CBO MEMORANDUM

PROJECTING CAPITAL
GAINS REALIZATIONS

November 1995

CONGRESSIONAL BUDGET OFFICE
SECOND AND D STREETS, S.W.
WASHINGTON, D.C. 20515

NOTE

Unless otherwise indicated, all years in this memorandum refer to calendar years.

This Congressional Budget Office (CBO) memorandum is part of CBO's ongoing documentation of its procedures for projecting federal revenues.

Larry Ozanne of CBO's Tax Analysis Division prepared the memorandum under the supervision of Rosemary Marcuss and Frank Sammartino. Kim Kowalewski and William Randolph of CBO and Patrick Driessen of the Joint Committee on Taxation provided valuable comments. Peter Ricoy prepared the figures.

Paul L. Houts edited the memorandum, with the assistance of Chris Spoor, and Simone Thomas prepared it for publication.

CONTENTS

TABLES

FIGURES

At the end of each calendar year, the Congressional Budget Office (CBO) projects the net value of capital gains that taxpayers will report on their tax returns for the year ending and for the coming six calendar years. Those projections represent one component in CBO's projection of tax revenues for the current and coming five fiscal years. The Joint Committee on Taxation also uses CBO's capital gains projections as the baseline against which it calculates the revenue effects of changes in capital gains taxation.

CBO's projections are based on the historical pattern of capital gains realizations during the past 40 years. In that period, the amount of capital gains that taxpayers realize has generally increased along with the size of the economy, but has shown substantial fluctuations from year to year. Some of that fluctuation follows fluctuations in the values of assets such as corporate stocks and real estate. Changes in the federal individual income tax rates on realized gains account for other shifts in the level of realizations. Equations that explain the growth of realized gains in terms of the above factors can account for 80 percent to 90 percent of the growth of realized gains between 1955 and 1993.

CBO uses a three-step procedure to project the amount of capital gains that taxpayers will report. In the first step, the equations that explain the historical pattern of realizations are used to forecast the growth of realizations for the year ending. Actual data on those realizations do not become available for over a year because of the time allowed for people to file tax returns and for the Internal Revenue Service to process the returns and tabulate the results. Because indicators of the data used by the equations to explain realizations are available near year's end, the equations can be used to forecast realizations.

In the second step, CBO projects realizations over the coming six calendar years by assuming that realizations will gradually revert to their historical average relative to the size of the economy, adjusted for differences in the current tax rate on capital gains from its historical average. The equations are not used to forecast the six coming years because the explanatory variables in the equations must be forecast first, and some of those variables, such as stock prices, cannot be forecast accurately. Using an alternative method that requires less information about the future can reduce the forecast error and lessen the size of revisions to the forecast in subsequent baseline projections.

In the third step, CBO adjusts the entire seven-year projection for major effects on realizations that are not incorporated in the first two steps. For example, in the projection made for CBO's January 1995 economic and budget outlook, the projection resulting from the first two steps was adjusted to reflect a hypothesized

shift in realizations from 1994 to 1995 and 1996 as some taxpayers deferred sales of assets until the prospects for a capital gains tax reduction became clearer.

CBO has been forecasting realized gains since the end of 1987. The three-step procedure has been used since the end of 1991, and it incorporates refinements that have been suggested by experience with earlier forecasts.

INTRODUCTION

Once a year, the Congressional Budget Office (CBO) projects the total value of net gains realized on the sale of capital assets as part of its six-year economic and budget baseline projections. The economic and budget baseline provides the Congress with an estimate of federal revenues and spending that it uses in developing the federal budget resolution for the upcoming year. The Joint Committee on Taxation also uses the CBO projection of gains as a basis for estimating how legislated changes in the taxation of gains will affect revenues.

Net gains represent a modest component of the personal income tax base. Between 1980 and 1992, for example, aggregate realized gains averaged 5.6 percent of total adjusted gross income, and taxes paid on realizations accounted for a slightly higher percentage of personal income tax receipts.

Although modest in size, net gains are a volatile component of the tax base and hard to predict. Between 1980 and 1992, the ratio of realized gains to adjusted gross income varied from a low of 3.2 percent in 1991 to a high of 13.1 percent in 1986, just before an increase in the tax rate on gains went into effect. The standard deviation of the ratio over those years is 2.4 percentage points. Few components of the personal income tax base are anywhere near as volatile. Furthermore, the dynamic that leads the millions of taxpayers to realize more or fewer gains in one year than in another is poorly understood.

Because of its volatility and uncertainty, the value of realized gains is difficult to project, which means that the annual projections of gains are subject to relatively large errors. Although some past errors have been large compared with the actual level of gains, they have seldom been sizable enough to cause large errors in projections of overall tax receipts.

Because the gains component of the baseline revenue projection is unusually prone to error, the Congressional Budget Office devotes considerable effort to learning about the subject; it tracks economic indicators and tax-return data that provide insights into realizations of gains, refines its projection methods, and explains the phenomenon of aggregate gains realizations to policymakers. This memorandum is part of that effort. It describes the method CBO now uses to project gains. It also explains as a case in point the projection of gains developed in December 1994 as part of the CBO baseline budget projections released in January 1995.

CBO's Economic and Budget Outlook

In January of each year, CBO presents its outlook for the economy for the coming six calendar years and for federal spending and revenues for the current and next five fiscal years.

The economic outlook is composed of a two-year forecast followed by a four-year projection.[1] The forecast predicts the actual path of the economy, including cyclical movements, based on economic models. Because of the uncertainty of model forecasts beyond two years, a simple projection is used for the final four years. The projection does not predict cyclical swings and instead assumes that the various components of the economy move steadily to their historical relationships by the end of the projection. Both the forecast and the projection show paths for major components of the economy, such as gross domestic product (GDP), the price level, the unemployment rate, interest rates, and many components of GDP.

CBO combines the paths of the macroeconomic variables with other information to project spending under particular programs and revenues from each major tax. For the personal income tax, the macroeconomic paths of wage and salary incomes, interest income, inflation, and other variables are used to project amounts of income and deductions that taxpayers will report on their income tax returns. Revenues are then computed by applying existing tax law to projected incomes and deductions on a sample of tax returns.

The Capital Gains Projection

The amount of capital gains that taxpayers will realize is one of the income components that must be projected. In each December, when the macroeconomic outlook is completed, variables from it are combined with other information to project the amount of capital gains that taxpayers will realize in each of the next six calendar years. Because the amount of capital gains that have been realized in the year just ending are unknown in December, they must be forecast as well. Not only are realizations of the closing year unknown, only preliminary information about the previous year is available. Final data on tax returns cannot be tabulated for more than a year after the taxable year closes because of the time allowed for tax returns to be filed and the time required for the Internal Revenue Service to process the returns and prepare tabulations.

CBO currently uses a three-step procedure to project realized gains. In the first step, it uses a set of equations to forecast the rate of growth of capital gains from the year for which preliminary data are available to the year just ending. CBO has developed the equations using historical data that explain the growth of realizations in terms of its major observable determinants. In the second step, CBO projects realizations to grow steadily in the coming six years to reach their historical average

---

1. In everyday language, forecasting and projecting are nearly interchangeable. In CBO's annual report, and to some extent among economists more generally, the two terms have distinct meanings. Forecasting an economic outcome is done from a model whose data inputs and equations reflect the process that determines the outcome. Projecting an economic outcome employs a simpler rule of thumb or historical trend. This memorandum employs a similar distinction, except that the multiyear combination of forecasts and projections also are referred to as a projection.

relative to GDP, conditional on current tax rates. In a final step, CBO adjusts the projection in response to current conditions not accounted for in the first two steps.

The first two steps parallel those of the economic outlook, but for different periods. The use of equations to forecast realizations for the year ending is similar to the use of economic models to forecast the path of the economy over the first two years of the economic outlook. The projection that realizations will return to their historical level relative to GDP and tax rates during the following six years is similar to the projection that the economy will return to its historical average relationships during the last four years of the outlook.

Which Capital Gains Are Projected?

CBO projects those capital gains that are subject to personal income taxation. In general, capital gains are the amount that a person gets from selling an asset over and above what he or she invested in the asset. What a person gets from selling an asset is the actual sales price less the costs of the sale, such as brokerage fees. What he or she invested in the asset is termed the asset's basis. Basis includes the actual purchase price or construction costs. It also includes the costs of making the purchase and costs of any improvements the owner makes. Deductions for depreciation or depletion that the owner claims on the asset must be subtracted from the basis. Of course, if a person gets less from selling an asset than its basis, he or she has a capital loss instead of a gain.

Gains and losses accrue while an asset is owned as the value of the asset changes. Values change unevenly, depending on prospects for a particular business or property and on trends in the broader economy. That variability makes projecting capital gains and losses difficult. Accrued gains and losses are not subject to taxation until they are realized by the sale of the asset. Projecting realized gains is even more difficult than projecting accrued gains because many additional factors influence people's decisions to sell and buy assets, and the timing of trades can frequently be shifted from one year to another when that is advantageous.

When people sell more than one asset in a year, they add the gains, subtract the losses, and report only the net gain or loss on their tax return. CBO projects the sum of net gains on all returns reporting a net gain. The sum includes net gains reported on Schedule D and capital gains distributions reported directly on Form 1040.

The sum of net losses on all returns reporting a loss is projected separately and is not discussed in this memorandum. The amount of net losses taxpayers can claim in a year is currently limited to $3,000, and the total of net losses is small compared with net gains. In 1992, for example, taxpayers claimed net losses of $8.4 billion compared with net gains of $126.7 billion.[2]

For tax purposes, realized gains are classified as short term or long term. Currently, assets held less than one year are short term, and those held longer are long term. Most realized gains are long term. In 1992, returns with net long-term gains reported $134.7 billion of long-term gains and returns with net short-term gains reported $22.1 billion of short-term gains.[3] CBO forecasts the sum of short- and long-term gains.

Since 1987, taxpayers have been required to include 100 percent of their long-term gains in their adjusted gross income. Between 1942 and 1978, 50 percent of long-term gains were excluded, and between 1979 and 1986, 60 percent were excluded.[4] The historical data presented below include 100 percent of net realized gains in all years.

Not all accrued capital gains are eventually realized for income tax purposes. The gains on assets held at the time a person dies are excluded. That exclusion is accomplished by "stepping up" the basis of the assets that heirs receive to their current market value. Between one-half and two-thirds of all accrued gains have been estimated to escape taxation through that route.[5] Capital gains on assets contributed to charitable institutions are not taxed either. The owner generally can claim a deduction for the full market value of the asset and the charitable institution generally is not taxed if it sells the asset.

The gains on owner-occupied homes are often excluded as well. When people sell one home and buy another, they can roll the entire gain on the sold home over to the purchased home as long as the purchased home is more expensive, which typically is the case. In addition, a one-time exclusion of a dollar amount of gains, currently $125,000, is allowed on a principal residence as long as the seller is over age 55. Finally, if a home is held until the owner dies, no gains are realized because of the step-up in basis that is allowed all assets held until death. The historical data and projections in this report cover only those capital gains subject to income taxation.

---

2. Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns 1992* (1995), p. 38.

3. Ibid. The $134.7 billion and the $22.1 billion cannot be added to arrive at net gains in taxable income. On some returns, the net long-term gains are reduced by net short-term losses, and on other returns, the net short-term gains are reduced by net long-term losses.

4. Between 1934 and 1941, the fraction of gains excluded from taxation increased in several steps with the length of the holding period. Between 1922 and 1933, capital gains held over two years were subject to a maximum rate below the top rate applicable to other income. No preferences existed before 1922. The size of the exclusion is one of several special provisions affecting the tax on capital gains income. The tax rates on capital gains are discussed later.

5. Alan J. Auerbach, "Capital Gains Taxation and Tax Reform," *National Tax Journal*, vol. 42, no. 3 (September 1989), p. 394; and Jane G. Gravelle and Lawrence B. Lindsey, "Capital Gains," *Tax Notes* (January 25, 1988).

EXPLAINING PAST CAPITAL GAINS REALIZATIONS

The total amount of capital gains that taxpayers realize in a year can be expected to follow the total amount of capital gains that are outstanding. The amount outstanding is the difference between the amount that has accrued in that and previous years and the amount that has previously been realized or stepped up through transfer at death. Realized gains can also be expected to vary inversely with the costs of selling assets.

Although direct measures of outstanding capital gains are not available, several indicators of total accruals are. Accrued gains should increase with the prices of assets and the size of the capital stock. Corporate equities are an important source of realized gains, and their past prices are available from the stock exchanges. Real estate is another major component of realized gains, and though national measures of real estate prices have not been available historically, data on housing starts provide a rough indication of movements in real estate prices. Gross domestic product moves in the same direction as the prices of many assets, and it also provides an indicator of the size of the capital stock.

The cost of selling assets with substantial accrued capital gains will depend partly on the rate at which those gains are taxed. Federal tax rates are the major source of taxation, and data on federal tax rates have been collected by the Treasury Department and the Congressional Budget Office.

The relationship between realization of capital gains and the observable determinants listed here can be seen by examining historical patterns. After CBO reviews those patterns, the relationships are specified more precisely with equations estimated from the historical data.

Realized Capital Gains and Indicators of Their Determinants: 1954-1993

Realized gains have grown substantially since 1954, but not at a steady rate. In 1954, \$7 billion was realized, compared with \$153 billion in 1993 (see Table 1). Realizations in 1993 were far from the peak, however. The peak occurred in 1986 when \$324 billion was realized. Realizations in 1986 were boosted far above realizations in other years because people sped up the sale of assets into the end of 1986 to escape the higher taxes scheduled to begin in 1987. In 1985, the previous high-water mark, realizations had been \$171 billion, and in 1987 they fell back to \$144 billion.

TABLE 1. REALIZED CAPITAL GAINS AND GROSS DOMESTIC PRODUCT

Year	Realized Gains (In billions of dollars)	GDP (In billions of dollars)	Ratio of Gains to GDP (In percent)
1954	7.16	371.5	1.93
1955	9.88	404.9	2.44
1956	9.68	427.1	2.27
1957	8.11	449.9	1.80
1958	9.44	455.6	2.07
1959	13.14	494.2	2.66
1960	11.75	513.4	2.29
1961	16.00	531.8	3.01
1962	13.45	571.6	2.35
1963	14.58	603.1	2.42
1964	17.43	648.1	2.69
1965	21.48	702.7	3.06
1966	21.35	769.8	2.77
1967	27.54	814.3	3.38
1968	35.61	889.3	4.00
1969	31.44	959.5	3.28
1970	20.85	1,010.7	2.06
1971	28.34	1,097.2	2.58
1972	35.87	1,207.0	2.97
1973	35.76	1,349.6	2.65
1974	30.22	1,458.6	2.07
1975	30.90	1,585.9	1.95
1976	39.49	1,768.4	2.23
1977	45.34	1,974.1	2.30
1978	50.53	2,232.7	2.26
1979	73.44	2,488.7	2.95
1980	74.58	2,708.1	2.75
1981	80.94	3,030.6	2.67
1982	90.15	3,149.6	2.86
1983	122.01	3,405.1	3.58
1984	138.66	3,777.2	3.67
1985	170.57	4,038.7	4.22
1986	324.45	4,268.6	7.60
1987	144.17	4,539.9	3.18
1988	161.87	4,900.4	3.30
1989	153.51	5,250.8	2.92
1990	123.78	5,546.1	2.23
1991	111.59	5,724.8	1.95
1992	126.69	6,020.2	2.10
1993	153.26	6,343.3	2.41
1994	n.a.	6,734.8	n.a.
Average, 1954-1993	*	*	2.79
Average Excluding 1986	*	*	2.67

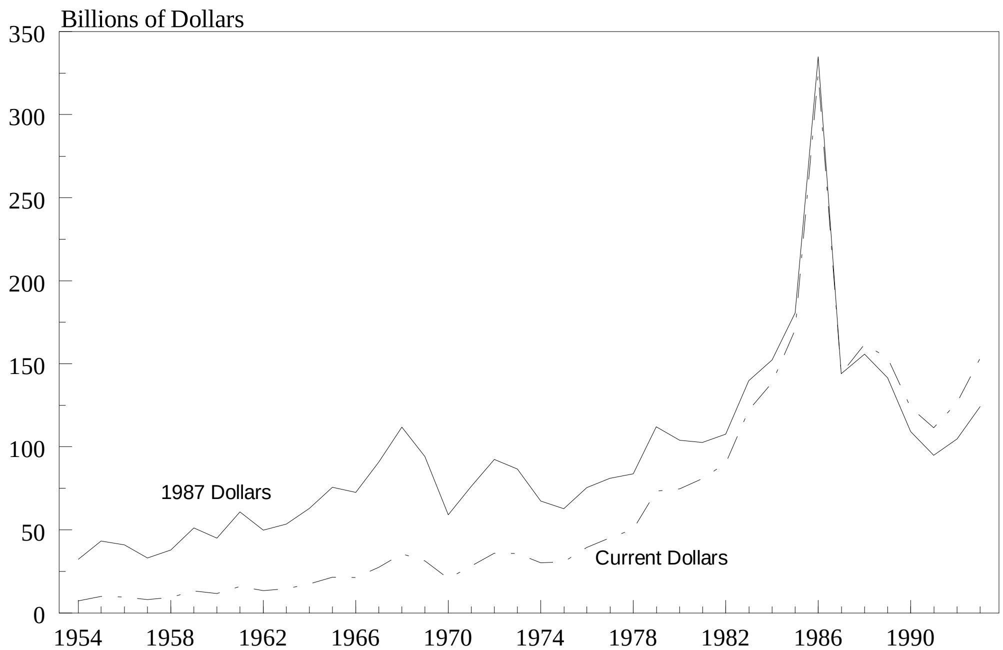
SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years1954 through 1992; unpublished Internal Revenue Service data for 1993; and Department of Commerce, Bureau of Economic Analysis.

NOTE: n.a. = not available; * = not applicable.

Although the peak in realizations in 1986 is exceptional, growth of realizations has been uneven throughout the last 40 years. Inflation overstates realizations in recent years relative to those in earlier years. When the amount of realized gains in all years is measured in constant 1987 prices, the growth of realized gains can be seen to be as uneven in the early years as it has been recently (see Figure 1).

Measured in constant dollars, realized gains grew at accelerating rates through 1968, and then declined sharply between 1968 and 1970. They grew much less rapidly during the 1970s than in the 1960s, and did not regain their 1968 level until 1980. Then, between 1982 and 1985, they grew very rapidly. After the 1986 peak, realizations grew between 1987 and 1988, but declined steeply from 1988 through 1991. Realizations turned up in 1992 and rose again in 1993.

FIGURE 1. CAPITAL GAINS IN BILLIONS OF CURRENT AND 1987 DOLLARS



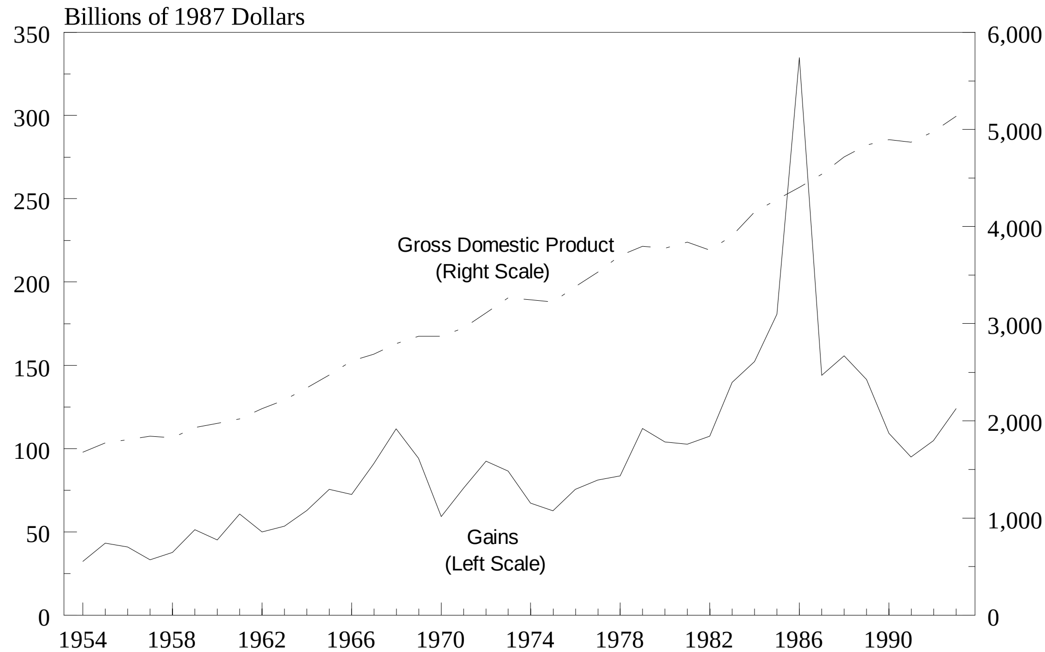
SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years 1954 through 1992; unpublished Internal Revenue Service data for 1993; and Department of Commerce, Bureau of Economic Analysis.

NOTE: The implicit price deflator for gross domestic product is used to convert realized capital gains to constant 1987 dollars.

Gross Domestic Product. As mentioned above, growth of realized gains should follow growth of total output because growth of output reflects growth of both the capital stock and asset prices generally. Fluctuations in realized gains may also follow fluctuations in output over the business cycle if the buying and selling of assets rises and falls as businesses alternately expand and cut back. Data on total output, as measured by GDP, are available in the national income and product accounts.

When GDP and realized gains are both measured in constant 1987 dollars, the two variables follow a similar trend from 1954 through 1985 (see Figure 2). The trend is disrupted between 1986 and 1991, but appears to reemerge in 1992 and 1993.

FIGURE 2. CAPITAL GAINS AND GROSS DOMESTIC PRODUCT



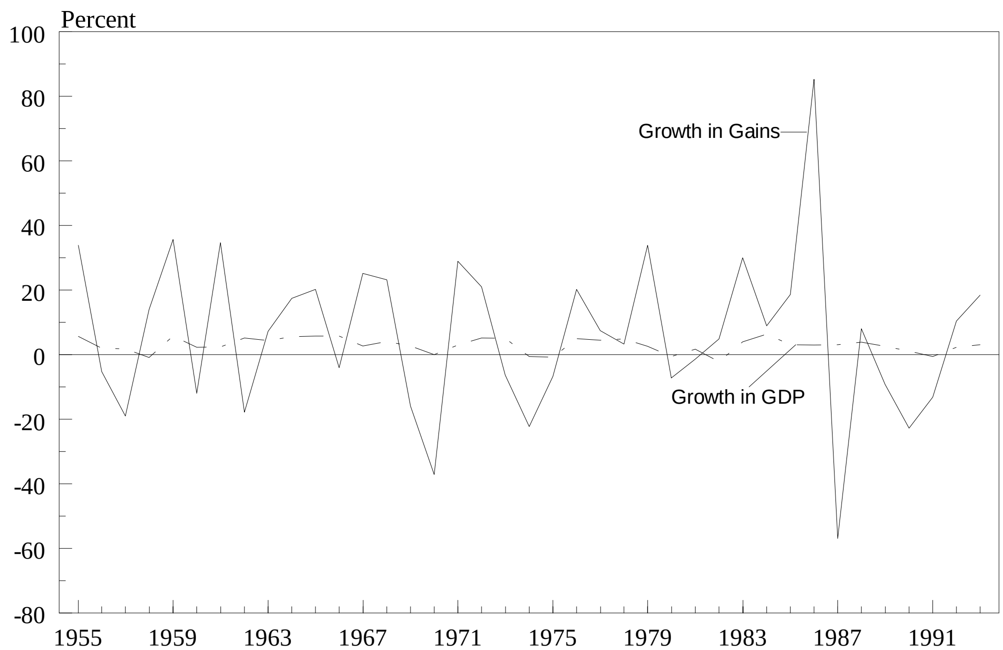
SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years 1954 through 1992; unpublished Internal Revenue Service data for 1993; and Department of Commerce, Bureau of Economic Analysis.

NOTE: The implicit price deflator for gross domestic product is used to convert realized capital gains and gross domestic product to constant 1987 dollars.

In addition to the similar trend, the annual rates of growth of the two variables tend to fluctuate together. The growth rate of realized gains fluctuates much more widely than that of GDP, measured in real 1987 dollars, but both tend to grow faster or slower in the same years (see Figure 3).

Because realizations grow more irregularly than GDP, the ratio of realizations to GDP fluctuates from year to year. The ratio dipped below 2 percent in 1954, 1975, and 1991, and topped 4 percent in 1968, 1985, and, of course, 1986. Excluding 1986, the ratio has averaged 2.67 percent. A straight line fitted to the ratio has a slope that is positive but not statistically different from zero. The absence of a significant positive or negative slope indicates that realizations have not been growing markedly faster or slower than GDP.

FIGURE 3. ANNUAL GROWTH RATES OF CAPITAL GAINS AND GROSS DOMESTIC PRODUCT



SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years 1954 through 1992; unpublished Internal Revenue Service data for 1993; and Department of Commerce, Bureau of Economic Analysis.

NOTE: Annual growth rates are measured as the change in the logarithms of annual capital gains and gross domestic product. Both are measured in constant 1987 dollars. The implicit price deflator for GDP is used to convert realized capital gains and gross domestic product to constant 1987 dollars.

Stock Values. Corporate stocks are a major source of realized gains. For selected years in the 1970s and 1980s, the Internal Revenue Service has prepared data on the types of asset sales on which gains are realized. In those years, the fraction of gains accounted for by the sale of corporate stocks ranges from just below one-fifth to almost one-half.[6]

The prices of corporate equities are available from the stock exchanges, but those prices alone do not capture the effects of price increases on the accruing gains of taxpayers. The share of stocks owned by individuals has been on a long-term downward trend because pension funds have been increasing their share of equities. Annual shares held by individuals also fluctuate from year to year for other reasons. A conceptually better indicator of capital gains accruing to individuals is the value of corporate equities held by households, which is collected by the staff of the Board of Governors of the Federal Reserve System.[7]

The rise and fall in the value of stocks held by households is matched by fluctuations in realized gains in many years between 1954 and 1993 (see Figure 4). That pattern suggests that increases in accrued gains on stocks show up quickly in realized gains. In contrast, the major decline in the value of stocks held by households between 1968 and 1974 is only faintly reflected by the path of realized gains, and the sharp increase in the value of holdings between 1988 and 1993 coincides with a major decline in realizations. Either taxpayers changed the rate at which they realized accruing gains on stocks or they changed the amount they realized on other assets during those periods.

Housing Starts. Real estate accounts for another large share of realized gains. Internal Revenue Service data for 1985 indicate that gains on real estate accounted for at least one-third of gains in that year.

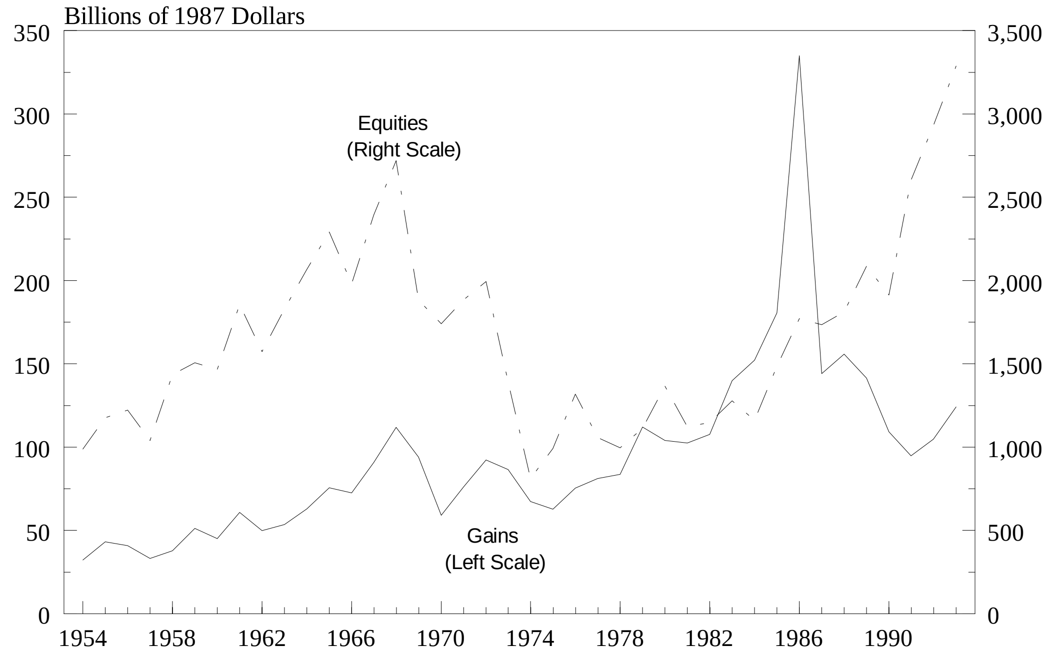
A national series on real estate prices is not available for the historical period. CBO uses Bureau of the Census data on housing starts as a replacement. Housing construction in local markets becomes profitable when real estate prices in those markets rise sufficiently to cover the costs of new construction. As a result, national data on housing starts should indicate in how many markets the price of housing has risen above construction costs. The greater the number of starts, the more wide-spread the rising prices of residential buildings, and therefore the greater should be the accrued gains of people holding residential real estate.

---

6. Shares of realized gains by asset type are reported for the years 1973, 1977, and 1981 in Michael J. Coleman, "Statistics of Income Studies of Individual Income and Taxes," *SOI Bulletin*, vol. 7, no. 3 (Winter 1987-1988), p. 30. CBO has adjusted the percentages to remove gains on personal residences, which are largely untaxed. Realized gains by asset type for 1985 are reported by Dan Holik, Susan Hostetter, and John Labate, "1985 Sales of Capital Assets" (paper presented at the 150th annual meeting of the American Statistical Association, Washington, D.C., August 6-10, 1989).

7. Board of Governors of the Federal Reserve System, *Balance Sheets for the U.S. Economy* (October 1994). The values CBO uses are the sum of direct holdings and holdings through mutual funds by the household sector. A shortcoming of the data for explaining taxable realizations is that the household sector includes holdings of nonprofit organizations such as private universities.

FIGURE 4.   CAPITAL GAINS AND CORPORATE EQUITIES



SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years 1954 through 1992; unpublished Internal Revenue Service data for 1993; and the Federal Reserve Board.

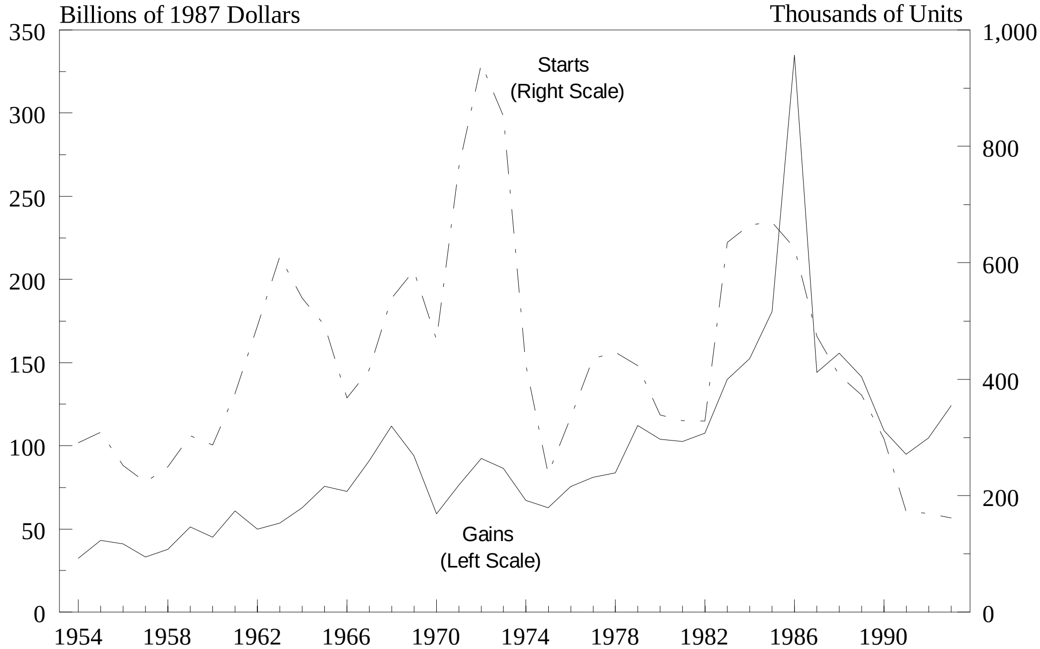
NOTE: Corporate equities are the values held by the household sector in the flow of funds accounts. The implicit price deflator for gross domestic product is used to convert realized capital gains and the value of household holdings of corporate equities to constant 1987 dollars.

Because individual income taxes largely exclude capital gains on owner-occupied homes, the most relevant measure of housing starts excludes starts of single-family housing units. Most buildings with two or more units are owned by investors who are subject to tax on gains when the structures are sold.[8]

The variable for housing starts should reflect price movements for commercial as well as residential real estate because markets for the two types of real estate are closely linked. Residential and commercial real estate are produced with similar inputs and construction technology, and demands for each type of real estate are strongly influenced by interest rates and local population growth. Those close links should cause prices for the two types of real estate to move in similar directions. Data

8. CBO has removed starts of units that were subsidized under federal programs because starts of those units would be less sensitive to market prices than starts of unsubsidized units. The subsidized starts occurred under the Section 202 and Section 221(d)(3) programs with below-market interest rates, under the Section 236 program, and under the Section 8 new construction program. Starts subsidized through the low-income housing tax credit have not been removed and may be important in the 1990s.

FIGURE 5. CAPITAL GAINS AND MULTIFAMILY HOUSING STARTS



SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years 1954 through 1992; unpublished Internal Revenue Service data for 1993; and Department of Commerce, Bureau of Economic Analysis and Bureau of the Census.

NOTE: Housing starts are of units in structures with two or more units per project. Starts of units subsidized by the Department of Housing and Urban Development under programs operating between 1961 and 1982 have been removed. The programs are the below-market interest rate loans under the Section 202 and Section 221(d)(3) programs, the Section 236 program, and the Section 8 program. The implicit price deflator for gross domestic product is used to convert realized capital gains to constant 1987 dollars.

provided by the Census Bureau for 1964 through 1990 shows that construction activity in the two markets has tended to rise and fall together, confirming the close links between the two markets and further suggesting that prices for the two types of real estate should tend to move together.[9]

The data on housing starts suggest that prices of real estate are volatile. The fluctuations in housing starts themselves are even larger than the fluctuations in realized capital gains (see Figure 5). The changes from peak to trough are steeper, and no upward trend is apparent between 1954 and 1993. The spike of housing starts in the first half of the 1970s is larger in percentage terms than the spike in realizations in 1986, and the decline in housing starts from their 1980s peak to their lows in the early 1990s is more pronounced than the decline in realizations between 1989 and

9. Bureau of the Census, *Construction Reports--Value of New Construction Put in Place* (May 1986 and March 1991).

1991. In fact, housing starts in the 1990s have been at lower rates than in any previous year of the 40-year period.

Although more volatile, the peaks and troughs of housing starts are echoed in the fluctuations of realized gains. The declines in starts since 1986, in particular, may explain the decline in realizations between 1989 and 1991 when rising stock prices and GDP suggest that realizations should be increasing. The decline in housing starts since 1986 reflects the high vacancy rates in residential and commercial buildings and the bankruptcies of many builders and lenders, all of which indicate that real estate values were falling and probably reducing realizations of gains.

Tax Rates. Federal income taxes are one of the main costs of selling assets with accrued gains. Therefore, federal income tax rates on gains are likely to influence whether and when people decide to realize gains.

The federal tax rates on gains realized by individuals in a particular year are based on the rate structure applying to all income because gains are generally treated as a component of adjusted gross income. Tax rates on gains have rarely been the same as on other income, however, since special provisions have limited the applicability of the general rate structure to realized gains. Those provisions have included the 50 percent and 60 percent exclusions mentioned earlier, the alternative tax for capital gains between 1954 and 1978, and the separate maximum tax rate on capital gains that has been in effect since 1987. The effects of those special provisions for capital gains have themselves been limited at various times by additional provisions, such as the alternative minimum tax, which have further altered the tax rates on capital gains paid by some taxpayers.[10]

The wide range of statutory tax brackets in many years and the existence of special provisions affecting capital gains might suggest that the best single representation of capital gains tax rates would be an average of rates paid by all individuals. The drawback of using an average rate, however, is that changes in statutory rates would not be precisely reflected in changes in the average rate unless taxpayers in all tax brackets responded similarly to the statutory change.[11]

---

10. Histories of capital gains taxation before 1986 are provided in Congressional Budget Office, *How Capital Gains Tax Rates Affect Revenues: The Historical Evidence* (March 1988); Department of the Treasury, *Report to the Congress on the Capital Gains Tax Reductions of 1978* (September 1985); and Joseph A. Pechman, *Federal Tax Policy*, 5th ed. (Washington, D.C.: Brookings Institution, 1987).

11. To see that the average tax rate is influenced by differential responses of taxpayers, suppose that taxpayers in the highest tax brackets are more responsive to tax rate changes than other taxpayers. Then, if the tax rates on capital gains were reduced across the board (say, by 50 percent), people in the highest brackets would increase their realizations by more than others. That outcome would cause their share of gains to increase, which would tend to keep the average tax rate from falling by 50 percent. The net effect would be to understate the amount by which the statutory rates had declined.

An alternative measure of tax rates that is independent of people's response is the top statutory tax rate on long-term gains.[12] It reflects legislative changes only, and most legislative changes have changed other rates in the same direction. In addition, a large fraction of gains throughout the historical period have been received by people who are taxed at the top rate. Between 1954 and 1986, the fraction of long-term gains received by taxpayers who faced the top rate varied between one-third and one-half. That fraction has increased to around three-fourths since 1987.[13]

Although the top statutory tax rate on gains is a good indicator of the overall tax rate, it is not complete. The wide variation in the fraction of gains subject to that rate indicates that legislative changes have not affected all gains equally. In particular, the Tax Reform Act of 1986 raised the tax rate on gains normally taxed below the top rate by more than it raised the top rate, and it dropped the top tax rate on short-term gains to that of long-term gains. One measure simply cannot capture all the variation in the tax rates that apply to capital gains.

Nonetheless, changes in the top statutory tax rate during the past 40 years have coincided with shifts in the level of realized gains--as would be expected to occur as people responded to a changing cost of realizing gains. When the top tax rate on realized gains began rising in 1968, realizations stopped rising relative to GDP (see Figure 6).[14] During the next 10 years, while the top rate rose and then plateaued, realizations fell relative to GDP and then remained low. Only after the top rate started falling in 1978 did realizations return to the high levels relative to GDP that they had reached in the 1960s. The spike in realizations in 1986 anticipated the rate increase that occurred in 1987, as previously discussed, and some of the decline in realizations relative to GDP after 1988 may reflect the increase in rates in 1987 and the surge in realizations in 1986.

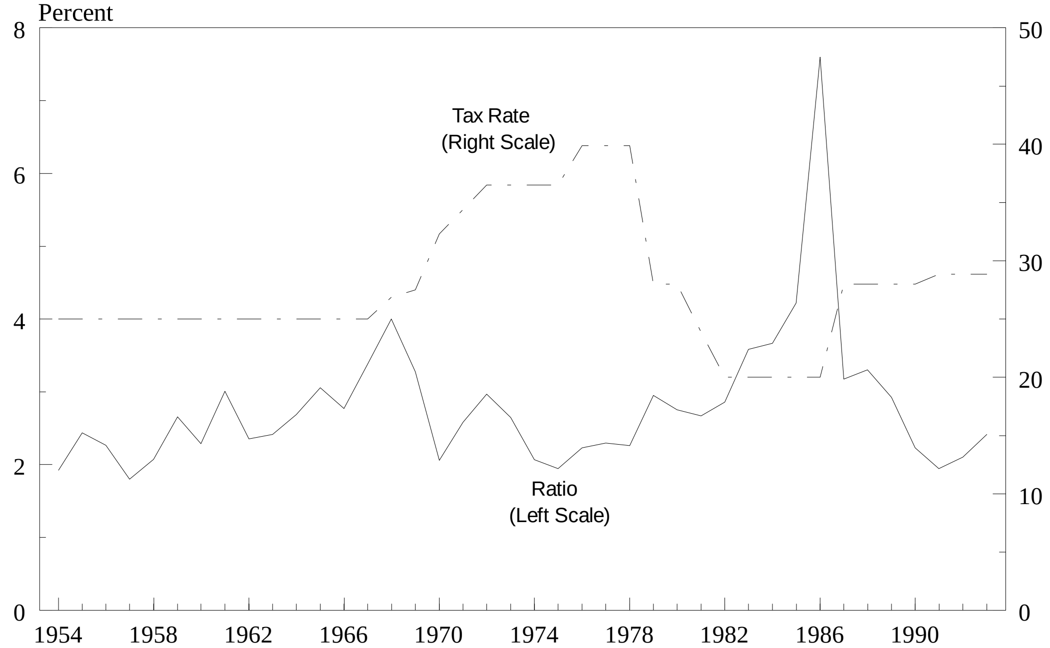
---

12. The top statutory tax rate for 1954 through 1988 is reported in Congressional Budget Office, *How Capital Gains Tax Rates Affect Revenues*, p. 36. Three alternative top rates are shown there. In the current analysis, CBO uses the middle one, which reflects the effects of the ordinary rate structure, the exclusion, and the alternative plus minimum tax. Since 1987, the top rate on capital gains stated in the tax code has been 28 percent. However, since 1991, the limitation on itemized deductions has raised the effective rate on taxpayers with the very highest incomes to nearly 29 percent.

13. The fraction of gains received by taxpayers paying the top rate between 1954 and 1986 is based on information on the fraction of gains received by the top 1 percent of taxpayers ranked by adjusted gross income (see Congressional Budget Office, *How Capital Gains Tax Rates Affect Revenues*, p. 31). The top 1 percent of taxpayers typically paid the top rate on capital gains. The fraction of gains received by taxpayers paying the top rate since 1987 is based on a tabulation of tax rates paid in 1992 and tabulations of total gains received by adjusted gross income classes for the years between 1987 and 1991.

14. Realizations are shown relative to GDP in this comparison to tax rates to keep the comparison between terms that are scaled in percentages. The text and Figure 6 both refer to realizations as a percentage of GDP. That percentage is the same as it would be if realizations and GDP were expressed in constant 1987 dollars.

FIGURE 6. RATIO OF CAPITAL GAINS TO GROSS DOMESTIC PRODUCT AND THE TOP STATUTORY TAX RATE ON CAPITAL GAINS



SOURCE: Congressional Budget Office based on Internal Revenue Service, Statistics of Income Division, *Individual Income Tax Returns* for years 1954 through 1992; unpublished Internal Revenue Service data for 1993; and Department of Commerce, Bureau of Economic Analysis.

NOTE: See text for information on construction of top statutory tax rate on capital gains.

Equations to Explain Realized Capital Gains

The discussion above suggests that the price level, real GDP, real stock values, housing starts, and capital gains tax rates would help to explain the amount of capital gains realized annually. A specific formulation of an equation with those variables is:

(1) $$\begin{aligned} \log(\text{GAINS}) = {} & c_0 + c_1 {*} \log(\text{PRICE LEVEL}) \\ & + c_2 {*} \log(\text{REAL STOCK VALUES}) \\ & + c_3 {*} \log(\text{REAL GDP}) \\ & + c_4 {*} \log(\text{STARTS}) \\ & + c_5 {*} \text{TAX RATE} \\ & + c_6 {*} \text{YEAR 1986} \\ & + \text{error term} \end{aligned}$$

That equation includes two refinements not yet discussed. One, a variable that takes on the value of 1 in 1986 and 0 in other years, is added to pick up the unique incentives to realize gains in that year. Two, the variables that grow over time are expressed in logarithms. That keeps the error term of the equation from growing over time and distorting the estimated standard errors of the parameters.[15]

The parameters of the equation could be estimated from the historical data described above using the statistical technique of ordinary least squares. The equation could then be used to forecast future realizations by using the estimated coefficients and forecasts of the explanatory variables.

Estimating equation (1) by ordinary least squares, however, could lead to over-estimates of the importance of the relationships between the explanatory variables and realized gains--that is, of the statistical significance of the parameters c_0 through c_6. In the extreme, causally unrelated variables could be found to be statistically related.[16]

The problem arises when the variables in the equation are not "stationary." Stationary is a term statisticians use to describe a variable whose mean and correlation with past values do not change as time passes. Realized gains, GDP, and stock prices have increased over time (as shown in Figures 2 and 4) and therefore are non-stationary in the period. Housing starts have not increased in the period (as shown in Figure 5) but starts may still be nonstationary if the correlations between current and past values are changing. If those variables are nonstationary, ordinary least squares could overstate the statistical significance of their coefficients.

The estimation problem can be avoided by transforming the variables until they are stationary. The appropriate transformation depends on the type of nonstationary process that is generating the observed path of the variables. Many economic time-series have been shown to follow a process called a random walk. In a random walk, the random influences on a variable in one year become fully embedded in the variable in the following year. Variables that follow a random walk have the following form:

$$X_t = X_{t-1} + e_t$$

That is, the value of the variable X in year t equals its value in the previous year plus a random element that is uncorrelated with previous levels of the variable.

---

15. Expressing dependent and explanatory variables as logarithms also causes the coefficients of those variables in equation (1) to measure the percentage change in realizations resulting from a 1 percent change in the explanatory variable.

16. C.W.J. Granger and Paul Newbold, *Forecasting Economic Time Series* (New York: Academic Press, 1977), pp. 202-207. This problem is also discussed in the context of time-series capital gains equations by Alan J. Auerbach, "Capital Gains Taxation in the United States: Realizations, Revenue, and Rhetoric," *Brookings Papers on Economic Activity*, no. 2 (1988), pp. 595-637; and Jonathan D. Jones, *An Analysis of Aggregate Time Series Capital Gains Equations*, Paper 65 (Department of the Treasury, Office of Tax Analysis, May 1989).

CBO's tests of several of the variables in equation (1) could not reject the hypothesis that they follow random walks.[17] The variables are realized gains, the price level, real stock values, real GDP, and housing starts (all expressed in logarithms). CBO did not test the tax rate because changes in it are likely to be influenced by the existing level of tax and because it has changed relatively infrequently.

Variables following random walks become stationary when transformed to measure first differences, which for these variables are the changes from one year to the next. Consequently, CBO transformed equation (1) to the following form:

$$
\begin{aligned}
(2) \qquad \Delta\log(\text{GAINS}) = {} & c_1 * \Delta\log(\text{PRICE LEVEL}) \\
& + c_2 * \Delta\log(\text{REAL STOCK VALUES}) \\
& + c_3 * \Delta\log(\text{REAL GDP}) \\
& + c_4 * \Delta\log(\text{STARTS}) \\
& + c_5 * \Delta\text{TAX RATE} \\
& + c_6 * \Delta\text{YEAR 1986} \\
& + c_7 * \Delta\Delta\log(\text{REAL GDP}) \\
& + c_8 \\
& + \text{error term}
\end{aligned}
$$

The term Δ indicates the change in a variable from the preceding year. Thus, the variable being explained by the equation is the change from the preceding year in the logarithm of realized gains. Because the change in logarithms of a variable approximates the percentage change in the variable itself, the equation can be thought of as explaining the annual percentage growth in realized gains. That growth is explained by the annual growth rates of the price level, real stock values, real GDP, and housing starts, as well as the percentage-point change in the statutory tax rate. An additional variable, the acceleration in real GDP, has been added to reflect how the speeding up and slowing down of economic activity over the business cycle affects the growth of realized gains. (Acceleration is measured as the change in the change in the logarithm of real GDP.)

Although transforming equation (1) to equation (2) reduces the risk that spurious relationships will be estimated, it also sacrifices valid information about the level of realized gains. Equation (1) provides the expected level of realized gains given the levels of the explanatory variables. It indicates that although random influences in particular years will cause realizations to diverge from their normal relationship to real GDP, stock values, and other determinants, realizations will tend to return to their normal level relative to those variables. Thus, equation (1) can be viewed as describing the determination of realizations over the long run.

---

17. The tests, which follow the procedures developed by D. Dickey and W. Fuller, are summarized in Appendix C.

In equation (2), the annual growth in realizations in one period depends only on the changes in that year in real GDP, real stock values, and the other variables. It does not depend on whether realizations in the preceding year were above or below their long-run level relative to the levels of the explanatory variables in equation (1). If realizations had been below their long-run level, equation (1) implies that growth in the current year would tend to be larger than indicated by the current changes in explanatory variables. That is, realizations would show a tendency to bounce back toward their long-run level.

The tendency for annual changes to move toward the long-run levels implied by equation (1) can be captured in an equation of annual changes by including error correction terms. In equation (2), that can be achieved by including the previous year's error term from equation (1). Tests have been developed to determine whether an equation describing a long-run relationship between the levels of variables, such as equation (1), is appropriately estimated by a change equation, such as equation (2), with an error correction term. CBO's application of the tests suggests that estimating change equations with error correction terms from equation (1) is likely to be appropriate.[18]

Consequently, in addition to estimating equation (2), CBO estimates equation (2) with the addition of the previous year's error term estimated in equation (1). That is accomplished by including the values from the previous year of all the variables in equation (1).

Including those variables nearly doubles the total number of parameters to be estimated. That number may be more than can be reliably estimated from the available data. To reduce the number of coefficients that must be estimated, CBO also specifies a more restrictive version of the long-run determinants of realizations than that embodied in equation (1).[19]

In the more restrictive version, realized gains grow at the same rate as GDP over the long run, and changes in tax rates on capital gains alter the level of realizations relative to GDP. The long-run effects of stock and real estate values on realizations are assumed to be reflected by GDP, and therefore those variables can be excluded from the equation. That restrictive version of equation (1) is broadly consistent with the experience during the past 40 years, as shown in Table 1 and Figure 6 above. The restrictive version of equation (1) can be written as:

---

18. The tests CBO uses are described in Robert F. Engle and C.W.J. Granger, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, vol. 55, no. 2 (March 1987), pp. 264-270. The test results are summarized in Appendix C.

19. An alternative approach would be to estimate equation (1) first and add its estimated residual vector as a single variable in equation (2). That alternative would not capture the potential advantages described below of using a more restrictive version of equation (1).

$$\log(\text{GAINS}) = d_0 + 1*\log(\text{GDP}) + d_1*(\text{TAX RATE}) + d_2*(\text{YEAR 1986}) + \text{error}$$

The variable YEAR 1986 is still included to capture the unique effects of that year. The assumption that realized gains grow at the same rate as GDP is shown by having the coefficient of log(GDP) set to unity. That can be imposed in estimating the equation by transforming the equation to the equivalent form in which the ratio of realized gains to GDP varies with tax rates.

$$(3) \qquad \log(\text{RATIO}) = d_0 + d_1*(\text{TAX RATE}) + d_2*(\text{YEAR 1986}) + \text{error}$$

Even if realized gains are assumed to follow equation (3) in the long run, they can still follow equation (2) in the short run. Over the business cycle, realizations can grow at a different rate than GDP. Furthermore, the value of stock and real estate holdings can grow faster or slower than GDP in particular years, and that difference in growth can affect realizations.

As long as equation (3) represents the long-run and equation (2) represents the short-run determinants of realized gains, adding the previous year's error term from equation (3) to equation (2) should pick up the tendency for realizations to bounce back toward their long-run level. An additional requirement for estimation is that the previous year's error be a stationary variable. That requirement also appears to be satisfied. The ratio of realized gains to GDP has not tended to rise or fall during the past 40 years, and CBO's statistical tests indicate that it does not follow a random walk (see Appendix C). Because the other variables in equation (3) are also stationary, the error term should be stationary as well.

In summary, CBO estimates three equations to explain the growth of realized gains. Those are equation (2), equation (2) with the error correction terms from equation (1), and equation (2) with error correction terms from equation (3).

In addition, CBO estimates each of those three equations without the variable of housing starts. Housing starts give an indirect measure of the value of real estate holdings, and the link between starts and real estate values could change over time. Changes would lead to biased predictions of how starts will influence realizations in the future. For example, housing starts in recent years could be affected by the introduction of the low-income housing tax credit, which could cause them to diverge from their historically estimated relationship to realizations. Estimating equations without starts allows CBO to make an alternative forecast that is free of the uncertain effect of housing starts on realized gains.

Statistical Estimates of the Equations

For the January 1995 baseline, CBO estimated the equations with data for the 39 years from 1955 through 1993. As of November 1994, when the equations were estimated, final data on realized gains in 1993 were unavailable. CBO used a value of $145 billion based on tax returns filed through August 1994. Use of the final figure for 1993 of $153.26 billion would not materially alter the estimated equations.

Least squares regression produces estimates of the coefficients for individual variables in equation (2) and statistics describing the goodness of fit of the entire equation. The estimated coefficients for individual variables are described first.

Estimated Coefficients. The estimated coefficients of equation (2) confirm the trends observed in Figures 2 through 6. Faster growth of real GDP, real stock values, and housing starts causes faster growth of realized gains. An increase in the tax rate on realized gains slows their growth. Those estimated coefficients are all more than twice the size of their estimated standard errors, which indicates that they are all significantly different from zero at the 95 percent confidence level. The estimated coefficients and related statistics are reported in the first column of Table 2 for the version of equation (2) with housing starts and in the first column of Table 3 for the version without starts.

In the version of equation (2) with housing starts, the estimated coefficients indicate that a 1 percent increase in the real value of corporate equities would cause realized gains to grow by one-half a percent, and a 1 percent increase in housing starts would cause realized gains to grow by one-quarter of a percent. A 1 percent increase in real GDP is estimated to cause realized gains to grow by over 3 percent. That strong effect by GDP suggests that growth in GDP is reflecting more than the growth in the stock of assets. It may also be reflecting the growth of asset values other than corporate stocks or real estate, and it may be reflecting a tendency for people to sell more assets during the upswing of the business cycle.

The tax coefficient in the first equation of Table 2 literally means that a change in the top tax rate of 1 percentage point--say, from 28 percent to 27 percent--would cause realizations in that year to grow 2.42 percent faster. Analysts frequently refer to the elasticity of the realizations' response to a tax change rather than the coefficient estimated in equation (2). The elasticity is the percentage change in realizations resulting from a 1 percent change in tax rates. The estimated coefficient implies that a 1 percent reduction in the tax rate, from 28 percent to 27.72 percent, would cause realizations to grow 0.68 percent faster. That elasticity of the tax rate becomes closer to zero at lower tax rates.

Although the tax rate coefficient is statistically significant, it provides only a rough guide to the long-term effects of changing tax rates on capital gains, for several reasons. First, the top statutory rate is only a proxy for the full tax schedule. A broader representation of the full schedule would be needed to give an accurate estimate of how a particular legislative change would affect all realizations.[20] Second, the equation assumes that a tax change causes an immediate shift to a new long-run level of realizations. The initial response will not necessarily be the same as the long-run response, however, because the level of accrued gains held by taxpayers will change as realizations adjust to the new tax rate. The estimated coefficient is likely to be a combination of initial and long-term responses. Third, previous studies have found that the estimated coefficient of the tax rate in equations like equation (2) is sensitive to the specification of other variables, such as wealth, accrued gains, and anticipated future changes in tax rates.[21]

The estimated coefficients for the rate of inflation and for the unique effect of the Tax Reform Act of 1986 are large and more than twice their standard errors in the first equation of Table 2. The coefficient and standard error of the inflation rate, however, fluctuate considerably in other versions of the equation and therefore are not precisely estimated (see Tables 2 and 3).

A coefficient for the acceleration in real GDP is not reported in Table 2, although the variable appears in equation (2). That variable was omitted from the reported regression because it had a statistically insignificant effect in the original version of equation (2).

When equation (2) is estimated without housing starts, the coefficients of most other variables change little (see the first column of Table 3). Exceptions are the coefficients for the inflation rate and the acceleration of GDP. The inflation rate becomes statistically insignificant, and the acceleration of real GDP becomes marginally significant.

---

20. Estimating the equations with an average tax rate is one way to represent the full schedule of tax rates on capital gains, but that would make the tax variable dependent on taxpayer behavior (see footnote 11). Proper estimation with such a tax rate requires estimation with instrumental variables, as pointed out in Robert Gillingham and John S. Greenlees, "The Effect of Marginal Tax Rates on Capital Gains Revenue: Another Look at the Evidence," *National Tax Journal*, vol. 45, no. 2 (June 1992), pp. 167-177. For purposes of forecasting, which CBO undertakes, instruments such as the top statutory tax rate can be entered directly into the equation. To represent the tax effects fully, however, additional instruments are needed.

21. Congressional Budget Office, *How Capital Gains Tax Rates Affect Revenues*, pp. 73-95; Auerbach, "Capital Gains Taxation in the United States"; and Jones, *An Analysis of Aggregate Time Series Capital Gains Equations*. For a recent analysis of how tax rates affect realized gains, see Leonard E. Burman and William C. Randolph, "Measuring Permanent Responses to Capital-Gains Tax Changes in Panel Data," *American Economic Review*, vol. 84, no. 4 (September 1994), pp. 794-809.

TABLE 2. EXPLAINING ANNUAL CHANGE IN LOGARITHM OF REALIZED GAINS EQUATIONS WITH HOUSING STARTS, REGRESSION ESTIMATES, 1955-1993

	Pure Change	Flexible Error Correction	Restricted Error Correction
Variable Coefficients (Standard errors in small type)			
Δlog(PRICES)	1.84 0.88	1.76 1.99	2.62 1.23
Δlog(RSTOCKS)	0.50 0.10	0.33 0.14	0.42 0.11
Δlog(RGDP)	3.35 0.96	3.74 1.08	3.77 1.00
Δlog(STARTS)	0.24 0.07	0.26 0.08	0.22 0.07
ΔTAX RATE	-2.42 0.70	-2.65 0.75	-2.12 0.73
ΔYEAR 1986	0.56 0.08	0.66 0.12	0.63 0.12
$\log(\text{GAINS})_{t-1}$		-0.56 0.17	
$\log(\text{RATIO})_{t-1}$			-0.23 0.12
$\log(\text{PRICES})_{t-1}$		0.19 0.32	
$\log(\text{RSTOCKS})_{t-1}$		-0.04 0.14	
$\log(\text{RGDP})_{t-1}$		1.30 0.66	
$\log(\text{STARTS})_{t-1}$		0.08 0.08	
TAX RATE_{t-1}		-1.65 0.73	-0.48 0.53
YEAR 1986_{t-1}		0.51 0.24	0.23 0.22
Constant	-0.11 0.06	-3.17 1.00	-0.87 0.38

(Continued)

TABLE 2. CONTINUED

	Pure Change	Flexible Error Correction	Restricted Error Correction
Summary Statistics			
R-Squared	0.85	0.90	0.87
Standard Error	0.10	0.10	0.10
Durbin-Watson	1.93	1.88	1.96

SOURCE: Congressional Budget Office.

NOTES: Δ= Difference between current and previous year of variable
log = Natural logarithm of variables
PRICES = GDP deflator
RSTOCKS = Value of corporate equities held by households divided by the GDP deflator
RGDP = Real gross domestic product
STARTS = Starts of dwellings in structures with two or more dwellings
TAX RATE = Top statutory tax rate on capital gains, in percent
YEAR 1986 = Variable with value of 1 in 1986, 0 in other years
GAINS = Capital gains reported on tax returns with net gains

TABLE 3. EXPLAINING ANNUAL CHANGE IN LOGARITHM OF REALIZED GAINS EQUATIONS WITHOUT HOUSING STARTS, REGRESSION ESTIMATES, 1955-1993

	Pure Change	Flexible Error Correction	Restricted Error Correction
Variable Coefficients (Standard errors in small type)			
Δlog(PRICES)	1.30 1.02	1.07 2.11	2.86 1.38
Δlog(RSTOCKS)	0.52 0.11	0.33 0.15	0.47 0.12
Δlog(RGDP)	3.42 1.29	5.64 1.01	5.22 0.99
$\Delta\Delta$log(RGDP)	1.43 0.91		
ΔTAX RATE	-2.76 0.79	-2.66 0.86	-2.05 0.82
ΔYEAR 1986	0.57 0.09	0.64 0.13	0.62 0.13
log(GAINS)$_{t-1}$		-0.53 0.19	
log(RATIO)$_{t-1}$			-0.28 0.13
log(PRICES)$_{t-1}$		-0.04 0.23	
log(RSTOCKS)$_{t-1}$		-0.14 0.13	
log(RGDP)$_{t-1}$		1.62 0.69	
TAX RATE$_{t-1}$		-1.96 0.84	-0.72 0.59
YEAR 1986$_{t-1}$		0.41 0.28	0.18 0.25
Constant	-0.09 0.07	-2.68 0.99	-1.01 0.43

(Continued)

TABLE 3. CONTINUED

	Pure Change	Flexible Error Correction	Restricted Error Correction
Summary Statistics			
R-Squared	0.81	0.85	0.83
Standard Error	0.12	0.11	0.12
Durbin-Watson	2.06	1.93	2.00

SOURCE: Congressional Budget Office.

NOTES: Δ= Difference between current and previous year of variable
log = Natural logarithm of variables
PRICES = GDP deflator
RSTOCKS = Value of corporate equities held by households divided by the GDP deflator
RGDP = Real gross domestic product
TAX RATE = Top statutory tax rate on capital gains, in percent
YEAR 1986 = Variable with value of 1 in 1986, 0 in other years
GAINS = Capital gains reported on tax returns with net gains

When the error correction terms are added to equation (2), realizations are found to rebound from years in which the level of realizations is particularly low or high relative to its major determinants. The estimated rebound is both large and statistically significant.

When the error correction terms are taken from equation (1)--labeled flexible error correction--the coefficient of $\log(\mathrm{GAINS})_{t-1}$ estimates the size of the rebound. Its value of 0.56 in the equation with housing starts indicates that 56 percent of last year's deviation of realized gains from their long-run level, as described in equation (1), is made up in the current year (see Table 2). The estimated bounce back is of similar size in the corresponding equation without housing starts (see Table 3). The coefficients of $\log(\mathrm{GAINS})_{t-1}$ in these two equations are about three times their standard errors, indicating that the estimates of a sizable rebound are unlikely to be a statistical fluke.[22]

When the error correction terms are taken from equation (3)--restricted error correction--the coefficient of $\log(\mathrm{RATIO})_{t-1}$ estimates the size of the rebound. Its size indicates that for each 1 percent that log(RATIO) in the previous year is below its long-run level, realized gains in the current year will grow 0.23 percent faster (see Table 2). The estimated coefficient is about twice its standard error, which is large enough to make its size unlikely to result from a statistical fluke.[23] The estimated size of the rebound and its statistical significance are similar in the equation without housing starts (see Table 3).

The size of the rebound is about half as large when the restricted instead of flexible error correction terms are used. The main reason is that the rebound with the restricted error correction terms is the rebound in gains relative to GDP, whereas the rebound with the flexible error correction terms is in the level of gains itself. One estimate is not clearly preferable to the other, but either is preferable to an estimate of no rebound.

As already noted, housing starts make a statistically significant contribution to equation (2) when error correction terms are omitted. It also makes a statistically significant contribution when either set of error correction terms is included.

---

22. The significance of the full set of error correction terms needs to be judged from estimating equations in which the actual residuals from equation (1) are used in place of the full set of equation (1) variables. When the actual residuals are substituted for the error correction terms reported in Table 2 and Table 3, the coefficients of the residuals are of similar size and statistical significance as the coefficients of $\log(\mathrm{GAINS})_{t-1}$. The inadequacy of judging statistical significance of the error correction terms in the equations reported in Table 2 and Table 3 is discussed in Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics* (New York: Oxford University Press, 1993), p. 725.

23. Reestimation of the equations with the actual residuals from equation (3) instead of the list of variables gives similar estimates of the size and statistical significance of the rebound.

The Equations' Goodness of Fit. The equations are able to explain a large fraction of the growth of realized gains. The yearly variation in the growth of gains is so large, however, that the unexplained residual variance remains sizable. Goodness of fit statistics are shown at the end of Tables 2 and 3.

Equation (2) with housing starts but no error correction terms can explain 85 percent of the growth of realized gains, as measured by the R-squared statistic. If the full set of flexible error correction terms is included, the equation can explain up to 90 percent. If housing starts are omitted, the fraction of growth explained drops by 4 to 5 percentage points. Thus, including housing starts contributes as much to explaining the growth of realized gains as does including all the error correction terms.

In spite of the large fractions of growth explained by the equations, the annual variation in growth is so large that the remaining unexplained variation is problematic for forecasting revenues. As shown in Figures 1 and 3, the variability of realized gains is not limited to the extraordinary growth and decline of 1986 and 1987 caused by the Tax Reform Act of 1986. The standard deviation of the growth of gains in years other than 1986 and 1987 is 18 percent compared with a mean of 8.6 percent.

The standard error statistic gives a measure of the typical error of an equation. The standard errors of the equations reported in Tables 2 and 3 range between 10 percent and 12 percent. Assuming that an equation's errors are normally distributed, a 10 percent standard error means that an equation predicting a typical growth rate of 8.6 percent in a particular year between 1955 and 1993 can expect to have the actual growth rate lie between -1.4 percent and 18.6 percent two-thirds of the time. That is a wide range, and one-third of the time actual growth is expected to fall outside the range.

The estimated equations have particularly large errors in 1989 through 1991, during which time realized gains dropped 37 percent. The version of equation (2) with housing starts but no error correction term can explain a decline of just 4 percent. When the full error correction terms are added, the equation can explain a decline of 19 percent, and when the restrictive set is added the equation can explain a decline of 12 percent. The versions of the equations without starts have much larger errors because housing starts are the only predictor of realized gains that declines throughout the period.[24]

The goodness of fit of an equation estimate is also measured by how independently the estimated errors vary from year to year. If the errors in one year

24. The equation without housing starts or error correction predicts 9 percent growth during the three years; the equation with the full set of error correction terms added predicts a 10 percent decline; and the equation with the restricted set added predicts no net change.

depend on the errors in the preceding year, for example, the variance of the estimated parameters is increased. Estimating equations as annual changes, as CBO does, frequently removes such correlations, and that appears to be the case in CBO's equations. A test of the hypothesis that the errors in one year are independent of those in the preceding year (that is, have no first-order autocorrelation) is provided by the Durbin-Watson statistic. Its value in equation (2) estimated with housing starts is 1.93 and without housing starts is 2.06. Those are both high enough that the hypothesis of no first-order autocorrelation cannot be rejected. A visual scan of higher-order autocorrelations and partial correlations in the errors suggests that they are not important in CBO's equation estimates either.

PROJECTING FUTURE REALIZATIONS OF CAPITAL GAINS

In early December, when the Congressional Budget Office's new macroeconomic forecast becomes available, CBO combines information from that forecast and other sources with the estimated equations in order to forecast realized gains for the year just ending. Then, CBO projects gains for the next six years using a trend method based on the macroeconomic forecast of gross domestic product and on current tax rates. Finally, CBO adjusts the seven-year projection for events that seem likely to affect realized gains but that are not adequately captured in the first two steps.

Forecasting with the Equations

The process of forecasting with the equations is illustrated by describing the process for forecasting realizations in 1994. In November 1994, when the equations reported in Table 2 were estimated, all historical data for estimating the equations through 1993 were available except for realized capital gains in 1993. As noted above, CBO estimated the equation using a preliminary figure of $145 billion.

When its new macroeconomic forecast became available in early December, CBO used the equations to forecast the growth of realized gains in 1994. The values of the explanatory variables were taken from available information for the year to date plus CBO's forecast of variables for the remainder of the year. In particular, CBO used the year's growth to date of the New York Stock Exchange composite index to measure the growth of stock values for all of 1994. CBO used the level of GDP, prices, and housing starts as reported for the first three quarters of 1994 combined with CBO's macroeconomic forecast for those variables in the fourth quarter. By early December, it was apparent that statutory tax rates on capital gains would remain unchanged through December.

Using the 1994 inputs for the explanatory variables, CBO forecast the rate of growth in realizations for 1994 with each of six equations (see Tables 2 and 3). CBO

TABLE 4. EQUATION FORECASTS OF REALIZED CAPITAL GAINS IN 1994

Equation Specification		Growth Rate	Gains
Housing Starts	Error Correction	(In percent)	(In billions of dollars)
Yes	No	12.3	164
Yes	Flexible	13.6	166
Yes	Restricted	13.8	167
No	No	4.0	151
No	Flexible	5.8	154
No	Restricted	5.1	153

SOURCE: Congressional Budget Office.

then applied each growth rate to the preliminary 1993 level of realizations to determine a level for 1994. CBO arrived at a single forecast by taking the mean of the six forecasts.[25]

Equation (2) with housing starts but no error correction term predicted that realizations grew by 12.3 percent (see Table 4). The corresponding equation without housing starts but with the acceleration of GDP predicted growth of just 4 percent. That large difference is primarily attributable to the rapid growth of housing starts. As of December 1994, housing starts were on a path to grow 41 percent for the year. The equations with error correction terms predicted faster growth than the simpler versions of equation (2). The faster growth reflects a predicted rebound in gains from their below-normal level in 1993. Realized gains in 1993, judging from the preliminary data, were 2.29 percent of GDP, which is below the long-run average of 2.67 percent.

The growth rates predicted by the six equations imply dollar levels of realized gains in 1994 that range from $151 billion to $167 billion. The mean growth rate is 9.1 percent--which, based on the 1993 preliminary realizations of $145 billion, implies

25. In other years, when four or five of the forecasts were similar, CBO used the median. The mean and the median gave nearly the same result in December 1994.

1994 realizations of just under $160 billion. CBO rounded its 1994 forecast to $160 billion.[26]

Each equation's forecast is accompanied by an estimated standard error based on the standard error of the estimated equation and the values of the explanatory variables. The standard error of the forecast indicates the range within which the actual growth rate is likely to lie. The standard errors of the six forecasts are between 11 percent and 13 percent, slightly higher than the standard errors within the estimation period. Those forecast errors imply that about two-thirds of the time the actual growth rate should be within plus or minus 11 percent to 13 percent. For a predicted 9 percent growth rate, for example, a 12 percent forecast error implies that two-thirds of the time the true growth rate should lie between -3 percent and 21 percent. The amount of gains for 1994, based on that range, should be between $140 billion and $175 billion two-thirds of the time. The error range is based on realized gains in 1993 of $145 billion to maintain consistency with the predicted growth rates.

The actual error range for an equation's forecast could be even larger than the range predicted by the regression. The regression estimate of the error assumes that the estimated equation accurately describes the process by which realized gains are determined, the explanatory variables are fully known for the forecast year, and the equation's errors are independently and normally distributed. Those conditions are not fully met, although the errors appear to be independent from year to year. Because the regression's predicted error of the forecast may underestimate the actual error of the forecast, CBO has calculated alternative measures of each equation's forecast error for earlier years. Those measures were obtained by making forecasts one year ahead for 1989 through 1993 and comparing them with actual outcomes. The root mean squared errors for the equations range from about 9 percent to 20 percent, with most equation errors between 11 percent and 14 percent (see Appendix A for details). Given the small number of years included, those errors are quite close to the standard errors calculated from the regression statistics. Nonetheless, both estimates of the forecast errors make it clear that a large degree of uncertainty surrounds the forecasts.

Projecting Realizations for Six Future Years

Forecasting realizations with the equations is more difficult for future years than for the year just ending. Much information about the explanatory variables is available

26. If the actual value of realized gains for 1993 had been available, the 1993 base to which the predicted growth rates are applied would have been $8 billion higher. Applying the predicted growth rates to the higher base would lead to proportionately higher predicted values for 1994. The growth rates estimated from the equations with error correction terms, however, would have been lower. The higher 1993 base would lead to a smaller predicted rebound for 1994. The net effect would have been a somewhat higher dollar forecast for 1994.

for the year just ending, but their values in future years must be forecast. Forecasting stock values and housing starts is particularly difficult. Changes in stock values cannot be reliably predicted even one year ahead, and although changes in housing starts are more predictable one year ahead, their pattern through the next six years cannot be reliably predicted.

When a large degree of uncertainty about the equation inputs must be added to the large standard errors of the equation forecasts, the advantage of forecasting with the equations is likely to disappear. Alternative methods that require less exacting information about the future may have smaller ranges of error.

In addition, alternative methods are likely to give more consistent projections from one baseline to the next. Because of the great uncertainty about future housing starts and stock market values, their forecasts for a particular future year tend to differ considerably from one forecast time to the next, even though the newer forecast is not much more likely to occur than the older one. The changing forecast causes large year-to-year changes in the level of realizations predicted by the above equations for a single future year. Those differences in realization must be explained with each new forecast, but little confidence can be placed in the greater accuracy of the newer forecast.

As an alternative to forecasting future realizations with the equations, CBO uses an assumption that future realizations will tend to follow GDP. The overall path of the economy, as measured by GDP, is more predictable than the paths of stock prices or housing starts, and, as shown in Figure 2, realizations of capital gains have generally followed the trend of GDP. CBO's specific assumption is that realizations in the coming six years will return to their average size relative to GDP.[27]

Realizations would not be expected to return to their normal size relative to GDP if tax rates on realizations were above or below their average. The ratio of realizations to GDP has tended to be below average when tax rates are high and above average when tax rates are low. Currently, CBO does not make an explicit adjustment in its target ratio because the current top statutory rate is nearly identical to its historical average (see discussion on next page).

Between 1954 and 1992, excluding the peak of 1986, realized gains averaged 2.67 percent of GDP. As pointed out above, the equations forecast that realizations will be $160 billion in 1994. That amount represents 2.38 percent of GDP as predicted in CBO's macroeconomic forecast. The percentage suggests that realiza-

---

27. CBO does not expect that realizations will steadily return to their historical average ratio because realizations have not followed a simple trend relative to GDP in the past. Nonetheless, because of the difficulty of forecasting the actual path of realizations in the coming years, CBO's projection may have smaller expected errors over the entire period than a forecast that attempts to predict the annual ups and downs of realized gains.

tions in 1994 continued to recover relative to GDP from the low value reached in 1991 (see Table 1). The projection for realized gains assumes that the ratio will rise linearly from its 1994 value to 2.67 percent of GDP by the year 2000. The macroeconomic forecast has GDP reaching $9,128 billion in that year, which puts realizations for 2000 at $243.7 billion (see Table 5 for the year-by-year path of GDP and realizations).[28]

Two influences on realized capital gains will tend to push gains away from their historical relationship to GDP in the coming six years. First, the average taxpayer is paying a higher rate today than was typically paid in the past. Second, realizations show a slight upward trend relative to GDP. Those influences push in opposite directions and therefore may offset each other. That would leave capital gains tending toward their historical relationship to GDP.

Higher average tax rates would tend to push gains below their average as a percentage of GDP. The current top statutory rate of 28 percent equals the historical average of the top rates, but the tax rate among all taxpayers realizing gains in recent years has been higher than in most previous years. The average rate has been high since 1987, when the capital gains exclusion was replaced by a cap on the tax rate applicable to capital gains. If tax rates remain unchanged through 2000, as is assumed for the baseline forecast, those higher rates should dampen realizations relative to GDP.

The slight upward trend in the ratio of gains to GDP would tend to push gains above their average as a percentage of GDP. The trend, though barely apparent in Figure 6, is large enough that if it were followed through 2000, realizations in that year would be 2.93 percent of GDP. That trend may result from taxpayers' becoming more and more proficient at characterizing ordinary income as capital gains to take advantage of the lower tax rate. Or it may reflect the falling transaction costs of trading assets, which would lead people to realize more of what they accrue.

The above-average tax rates in effect since 1987 may keep realized gains from growing to 2.93 percent of GDP, but they will not necessarily lower the ratio below its historical average of 2.67 percent. Since tax rates were raised in 1987, realizations have been above that percentage of GDP in 1987, 1988, and 1989.

---

28. The smoothness of the trend in realizations caused by assuming that they return to their average size relative to GDP is accentuated by the smoothness in CBO's projection of GDP. CBO's projection of GDP after the first two years is itself a steady trend. CBO's macroeconomic forecasters try to predict the actual path of GDP during the first two years (1995 and 1996 in the January 1995 forecast), but for the remaining years (through 2000 in this case), they assume that real GDP will grow so that by the end of the forecast period the gap between real GDP and potential GDP reaches its historical average. In the January 1995 forecast, real GDP is assumed to grow 2.3 percent from 1997 through 2000. Inflation is also projected to proceed at a steady rate in this period, causing actual GDP to rise by about 5.2 percent a year.

TABLE 5. TREND FORECAST OF REALIZED GAINS

Year	Gross Domestic Product (In billions of dollars)	Ratio of Gains to GDP (In percent)	Gains (In billions of dollars)
1994	6,735	2.38	160.0
1995	7,127	2.42	172.6
1996	7,456	2.47	184.2
1997	7,847	2.52	197.6
1998	8,256	2.57	212.0
1999	8,680	2.62	227.3
2000	9,128	2.67	243.7

SOURCE: Congressional Budget Office.

Forecasting or Projecting the First Year

The method used to project realizations for the six future years of the forecast could easily be extended to project realizations for the year ending as well. The purpose of using the equations instead is to improve the accuracy of the first year's estimate by using more indicators of realizations activity. It could turn out, however, that the equations are no better able to predict realizations in that first year than would an extension of the projection method. After all, the equations have relatively large forecast errors. A comparison of the equations' and the projection method's predictions one year ahead for the years 1989 through 1993 is described in Appendix A. In that comparison, most of the equations have lower root mean squared errors than the projection method. Using equations, at least in those years, apparently does give a more accurate prediction.

Adjusting for Omitted Influences

Once the forecasts from the equation and the trend projection are constructed, CBO modifies the entire seven-year projection to account for other changes that those methods do not capture. For example, CBO adjusted the January 1995 projection to reflect an increased likelihood of a future change in taxation of capital gains.

The outcome of the November 1994 elections substantially increased the likelihood of a capital gains tax reduction in the near future. In response, some people who were planning to sell assets at the end of 1994 probably postponed those sales in hopes of paying less tax on their gains. The sharp increase in realizations in 1986 clearly shows that people will adjust the timing of their asset sales to take advantage of anticipated changes in capital gains taxes.

In December 1994, CBO modified the above projection to reflect the deferral of some sales from 1994 to 1995 and 1996. CBO's equations predicted that with no anticipated change in tax rates, about $160 billion of gains would have been realized in 1994. Based on data from 1985, about 17 percent, or $28 billion, of those gains would have been realized in the final two months of the year.[29] CBO anticipated that $10 billion of those gains were deferred because of the prospect for a lower tax rate in 1995 or 1996. That reduced the 1994 forecast to $150 billion.

Because CBO forecasts the baseline under current law, it must assume that no change in legislation occurs. CBO assumes that if no change occurs, the sales postponed in 1994 will occur in 1995 and 1996. CBO projects that realizations would be $5 billion higher in each year than they otherwise would have been. (The net effect of the $10 billion shift in realizations is incorporated in the projection shown in Table 6.) This final projection is the one used in CBO's January 1995 baseline projection of income tax revenue.[30]

The projection method described in this paper has been used for the January baselines of 1992 through 1995. Information about earlier CBO projections and methods can be found in Appendix B.

---

29. A sample of tax returns for 1985 shows that 17.6 percent of capital gains on stocks were claimed in the last two months. See Leonard Burman, Kimberly Clausing, and John O'Hare, "Tax Reform and Realizations of Capital Gains in 1986," *National Tax Journal*, vol. 47, no. 1 (March 1994), p. 13. Sales of stocks are likely to be easier to delay than sales of other assets yielding capital gains, such as businesses and real estate. Data for recent years and for additional asset categories have not been tabulated.

30. Preliminary tax return data through August 1995 suggest that realized gains will be near $137 billion for 1994. That lower amount could reflect the error range predicted by the equations, a larger deferral of realizations in response to the anticipated tax changes, or a combination of the two. The projections for the following years would be affected as well, with the size of the effect being largest in the most recent years, especially if the lower figure is a response to anticipated tax changes. Final data on realized gains for 1994 could tell a different story from the current one, and they are not likely to be available until September 1996.

TABLE 6. PRELIMINARY AND PROJECTED CAPITAL GAINS, JANUARY 1995
(In billions of dollars)

Year	Gains
1993 (Preliminary)	145.0
1994	150.0
1995	177.6
1996	189.2
1997	197.6
1998	212.0
1999	227.3
2000	243.7

SOURCE: Congressional Budget Office.

APPENDIX A: TESTS OF EQUATION FORECASTS

The Congressional Budget Office's procedure for projecting realized gains starts with a forecast one year ahead using the equations discussed in the body of the paper. This appendix reports the results of new tests of the equations' forecasting ability.

The tests compare the forecasts of the equations for past years to actual outcomes and to alternative forecasts made using a simpler method of projecting a trend. The comparison to actual outcomes measures the size of the forecasting error of the current method. The comparison to the projection of trend measures what, if anything, is gained by using the equations instead of a simpler method.

In the tests, CBO estimates each equation over the historical period ending in a particular year, and then uses that estimated equation to forecast the next year. For example, equation (2) is estimated from 1955 through 1988, and that estimated equation is used to predict the growth of realizations in 1989. That prediction is then compared with actual growth in 1989 to measure its accuracy. It is also compared with the trend projection to measure its forecasting advantage, if any.

The test process is repeated four times with each equation. Each repeat adds the next year to the estimation and then again forecasts one year ahead. Forecasts are made for a total of five years,1989 through 1993.

CBO tests eight equations. In addition to the six equations of Tables 2 and 3, it includes two equations explaining the level of realized gains. One of those is similar to equation (1) and the other is similar to equation (1) without housing starts. The only difference is that both of those equations include the change in the logarithm of gross domestic product to explain fluctuations in realized gains over the business cycle. For simplicity, the equations are referred to as equation (1) in the rest of this appendix.

Thus, one group of four equations includes equation (1) plus the three change equations of Table 2, all of which include housing starts. The other set of equations includes equation (1) without housing starts plus the three change equations of Table 3, none of which includes housing starts.

The forecasts for one year ahead use actual year-end data for the explanatory variables. CBO's baseline forecasts are made in December of the year being forecast, and by then CBO has a good but not perfect estimate of what the year-end values will be. Consequently, the accuracy of the forecasts in the test will be slightly better than for CBO's baselines.

The ultimate purpose of using the equations is to forecast the level of realized gains. But the equations used for forecasting since 1992 predict only the annual growth of realizations. CBO shows how reliably the equations forecast both the level and the growth rate of realized gains.

In evaluating the tests, one should keep in mind that five forecasts are too few to draw definitive conclusions about how well an equation is doing. Random events can affect realizations in such a short period in ways that misstate the ability of all equations to predict and the relative ranking of abilities among the equations. Nonetheless, the tests do provide a picture of how the equations would perform in circumstances like those of the recent past.

Accuracy in Predicting the Level of Realized Capital Gains

All equations predicted that realizations would be higher than they actually were in 1989, 1990, and 1991 (see Tables A-1 and A-2). No equation predicted the depth of the decline that occurred when realizations plunged from $162 billion in 1988 to $112 billion in 1991.

All of the change equations did better than the level equations between 1990 and 1992. The change equations did better in part because they only predict growth from last year's actual level of realizations. When an unexplained dip in realizations persists from year to year, forecasts of the change equations will incorporate that dip even if they cannot explain the growth rate more accurately than the level equations can. The level equations, in contrast, give no greater weight to last year than to any other year.

The errors of the change equations were much smaller in the last two years, when realizations were rising, than in the first two years. Not only were errors in the last two years smaller, they were negative instead of positive as in the first three years. In the last two years, the change equations consistently predicted less growth than occurred. The change equations that included housing starts actually improved their accuracy for the middle year as well as for the last two years.[1]

---

1. Errors for 1993 are calculated from the preliminary level of realized gains used to estimate the equations (see Table A-1). If the final figure was used, the size of the errors would most likely be larger in 1993, but the relative rankings of the equations would probably be unchanged.

TABLE A-1. ACTUAL AND FORECAST REALIZED CAPITAL GAINS
(By calendar year, in billions of dollars)

	1989	1990	1991	1992	1993
Actual	154	124	112	127	145[a]
Forecasts Without Starts					
Equation (1)	190	180	162	164	168
Pure change	184	147	133	126	142
Flexible e.c.	196	163	141	124	141
Restricted e.c.	178	142	133	126	144
Forecasts With Starts					
Equation (1)	184	173	146	145	149
Pure change	182	142	117	119	137
Flexible e.c.	180	147	119	220	135
Restricted e.c.	175	140	119	124	140
Trend Method	169	161	132	124	139

SOURCE: Congressional Budget Office.

NOTE: e.c. = error correction terms included.

a. Based on preliminary data. Final figure is $153 billion.

Although all equations forecast more accurately in the last two years than in the first two years, the equations without starts improved much more. In the first two years, the equations with starts all had smaller errors than their counterparts without starts. In the last two years, the equations without starts did slightly better than the equations with starts. Overall, the improvement from adding starts in the first three years outweighed the drawback in the last two years. The root mean squared error is lower for each equation with starts than it is for its counterpart without starts.

Finally, the resurgence of capital gains realizations in the last two years is bringing annual amounts back into the range predicted by the level equation with starts. In 1993, for the first time in the five-year test period, the level equation with starts predicted a level of realizations that is both close to the actual level and as close as the predictions of the change equations. The prediction of equation (1) without starts is also improving in the last two years, but is still large.

TABLE A-2. ERRORS IN FORECAST OF REALIZED CAPITAL GAINS
(Forecast minus actual, by calendar year, in billions of dollars)

	1989	1990	1991	1992	1993	Root Mean Squared Error
Forecasts Without Starts						
Equation (1)	37	57	50	37	23	42.2
Pure change	30	23	22	-1	-3	19.7
Flexible e.c.	42	39	30	-3	-4	29.0
Restricted e.c.	25	19	21	-1	-2	16.8
Forecasts With Starts						
Equation (1)	31	49	35	18	4	31.4
Pure change	29	18	6	-8	-8	16.1
Flexible e.c.	26	23	8	-7	-10	16.9
Restricted e.c.	21	16	8	-2	-5	12.5
Trend Method	16	37	20	-3	-6	20.3

SOURCE: Congressional Budget Office.

NOTE: e.c. = error correction terms included.

Accuracy in Predicting the Growth of Realized Gains

The six change equations predict the growth rate of realized gains rather than the level. Consequently, their ability to predict the annual growth of realizations is a clearer test than is their ability to predict the level of realizations. (For actual growth and the growth predicted by the equations, see Table A-3; the errors are shown in Table A-4.)

As mentioned in the text, the change equations literally predict the year-to-year change in the logarithm of realized gains. That difference approximates the percentage growth rate, and for simplicity the following discussion refers to the differences as if they were annual growth rates. Nonetheless, both the predictions and the actual are annual changes in the logarithms of gains, multiplied by 100 to approximate percentage changes. References to predicted growth by the level equations refer to the difference between the predicted logarithm of realizations in the prediction year and in the last year of estimation.

TABLE A-3. ACTUAL AND FORECAST GROWTH OF REALIZED CAPITAL GAINS
(By calendar year, in percent)

	1989	1990	1991	1992	1993
Actual	-5.3	-21.5	-10.4	12.7	13.5[a]
Forecasts Without Starts					
Equation (1)	9.1	0.6	-2.8	15.1	10.0
Pure change	12.6	-4.3	7.4	12.0	11.1
Flexible e.c.	18.9	6.0	13.1	10.4	10.7
Restricted e.c.	9.7	-7.5	6.9	11.8	12.1
Forecasts With Starts					
Equation (1)	8.6	-0.8	-8.1	11.1	7.3
Pure change	11.8	-8.1	-5.3	6.6	7.6
Flexible e.c.	10.4	-4.3	-3.7	7.0	6.6
Restricted e.c.	7.6	-9.5	-3.8	10.9	10.0
Trend Method	4.4	4.5	6.2	10.4	9.1

SOURCE: Congressional Budget Office.

NOTE: e.c. = error correction terms included.

a. Based on preliminary data. Final data indicate growth of 20.9 percent.

The predictions of growth rates show much the same pattern as the predictions of levels. The equations are better at predicting the upswing of realizations in the last two years than the downturn in the first three. All equations predict growth greatly in excess of what occurred in the first two years, and the change equations predict growth slightly below what actually occurred in the last two years. The predictions of the change equations that include housing starts improve in the third year but not those that exclude starts. When the predictions of the change equations without starts improve in the last two years, they become more accurate than the change equations with starts.[2]

2. The errors for 1993 would be larger if the final data on realizations for 1993 were used (see Table A-3). The amount of the increase cannot be judged precisely from Table A-1 because the forecasts of growth would also change if the final data were used to estimate the equations. The relative success of the equations would probably be unchanged from that discussed here.

TABLE A-4. ERRORS IN FORECAST OF GROWTH OF REALIZED CAPITAL GAINS
(Forecast minus actual, by calendar year, in percent)

	1989	1990	1991	1992	1993	Root Mean Squared Error
Forecasts Without Starts						
Equation (1)	14.4	22.1	7.6	2.4	-3.5	12.4
Pure change	17.9	17.2	17.8	-0.7	-2.4	13.7
Flexible e.c.	24.2	27.5	23.5	-2.3	-2.8	19.5
Restricted e.c.	15.0	14.1	17.3	-0.9	-1.4	12.0
Forecasts With Starts						
Equation (1)	13.9	20.7	2.3	-1.6	-6.2	11.6
Pure change	17.1	13.4	5.1	-6.1	-5.9	10.7
Flexible e.c.	15.7	17.3	6.6	-5.7	-6.9	11.6
Restricted e.c.	12.9	12.0	6.6	-1.8	-3.5	8.6
Trend Method	9.7	26.1	16.6	-2.3	-4.4	14.7

SOURCE: Congressional Budget Office.

NOTE: e.c. = error correction terms included.

In spite of the broad similarities with the predictions of levels, a key difference stands out. The two level equations (equation (1) with and without starts) predict growth rates about as well as the change equations, even though they predict the level of realizations much less well. In Table A-4, the level equations have lower errors than at least one of the three corresponding change equation in most years. The change equations are nearly all more accurate than the level equations only in 1990, and are uniformly less accurate in 1991. Over the five years, the level equation without starts has a lower root mean squared error than two of the three change equations without starts. The level equation with starts ties for the highest root mean squared error among equations with starts, and its error is close to that of another equation.

The similarity of the level and change equations at predicting growth indicates that the greater success of the change equations in predicting the level of realizations arises because they start from the actual level of realizations in the preceding year. None of the equations are particularly good at predicting the decline in realizations in the first three years, but the change equations predict levels of realizations better in the second through fourth years because they start from the low levels of realizations reached in the preceding year.

Looking at predictions of growth rates instead of levels of realizations also highlights the advantage of including housing starts, at least in these years. The highest root mean squared error of the equations with starts is below the lowest root mean squared error of the equations without starts. As mentioned above, including starts helps pick up the decline in the first three years but actually worsens the forecasts slightly during the recovery of the last two years.

Unlike adding starts, adding error correction terms does not generally improve the predictions. The equations with the flexible error correction terms have the highest root mean squared errors of each group of equations, whereas the terms with the restricted error correction terms have the lowest. The equation with both starts and the restricted error correction term has noticeably lower errors over the full five years than any other equation.

The failure of the equations with flexible error correction terms to outperform the pure change equations is surprising. The tests of the equation variables indicate that including the error correction terms is appropriate and the estimated coefficients of the error correction terms are statistically significant. The failure may result from including too many variables given the number of years for which data are available. When a large number of explanatory variables are used to explain growth in past realizations, they are more likely to be mistakenly credited with explaining some of the random fluctuation in past realizations. Those mistaken estimates then distort the equation's ability to predict future realizations.

Advantage of Forecasting with Equations

All forecasts will have errors. An important question for evaluating the forecasts of the equations is whether they have smaller errors than alternative methods. A plausible alternative to using the equations is to extend the current method for projecting realized gains in the last six years to project gains in the first year as well. Realized gains have tended to return to their historical average level relative to realized gains during the past 40 years. Using that regularity to project realized gains one year ahead might predict realizations better than the equations can.

As explained in the body of the memorandum, realizations are projected six years into the future by assuming that the ratio of realizations to gross domestic product in the year in which the projection is made will grow linearly to its historical average during the next six years. The ratio of realizations to GDP in the year of the projection is obtained by forecasting the growth in realizations with the equations from the previous year and then dividing that forecast by CBO's macroeconomic forecast for GDP in the same year.

The projection method is extended to give a projection for the year in which the projection is made by assuming that the actual ratio of realized gains to GDP in the previous year will return to the historical average over the seven years of the projection period. More specifically, to project realizations for the first year (the year in which the projection is made), the ratio of realizations to GDP in the previous year is assumed to change by one-seventh of the gap between that year's ratio and the historical ratio up to that year.

The method of projecting by ratios instead of forecasting with equations can be applied retrospectively to obtain projections one year ahead. To project realizations in 1989, for example, the ratio of realizations to GDP in 1988 is changed by one-seventh of the difference between this ratio and the average ratio of realizations to GDP between 1954 and 1988 (excluding 1986). The ratio in 1988 was 3.30 percent, and the average through 1988 was 2.73 percent. Consequently, the ratio for 1989 is projected to drop to 3.22 percent. GDP in 1989 was $5,250.8 billion, so realizations in 1989 are projected to be $169 billion.

CBO repeated the procedure in the four succeeding years to obtain projections one year ahead for each of the years from 1989 through 1993. Those projections are reported in the bottom line of Table A-1, and their errors are shown in the bottom line of Table A-2. The trend method has a smaller error than any equation in 1989, a larger error than most change equations in 1990, and errors similar to the change equations without housing starts in the last three years. Overall, as summarized by the root mean squared error, it does better than the level equations but not as well as most change equations, especially those with housing starts.

The trend method does better than the level equations because it projects from the actual ratio in the preceding year, rather than because it does a better job of predicting the rate of growth in realizations. That advantage is similar to that of the change equations noted above. The change equations also base their forecast for the year ahead on the actual level in the year before, whereas the level equations give no more weight to the last year than any other year.

When growth rates for the year ahead are compared, the trend method loses its advantage over the level equations. Predicted growth rates are shown in Table A-3 and prediction errors in Table A-4. The trend method predicts less variation in the growth rate from one year to the next than any of the equations, and the greater variation predicted by the equations tends to reflect growth of actual realizations. As a result, all but one of the equations have lower root mean squared errors in predicting annual growth than does the trend method. The trend method outperforms only the equation without housing starts and with the specification for flexible error correction.

Implications for Improving Equations

Overall, the equations did better than the trend method, which argues for their continued use. Furthermore, the errors of the equations can be examined for ways in which the equations might be improved.

The equations had difficulty predicting the declines that occurred in 1989, 1990, and 1991. They had difficulty in 1989 because only the decline in housing starts suggested a decline in realizations in that year. But housing starts had been declining since 1986 without causing a decline in realizations. Furthermore, the growth of stock values and GDP in 1989 suggested continued growth in realizations. Those latter two changes carried more weight in the equations. Of course, the decline in housing starts carried no independent weight in the equations without starts, but even the equations with starts all predicted a growth in realizations in 1989.

Most of the change equations were able to predict a decline in 1990, which was indicated by the recession and the decline in the stock market. The predictions, however, were much smaller than the actual decline. The largest predicted decline was just 9.5 percent whereas the actual decline was 21.5 percent. The level equations predicted little change from their expected levels in the previous year.

The decline in 1991 was predicted by the change equations with starts but not the others. The stock market rebounded in 1991, but GDP was flat and starts continued their decline. The change equations without starts were apparently mislead by the robust growth of the stock market. Even the change equations with starts did not capture the depth of the decline in that year. In fact, the level equations did better at predicting the decline than did their respective change equations.

The failure of the equations to predict even half of the decline in realizations that occurred in 1989, 1990, and 1991 suggests that they are missing a major determinant of realized gains. The primary alternative--that the three-year decline was a chance occurrence--is unlikely but cannot be ruled out.

The equations might be missing an accurate indicator of the declines in value that hit real estate in the late 1980s and early 1990s. Housing starts are an imperfect measure. One reason is that housing starts respond to population growth, and greater population growth does not necessarily lead to proportionately greater housing prices. Another reason is that until the 1980s housing starts were influenced by the flow of funds in and out of depository institutions. The effects of population growth and the availability of loans undoubtedly affect the historically estimated relationship between starts and realizations and thereby can lead to inaccurate predictions about the effects of changes in starts between 1989 and 1993. A third reason is that the low-income housing tax credit may be distorting the relationship between housing starts and real estate prices in recent years.

The equations could also be missing some of the decline in realizations brought about by the Tax Reform Act of 1986. As mentioned in the body of this memorandum, people who faced capital gains tax rates below the top rate in 1986 experienced larger tax rate increases than CBO picks up with the increase in the top statutory tax rate. In addition, the act placed new restrictions on tax shelters and reduced the differential between tax rates on capital gains and other income. Those changes reduced the incentive to take income in the form of capital gains. Thus, they could have led to greater reductions in capital gains realizations than the increase in the top tax rate indicated.

The decline in realizations in 1989 might have resulted from people waiting for a capital gains tax reduction in the next year. Legislation reducing the tax rate on capital gains narrowly failed to pass in 1989, and at the end of the year prospects were positive for passage in 1990. That explanation by itself is insufficient to explain the decline in subsequent years, however, because passage of such a tax cut became less likely in 1990 and 1991, though realizations did not rebound.

To obtain additional information about the cause of the decline in 1989, 1990, and 1991, CBO is exploring capital gain and loss transactions on a sample of individual returns. If the decline in total realizations between 1989 and 1991 resulted from declining real estate values, realizations on such property should have fallen whereas realizations on corporate stocks should have continued to rise. If the decline resulted from a reduced incentive to shift income to capital gains, realizations should have declined for all types of assets, but other types of income should have increased by more than normal. If the decline resulted from restrictions on tax shelters, realizations should have declined more on sales of partnerships than on corporate equities. Further information about the source of the forecasting errors can serve as a guide for improving the existing equations.

APPENDIX B: PREVIOUS FORECASTS

The Congressional Budget Office first forecast realizations of capital gains for its February 1987 baseline. Before 1987, CBO forecast revenues directly from the macroeconomic forecast of economic activity. The earlier forecasts of personal income tax revenue were made using an equation that described the historical relationship between certain macroeconomic variables and revenue collected from the personal income tax. Because of the large changes in the tax law taking effect in 1987, the historical relationship between the economic aggregates and revenues was no longer applicable. In response, CBO began simulating income tax revenues using a sample of tax returns filed in previous years. Components of income and deductions were projected for the sample of taxpayers, and then the current tax laws were applied to estimate taxes paid. That process requires a forecast of total income of each type--such as wages, dividends, and capital gains--that is distributed among the sample. Consequently, CBO began forecasting total realized gains.

The forecast for the baseline of February 1987 was based on a one-time procedure that dealt largely with the transitory effects of the Tax Reform Act on realizations between 1986 and 1988. The forecasts for the baselines of February 1988 through January 1991 were made using equations similar to equation (1) in this memorandum. The current method was first used for the baseline of 1992. The forecasts of each baseline are shown in Table B-1.

Forecasts of February 1988 Through January 1991

CBO developed equations similar to equation (1) as part of a study of the historical relationship between capital gains realizations and tax rates.[1] Their specifications and those of alternative equations were extensively tested using data for the 1954-1985 period. The equations were able to explain 98 percent or more of the level of realized gains during those years, and their errors appeared to be random.

Data on realizations in 1986 and 1987 were not available when the equations were being developed. Consequently, the indicator variable shown in equation (1) was not included. The earlier equations also employed an average tax rate on capital gains instead of the top statutory rate and included only net long-term gains in the dependent variable. They did not include housing starts, but some did include the change in the logarithm of real gross national product to account for fluctuations in realized gains over the business cycle.

---

1. Congressional Budget Office, *How Capital Gains Tax Rates Affect Revenues: The Historical Evidence* (March 1988).

TABLE B-1. ACTUAL AND PROJECTED CAPITAL GAINS REALIZATIONS (In billions of dollars)

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Actual	171	324	144	162	154	124	111	127	153	*	*	*	*	*	*	*
CBO Projections																
February 1987 Baseline	*	223	118	127	149	174	193	212	*	*	*	*	*	*	*	*
February 1988 Baseline	*	*	144	132	154	181	197	212	222	*	*	*	*	*	*	*
January 1989 Baseline	*	*	*	151	183	207	225	242	248	254	*	*	*	*	*	*
January 1990 Baseline	*	*	*	*	224	254	269	287	295	301	315	*	*	*	*	*
January 1991 Baseline	*	*	*	*	*	170	189	213	231	251	270	300	*	*	*	*
January 1992 Baseline	*	*	*	*	*	*	132	144	160	177	195	215	237	*	*	*
January 1993 Baseline	*	*	*	*	*	*	*	119	132	146	161	178	196	215	*	*
January 1994 Baseline	*	*	*	*	*	*	*	*	131	145	159	175	193	212	232	*
January 1995 Baseline	*	*	*	*	*	*	*	*	*	150	178	189	198	212	227	244

SOURCE: Congressional Budget Office.

NOTE: * = not applicable.

When CBO needed to forecast realized gains, the newly developed equations seemed appropriate. They could explain past realizations, and the explanatory variables could be taken from CBO's macroeconomic forecast.

A few changes were made to tailor the equations to the forecasting task. The dependent variable was expanded to include returns with net short-term gains because they are included in taxable income. In addition, experimentation with the statistical fit suggested the addition of a time trend, interest rates, and a fourth-order autoregressive term.

For the baselines of February 1988 through January 1991, the modified equation was used to forecast realizations for the full seven years covered in each forecast, using CBO's regular macroeconomic forecasts of gross national product, inflation, and interests rates. A special projection of stock market prices was added to the macroeconomic forecast to use in projecting realized gains.

In the first two baselines in which the modified version of equation (1) was used, those of early 1988 and 1989, CBO modified the equation's forecasts to reflect the shifting of sales into 1986 from subsequent years. The equation did not include the dummy variable for 1986 that is shown in equation (1) or any other measure of the transitory effects of the 1986 tax law change. It essentially projected realizations for 1987 and after as if the new tax rate had permanently been in place. To reflect the shifting of sales into 1986, CBO lowered the equation's predictions for 1987 and 1988.

Evidence on how well the equation was forecasting realized gains was not available for the 1988, 1989, and 1990 baselines. Data on realizations in 1986 and 1987 became available during this period, but they were distorted by the short-term response to the tax changes enacted in 1986.

The short-term distortions in 1986 and 1987 realizations also precluded reestimating the equation to reflect more recent experience. Thus, the forecasts for January 1988 through January 1990 were made with equations that did not reflect any realizations behavior after 1985.

The forecasts were made using the most recent economic data on equation inputs and the current CBO forecast for future years. For the January 1990 forecast, which covered the years from 1989 through 1995, CBO had actual data through early December 1989 and the economic forecast that it had just made. In 1989, the economy grew and the stock market rose strongly, which led the equation to predict a high level of realized gains in that year. Furthermore, the CBO macroeconomic forecast was calling for continued economic expansion, falling interest rates, and a rising stock market, all of which contributed to the equation's forecasting strong growth in realizations through 1995.

As actual realizations for several of those years have become available, that forecast has been found to have substantially overestimated capital gains. Instead of growing rapidly, actual realizations fell in 1989, 1990, and 1991, before rising in 1992 and 1993. As a result, the January 1990 forecast was $70 billion too high for 1989, and too high by larger amounts for subsequent years, as shown in Table B-1.

By late 1990, when CBO was making the January 1991 forecast, the recession had become apparent. CBO's new macroeconomic forecast predicted slower growth in the near term, as well as lower stock market prices. The failure to forecast this recession in the January 1990 baseline accounts for a large portion of the error in that baseline's forecast of capital gains realizations for the years from 1990 through 1995.

The failure to forecast the recession was not the only reason that the 1990 baseline overestimated realizations, however. That baseline's forecast for 1989 was too high, even though the levels of economic output and stock prices were largely known at the time.

What caused the 1989 overestimate was not clear in late 1990, nor is it much clearer now. By late 1990, CBO had final data on realizations in 1988 and preliminary data on realizations in 1989. The 1988 data showed that the January 1989 forecast had predicted $11 billion less in realizations than the $162 billion that actually occurred. Coming that close so soon after the upheaval of 1986 suggested that the equation was performing well. In contrast, the preliminary data for 1989 indicated that the January 1990 forecast was substantially overestimating realizations for 1989.

In late 1990, one hypothesis for the 1989 overestimate was that people had postponed the sale of assets in 1989 in anticipation of a capital gains tax reduction in 1990. In 1989, President Bush proposed a reduction in the tax rate on capital gains, and the Congress narrowly failed to pass one. Prospects for passage in the next year were good. The forecasting equation was not expected to pick up responses to anticipated tax changes and therefore could be accurately estimating what people would have realized had no change in tax rates been anticipated. Another hypothesis for the 1989 overestimate was that the equation, having been developed on data only through 1985, was structured to place too much emphasis on continued growth in realizations.

In late 1990, CBO responded to the second hypothesis by reestimating its forecasting equation through 1988, adding an indicator variable for 1986 and eliminating the trend and interest rate variables that appeared to overstate the growth in realizations. The modified equation and the less robust macroeconomic forecast led to a prediction that realizations in 1990, the year just ending, would be $170 billion, or $84 billion below the amount forecast for 1990 one year earlier. However, even with the adjusted equation and the worsened economic experience, the January 1991 forecast for 1990 turned out to be $46 billion above the actual level of $124 billion.

Furthermore, the January 1991 forecast had realizations growing to higher levels in 1991, whereas future data would show that they declined in that year as well.

Forecasts of January 1992 Through January 1995

When preliminary data on realized gains in 1990 became available in late 1991, CBO could see that the modifications of its forecasting equation a year earlier had failed to capture the decline in realized gains in 1990. CBO reestimated an equation with data through 1990 and found that it predicted less growth when the recession was anticipated, but it still could not explain the decline of realizations in 1989 or the depth of the decline in 1990.

Those failures suggested that the forecasting equation might be missing important information about the determinants of capital gains or erroneously measuring the historical relationship between realized gains and the determinants that CBO had identified.

CBO explored both possibilities in preparing the January 1992 forecast. The result of those explorations was the three-step procedure described in the body of this memorandum. CBO is continuing to explore the sources of the 1989-1991 decline in realizations as discussed at the end of Appendix A.

APPENDIX C: TESTS FOR UNIT ROOTS AND COINTEGRATION

Econometricians refer to variables that follow random walks as having unit roots. If variables that have unit roots are used to explain variation in another variable that also has a unit root, the unexplained residual variation may be stationary. When that occurs, the set of variables is termed cointegrated. If variables have unit roots and are cointegrated, then estimating equations in first differences with error correction terms is appropriate.

Econometricians have developed several tests to determine whether variables individually have unit roots and whether a set of variables is cointegrated. The Congressional Budget Office has applied some of those tests to variables in equation (1) and equation (3) to determine the appropriateness of estimating equation (2) with the addition of error correction terms from those equations. The tests for unit roots in six variables are presented first, followed by tests for cointegration of the variables in equation (1). CBO does not test whether the variables of equation (3) are cointegrated because they appear to lack unit roots individually.[1] (The equations appear in the body of the memorandum.)

Tests for Unit Roots

CBO applies the Dickey-Fuller tests for unit roots as described by James D. Hamilton.[2] The tests are implemented by estimating an equation of the form:

$$\text{(C-1)} \qquad \Delta X_t = \alpha + \beta * YEAR_t + \gamma * X_{t-1} + u_t$$

The equation hypothesizes that the change in the variable X in year t can be explained by the value of the variable last year; a deterministic, linear time trend (YEAR); and a constant amount, referred to as drift. The variable X has a unit root if $\gamma=0$, and the test is to determine whether the value of γ estimated from an ordinary least squares regression is statistically less than zero. Dickey and Fuller have determined critical values that the standard t-statistic must be more negative than to reject the hypothesis that $\gamma=0$.

If the variable to be tested has no steady trend up or down, the equation can be estimated without the YEAR variable. If the variable to be tested is clearly centered about zero, the equation can be estimated without either the trend or drift variables. Different critical values for testing γ apply when the trend or drift variables are omitted. If the importance of the trend or drift terms are unclear, the joint

1. All tests involving realized gains use the preliminary figure for 1993 of $145 billion instead of the final figure of $153 billion. The test results are unlikely to be changed significantly.

2. James D. Hamilton, *Time Series Analysis* (Princeton: Princeton University Press, 1994), p. 502.

hypothesis that γ and one or both of β and α equal zero can be tested. Dickey and Fuller have determined critical values for tests of those joint hypotheses.

CBO tests five variables from equation (1) and one variable from equation (3) for unit roots--log(GAINS), log(PRICE LEVEL), log(REAL STOCK VALUES), log(REAL GDP), log(STARTS), and log(RATIO). The first four of those have risen fairly steadily over the 1954-1993 period (see figures in body of memorandum). They are very likely to have either positive drift or a time trend, and therefore CBO tests them using the full version of equation (C-1). If the test for one of those variables cannot reject the hypothesis that it has a unit root (γ=0), CBO tests the combined hypothesis that the variable has a unit root and has no time trend (γ=β=0).

The last two variables--STARTS and RATIO--do not show steady trends up or down, and therefore CBO tests each one using equation (C-1) without the YEAR variable. If the test cannot reject the hypothesis of a unit root (γ=0), CBO tests the joint hypothesis that the variable follows a random walk without drift (γ=α=0).

TABLE C-1. TESTS FOR UNIT ROOTS USING EQUATION (C-1)

	Test γ=0		Test γ=β=0	
Variable	t-statistic	95 Percent Critical Values[a]	F-statistic	95 Percent Critical Values[b]
log(GAINS)	-2.77	-3.50 to -3.60	3.97	6.73 to 7.24
log(PRICE LEVEL)	-1.98	-3.50 to -3.60	3.88	6.73 to 7.24
log(REAL STOCK VALUES)	-1.54	-3.50 to -3.60	2.22	6.73 to 7.24
log(REAL GDP)	-1.43	-3.50 to -3.60	1.84	6.73 to 7.24

SOURCES: Congressional Budget Office; James D. Hamilton, *Time Series Analysis* (Princeton: Princeton University Press, 1994).

a. Hamilton, p. 763, Case 4.

b. Hamilton, p. 764, Case 4.

The test statistics for the first four variables, as shown in Table C-1, come from equations estimated over the 39 years from 1955 through 1993. The two critical values shown for both the t-statistic and the F-statistic are for sample sizes of 50 and 25, respectively. For each variable, the t-statistic is closer to zero than either critical value, and therefore the hypothesis that each variable has a unit root cannot be rejected. In addition, each F-statistic is below its critical value. Consequently, the joint hypothesis that the variable has a unit root and no linear trend cannot be rejected.

The test statistics for the STARTS and RATIO variables, as shown in Table C-2, are from equation (C-1) estimated without the time trend variable. The t-statistic for log(STARTS) is closer to zero than either of its critical values. Thus, the hypothesis that log(STARTS) has a unit root cannot be rejected. In addition, the F-statistic is below its critical value, indicating that the joint hypothesis that log(STARTS) has a unit root and no drift cannot be rejected. Based on those tests and the four tests above, CBO concludes that the five variables of equation (1) can be treated as if they contain unit roots.

In contrast, the t-statistic for log(RATIO) is farther below zero than its critical value, so the hypothesis of a unit root is rejected. Log(RATIO) does not appear to follow a random walk. Because the coefficient of log(RATIO) is significantly below zero, the test of the joint hypothesis that it and the drift term are zero is unnecessary. Furthermore, because log(RATIO) does not have a unit root, and because the other variables in equation (3) do not either, the residuals of this equation will not follow a random walk. Testing for cointegration of the variables in equation (3) is unnecessary.

TABLE C-2. TESTS FOR UNIT ROOTS USING EQUATION (C-1) WITHOUT TREND VARIABLE

	Test γ=0		Test γ=β=0	
Variable	t-statistic	95 Percent Critical Values[a]	F-statistic	95 Percent Critical Values[b]
log(STARTS)	-1.54	-2.93 to -3.00	3.31	4.86 to 5.18
log(RATIO)	-3.20	-2.93 to -3.00	n.a.	n.a.

SOURCES: Congressional Budget Office; James D. Hamilton, *Time Series Analysis* (Princeton: Princeton University Press, 1994).

NOTE: n.a. = not applicable.

a. Hamilton, p. 763, Case 2.

b. Hamilton, p. 764, Case 2.

Tests for Cointegration

Robert Engle and C.W.J. Granger describe several tests for determining whether a set of variables is cointegrated.[3] CBO presents two tests developed by Dickey and Fuller and one test based on the Durbin-Watson statistic. A single test cannot be relied on because the power of those tests varies among data sets.

All tests are of the hypothesis that the variables are not cointegrated. Thus, if the hypothesis is rejected, the variables can be treated as cointegrated. CBO reports the tests for the five variables of equation (1) found above to have unit roots.

The two Dickey-Fuller tests are based on equations of the form:

(C-2) $$\Delta u_t = \varphi * u_{t-1} + b_1 * \Delta u_{t-1} + b_2 * \Delta u_{t-2} \ldots + error_t$$

where u_t is the estimated residual in year t from equation (1), φ and b are parameters estimated by ordinary least squares, and $error_t$ is the residual of the test equation. The tests are to determine whether those residuals (u_t) have a unit root, in which case the variables of the equation generating the residuals are not cointegrated. The residuals have a unit root if the coefficient on the lagged level of the residuals is zero (φ=0), which is tested by comparing its estimated t-statistic to a critical value. That is similar to the above test for a unit root in a single variable when γ=0, although the critical values differ.

The two variants of the Dickey-Fuller tests presented here differ in the number of lagged values of Δu_t that are included. CBO uses lags of 4 and 1 because it could find approximate critical values for tests with those lags. Critical values for the test statistics depend both on the sample size and on the number of variables being tested. The critical values CBO found apply to five variables but assume a sample size of 50 compared with the 39 that it has. As a result, the outcomes of the tests need to be interpreted with caution. The critical values CBO uses come from Anindya Banerjee and others.[4]

The results of the two Dickey-Fuller tests are shown in Table C-3. Both test statistics are farther from zero than their critical values, indicating that the hypothesis of no cointegration is consistently rejected. The test statistics are not far from their critical values, however, so the evidence of cointegration is not overwhelming. All

---

3. Robert Engle and C.W.J. Granger, "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, vol. 55, no. 2 (March 1987), pp. 251-276.

4. Anindya Banerjee and others, *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data* (New York: Oxford University Press, 1993), p. 209.

TABLE C-3. TESTS FOR COINTEGRATION OF VARIABLES IN EQUATION (1)

Test	t-statistic	95 Percent Critical Value[a]
Augmented Dickey-Fuller 4 Lags	-4.59	-4.15
Augmented Dickey-Fuller 1 Lag	-5.04	-4.51

SOURCES: Congressional Budget Office; Anindya Banerjee and others, *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data* (New York: Oxford University Press, 1993).

a. Critical values for the tests come from Banerjee and others, p. 209. They are for sample sizes of 50, compared with sample sizes of 35 or 38 (depending on the number of lags in the test) used to generate the test statistics. The critical values assume five variables with unit roots. Equation (1) has five variables that have unit roots plus two that do not.

lags in the two tests have t-statistics above standard critical values, indicating that their inclusion is appropriate.

Banerjee and coauthors also report a critical value for the Durbin-Watson statistic to test whether the equation variables are cointegrated. That critical value is 1.19, which is less than the Durbin-Watson statistic of 1.39 in equation (1).[5] Thus, that test also rejects the hypothesis that the five variables are not cointegrated.

CBO concludes that variables in equation (1) are cointegrated and that estimation of equation (2) with error correction terms from equation (1) is appropriate. The variables in equation (3) do not contain unit roots, and therefore its residuals can also be used as error correction terms in equation (2).

5. Ibid, p. 209.