly lowered the wages of those without high school degrees (who represent about 10 percent of the native-born labor force). If this finding is correct, then excluding foreign-born workers might give a small boost to the earnings of American high-school dropouts. But such a policy would be costly and counterproductive from the perspective of American consumers, businesses, and most native-born workers. Moreover, such a policy would not be a well-targeted or effective way to assist low-income Americans. The economic challenges facing low-income Americans are a serious concern, but sharp restrictions on immigration are not the remedy. A better policy is to ensure that all Americans have opportunities to acquire skills that will improve their labor market outcomes.

Comprehensive Immigration Policy Reform

Border security is a fundamental responsibility of a sovereign nation and an urgent requirement for our national security. Since 2001, funding for border security has more than doubled, from $4.6 billion in fiscal year 2001 to $10.4 billion in fiscal year 2007. We will have increased the number of Border Patrol agents by 63 percent, from 9,000 at the beginning of this Administration to nearly 15,000 at the end of fiscal year 2007, and we have deployed about 6,000 National Guard troops to assist our border security efforts at the southern border. We have also added 6,700 new detention beds, for a total of 27,500, and have been able to effectively end the practice of "catch and release" of illegal aliens apprehended at the border. The heightened efforts to control entry into the United States are one part of a larger strategy to improve the immigration system while bolstering national security.

The President believes that the best way to fix immigration policy is to adopt a comprehensive program that combines stronger border security, more effective worksite enforcement of employment eligibility laws, and expanded legal channels for the employment of foreign-born workers, including those who are not highly skilled. The comprehensive program would reduce the number of illegal workers and preserve the economic benefits associated with a flexible supply of hardworking foreign-born workers. The key features of comprehensive immigration policy reform would work together and reinforce one another to strengthen the incentives for both workers and employers to comply with immigration and employment laws.

A Legal Bridge between Employer Demand and Migrant Supply

By improving the technology that firms use to verify new workers' employment eligibility and expanding the channels for legal, temporary migration by less-skilled workers, comprehensive immigration policy reform can dramatically reduce incentives for illegal work. Effectively narrowing employment opportunities for illegal workers must be the keystone of immigration policy reform. Unfortunately, at present, it is often difficult for employers to verify the employment eligibility of migrant workers, some of whom have fraudulent documents or engage in identity theft. And as long as some firms employ illegal workers, other firms might do the same to compete on the basis of cost. The current situation with millions of illegal workers and many non-compliant employers is both unacceptable and unnecessary.

Comprehensive immigration reform should aim to establish an environment in which all employers can easily determine the legal status of newly hired workers, in which foreign-born workers can easily prove their identity and legal status, and in which firms can legally hire a foreign worker when no American worker is available to fill a given job. This reform requires an electronic employment eligibility verification system that is accurate, fast, and inexpensive. The Department of Homeland Security continues to refine and expand an internet-based system called the Basic Pilot Program that allows participating employers to verify the employment eligibility of their new hires by checking against Social Security Administration and immigration records. In addition, to curtail the use of fraudulent identity documents, the Department of Homeland Security now issues tamper-resistant, biometrically enhanced (with photograph and finger print) identity documents to most lawfully present foreign-born workers.

Employers also must be held accountable if they hire illegal workers. A rigorous system of verification checks in combination with strong enforcement and enhanced penalties can effectively promote compliance. In this regard, the new policy would remedy the comparatively lax enforcement of immigration law that followed the Immigration Reform and Control Act of 1986.

Electronic verification of new hires' eligibility, tamper-resistant and biometric identification cards for foreign-born workers, and stronger interior enforcement measures should be complemented by the establishment of a temporary worker program, initially proposed by the President on January 7, 2004. A temporary worker program would provide a legal channel for a foreign-born worker to enter the United States for a specific period of time, provided that the worker maintains a consistent work record, does not break the law, and follows the rules of the program. In addition, under certain conditions, some currently undocumented workers would be eligible to work here legally if they pay a substantial penalty for having violated the law. As

long as the costs for program participation are kept low, enforcement is robust, and the number of workers allowed to participate is sufficient, migrants and employers will choose this legal channel for finding matches rather than resorting to illegal means.

A temporary worker program should also endeavor to preserve the flexible role that foreign-born workers play in the American economy. Foreign-born workers are responsive to new economic opportunities and to variation in opportunities over time and space. This responsiveness tends to improve labor market efficiency and overall economic productivity. Administrative requirements that are burdensome for firms, migrants, or government agencies will raise the program's economic costs and, depending on the program's structure, result in either non-compliance with the new policy or a significant drain on government resources.

Finally, comprehensive immigration reform must ensure that highly-skilled immigrants are welcome to make contributions to the U.S. economy. For example, many of the world's best students come to American universities for advanced training in science and technology, and a large share of these students would like to stay and work in the United States after finishing their education. As discussed earlier in the chapter, their work helps keep the United States at the frontier of research and development, and their post-schooling employment depends upon their ability to acquire a temporary work visa or permanent resident status.

The Pitfalls of Partial Policy Reforms

Less-skilled workers are infrequently admitted to the United States unless they have a close relative who is already an American citizen or lawful permanent resident, or they are coming for a short-term, seasonal job. At the same time, America has a strong demand for the products and services that less-skilled workers provide and a declining number of less-skilled domestic-born workers to provide them. This combination acts as a powerful magnet for less-skilled foreign workers. While there is no excuse for breaking immigration and employment laws, the underlying economic forces that draw immigrants to the United States are powerful and deeply rooted. Comprehensive immigration reform can put the United States on a firm legal and economic footing to manage twenty-first century immigration, whereas partial reforms are likely to entail significant costs without yielding satisfactory results.

A policy that relies on more extensive border fencing or more intensive border patrols will make it more difficult for migrants to cross the border illegally. This is an important step in improving control over our borders. By itself, however, this approach will not undercut the existing demand from U.S. employers and consumers for the labor services of foreign-born workers. Therefore, it seems likely that in response to this partial reform the flow of

migrants would change its path rather than dry up completely. Building fences, for example, does not address the problem of lawful entrants over-staying their visas and working without permission, and according to a recent study, between one third and one half of all unauthorized migrants in the U.S. entered the country legally. Thus, even with substantial increases in border patrol resources and increases in the cost of "coyote" services (guides who lead illegal migrants across the southern border), the best efforts of our Border Patrol have not fully stemmed the tide of illegal immigration. Pursuing inten-sive fencing and patrolling approaches to extremes would be inordinately costly in terms of material and manpower, and still it would not achieve the goal of greatly reducing the employment of illegal workers.

A partial policy reform that targets current employers of illegal migrants might lower the demand for illegal workers, make it more difficult for illegal migrants to find work, and therefore lessen the illegal inflow. But if the supply of authorized foreign-born workers is not simultaneously augmented through a temporary worker program, this approach would hurt many American companies and consumers and, as discussed above, would hurt complemen-tary American workers. It would also slow the growth of the labor force and the overall economy.

Alternatively, a partial policy reform that focuses primarily on detecting, apprehending, and removing illegal workers who are already present in the United States might reduce migrants' desire to live and work here, but would be very costly to carry out. Moreover, fundamental economic forces would still drive many foreign workers to try their luck in America, illegally if neces-sary. The likely outcome of such partial reform is that there would still be many illegal workers and, more than ever, they would be unwilling to communicate with local law enforcement officials, prone to work in the underground economy, and subject to exploitation by criminals, smugglers, and unscrupulous employers.

Unlike partial reforms, the President's comprehensive approach can succeed because it combines a number of elements that reinforce one another. This comprehensive approach gives employers access to a source of legal foreign-born workers when they cannot find Americans to fill jobs, gives them better tools to verify the employment eligibility of persons they hire, and strongly punishes non-compliance with enhanced civil and criminal penalties. This approach also provides potential temporary migrants a more expansive legal route to employment in the United States that does not depend so heavily on having high levels of education and skills or on having relatives in the United States to sponsor them. It also makes illegal border crossing more difficult to accomplish and makes unauthorized employment more difficult to find. By simultaneously narrowing illegal channels for migration and employment and widening legal channels, the comprehensive approach to immigration policy reform can significantly improve upon the current system.

Conclusion

Immigrants make important contributions to the American economy. They help the economy grow by adding to the labor force; they fill in jobs at the lower end of the skill distribution where relatively few native-born Americans are available to work; they also fill in jobs at the highest end of the skill distribution and help keep the United States at the forefront of technological and medical innovation; they respond quickly and flexibly to shifts in labor demand; and they work hard to make better lives for themselves and their children. Immigration is both a reflection of and a contributor to our economy's prosperity.

The foreign-born proportion of the population has steadily increased in recent decades, and now stands at about 12 percent of the total U.S. population. Over the same period, the U.S. economy has performed well in comparison with other advanced economies. Still the large number of unauthorized workers has made it clear that our current immigration policy is inadequate. At the same time, the economic forces that drive international migration are as strong as ever and will remain so for the foreseeable future.

Comprehensive immigration policy reform can improve border security, significantly reduce the number of illegal workers, and yield economic benefits for employers, workers, and consumers in the United States. Achieving these policy goals requires better interior enforcement which, in turn, requires better tools for employers to verify worker eligibility. It also requires the creation of better legal channels for the migration of hard-working foreign-born workers who are eager to fill jobs that contribute to the American economy. Such workers tend to enhance the productivity of American factors of production, but they currently have few avenues, aside from family reunification, to gain legal entry and employment for a sustained period of time. By mutually reinforcing one another, the various components of comprehensive immigration policy reform can support a legally and economically viable immigration system.