Working Paper Series
Congressional Budget Office
Washington, D.C.

Testing Alternative Methods for Forecasting Capital Gains

Larry Ozanne

Tax Analysis Division

Congressional Budget Office

Washington, D.C.

larry.ozanne@cbo.gov

March 2005
2005-03

Abstract

The Congressional Budget Office (CBO) forecasts capital gains as part of its start-of-the-year budget outlook. Jan Kim and Preston Miller in a recent paper with Larry Ozanne (KMO) developed two models that forecast gains for the year ahead more accurately than CBO's existing method. The models were developed with data as revised through 2002, and gains were forecast recursively for the years 1971-2000 (tax changes and their effects were not forecast). This paper tests one of their models further by using data actually available to CBO at the time of its annual forecasts and by extending the forecast comparison through 2002. It finds that using data available at the time of the original forecast does not reduce the advantage of the model, but extending the comparison through 2002 eliminates 40 percent of its advantage. This paper also finds a previously unnoticed specification problem in the model. Those findings lead to suggestions for alternative equations that are developed and tested. The new equations offer a mixture of advantages and disadvantage when compared to the KMO or existing CBO forecasting methods. Testing of the new equations also suggests that stock prices may belong in equations for forecasting gains, even though they cannot forecast accurately themselves. Stock prices appear to be needed to ensure that coefficients on other variables are estimated accurately.

Table of Contents

ABSTRACT i

CHAPTER I: INTRODUCTION 1

CHAPTER II: ALTERNATIVES FOR PREDICTING GAINS IN YEAR ENDING 4

- Testing the Variables 5
- CBO Equation 6
- Estimation 1952-2002 7
- Predicting Gains with November 2003 Data Set 8
- Predicting Gains with Data Available in December 10
- Conclusion 11

CHAPTER III: FURTHER EVALUATION OF THE TWO-STEP MODEL 13

- The Two-Step Model 13
- Mean Reversion Method 14
- Replicating KMO Results 15
- The Equation for Forecasting Gains in the Two-Step Model 17
- Forecasting under Actual Conditions 20
  - Adjustments 20
  - Findings 23
- Conclusion 24

CHAPTER IV: ALTERNATIVES FOR FORECASTING CAPITAL GAINS 26

- Strategy 26
  - Strategy for Adapting the Combination Equation 26
  - Strategy for Testing Forecasts 29
- Development of Equations 30
  - Unit Root Tests 30
  - Tests of Variables 1952-1990 31
- Re-Estimating the Equations in Subsequent Years 32
- Tests of Forecasting 34
  - Forecasts of the Gains Ratio 34
  - Forecasts of the Growth Rate of Gains 36
- Conclusion 38

Table of Contents continued

Choosing among the Alternatives Tested	38
Thinking about New Alternatives	39
TABLES for CHAPTER 2	43
TABLES for CHAPTER 3	50
TABLES for CHAPTER 4	53
FIGURES	61
APPENDIX A: Additional Results for Equations Predicting Gains	65
APPENDIX B: Explaining the Estimated Negative Effect of Real GDP on Capital Gains	70
APPENDIX C: Tables of Annual Forecasts Used in Figures	74

I. Introduction[1]

At the beginning of each year, the Congressional Budget Office (CBO) provides the Congress with a forecast of how the economy will perform during the coming decade and a projection of how that forecast will affect federal revenues and spending under current laws. As part of projecting revenues, CBO projects the amount of capital gains individuals and corporations will realize and report on their tax returns. This paper focuses solely on the gains reported by individuals.

CBO constructs its forecast of the economy and its projections of revenues and spending during the final months of the year before each report is released. At that time, the amount of gains realized during the year ending is unknown. As a consequence, CBO needs to determined gains for the year ending as well as the 10 years ahead. In this paper, I refer to the process of obtaining gains for the year ending as predicting, and the process for obtaining gains in the years ahead as forecasting.

CBO predicts the growth rate of gains for the year ending using information about how the economy, stock markets, and tax rates have changed from the previous year. CBO does not use the same relationships to forecast the path of gains in subsequent years because forecasts of stock prices are so unreliable. Instead, changes in gains for coming years are forecast by assuming that gains gradually return to their normal level relative to the size of the economy, as measured by GDP. That assumption and CBO's forecast of GDP are used to provide a specific future path for gains. The assumption is based on the historical pattern in which gains have frequently risen faster or slower than output from one year to the next, but have not risen or fallen relative to GDP over the long run (see Figure 1). In the rest of the paper, I refer to this method as mean reversion.

Jan Kim and Preston Miller developed two new models for forecasting capital gains that are described in the paper by Kim, Miller, and Larry Ozanne, *Estimating and Forecasting Capital Gains with Quarterly Models* (hereafter, referred to as KMO).[2]

One innovation of KMO's was to recreate a history of macroeconomic forecasts that could be used to test the ability of various macroeconomic variables to forecast gains. Previously, all attention had been placed on identifying which variables could be used to explain gains when actual historical values of the variables were known. Those previous approaches ignored the question of which variables could still explain gains when only forecasts of their values were known, as would be the case when gains themselves have to be forecast.

KMO developed Bayseian vector autoregressive (BVAR) models to forecast macroeconomic variables. One of their models used those macroeconomic forecasts in a separate equation to forecast gains (their two-step model). The other model entered the capital gains equation directly into a BVAR which forecast all variables simultaneously (their integrated quarterly model).

---

1. Helpful comments were received from Roberton Williams, Robert Dennis, and Arlene Holen of CBO and from participants in seminars of the Tax Analysis Division of CBO. John McMurray edited the manuscript and Denise Jordan-Williams formatted it.

2. CBO Technical Paper 2004-14, September 2004, www.cbo.gov/tech.cfm.

KMO found that their models were able to forecast a measure of gains more accurately than CBO's mean reversion method or other simple methods between 1971 and 2000. Their measure of gains used exogenous information to estimate and remove the effect of tax changes on gains. KMO's focus on "tax-adjusted gains" did not seem problematic for CBO's use of their models, because CBO forecasts gains under current law, which historically has meant under fixed tax rates.

A major unanswered question about KMO's models has been how their accuracy would change when the models were used with CBO's macroeconomic forecasts. CBO's official forecasts of gains must be based on CBO's forecasts of the economy, and given the very different methods CBO uses to make its forecasts, the two sets of forecasts are likely to differ.

An initial test of the sensitivity of KMO's models to different forecasts was conducted for the winter 2003 and winter 2004 baselines. Jan Kim forecast gains with the two-step and integrated quarterly models using macroeconomic variables from both the BVAR models and from CBO's forecasts. In each year, the integrated quarterly model changed its forecasts of gains substantially when it switched macroeconomic forecasts. Because of this sensitivity, I concluded that more basic development of that BVAR model would be needed to make it work well with CBO's forecasts. So far, the work of BVAR modeling has deterred me from pursuing that approach further. The two-step model changed its forecasts much less when the BVAR and CBO forecasts were substituted, which has led me to consider further its usefulness for forecasting.

The basic motivation for this paper has been to test further the two-step forecasting approach under actual CBO forecasting conditions. One question is by how much does the substitution of CBO's macroeconomic forecasts affect the forecasting advantage of the two-step model over the mean reversion method. Another is how has relative forecasting accuracy been affected by additional experience. KMO completed their model development with data through 2000; since then, the stock market has fallen substantially, a recession occurred, and gains dropped dramatically. KMO did report forecasts of their models for 2001, but did not integrate those forecasts with their measures of forecasting accuracy in earlier years. In addition, data for 2002 have since become available. Beyond those questions, I wanted to explore the two-step method in greater detail than KMO reported in preparation for possible use by CBO.

Additional motivation for this paper comes from KMO's success at predicting tax-adjusted gains for the year ending. As part of their model development, KMO experimented with alternative variables and found a specification which predicted tax-adjusted gains more accurately than one of the equations CBO had been using. That finding encouraged me to determine whether the equation KMO developed could be used to improve CBO's prediction of actual gains.

A key part of my method is to apply KMO's models under actual conditions faced by CBO. Annual winter baselines containing the range of variables used by KMO are readily available starting with the winter 1992 baseline finalized in December 1991. Annual winter baselines are available for each subsequent year, and I use those through the winter 2003 baseline. Each baseline has with it the historical values of its macroeconomic variables, as they were known at the time. Because of this historical information, I can essentially apply alternative approaches as if they were being applied at the time of each baseline. Because actual information on gains is available through year 2002, I can also determine the relative accuracy of the alternatives had they been applied over the last dozen years. More precisely, with these baselines, I can test twelve predictions for 1991-2002 and eleven year-ahead forecasts for 1992-2002.

The tests indicate that one of the variables KMO introduced can be used to improve slightly CBO's prediction of gains for the year ending. The tests also show that the two-step method's forecasting advantage over mean reversion is sensitive to the years being forecast. Adding a forecasts of 2001 cuts the two-step method's advantage in accuracy by 40 percent, and adding 2002 reduces that advantage slightly further. Investigation of the forecasting equation used in the second step also reveals a specification problem: forecasts of the growth rate of real GDP are estimated to have a negative impact on the growth rate of gains.

The experience gained from testing the two-step and mean reversion forecasting methods leads me to consider an alternative approach to develop forecasting equations. The equations I develop and test fail to forecast more accurately than mean reversion and the two-step methods, but they have other advantages. The equations are not without their disadvantages, however. In particular, their estimation with historical data changes significantly as additional years of data are added. That disadvantage leads me to consider forecasting with equations that, like CBO's equations for predicting gains, incorporate stock market variables.

The work is presented in the three following chapters. Chapter 2 evaluates the equation developed by KMO for predicting gains in the year ending. Chapter 3 explores the two-step forecasting model as an alternative to mean reversion for forecasting one year ahead. Chapter 4 develops new equations for forecasting gains, tests them, and draws conclusions about forecasting gains.

II. Alternatives for Predicting Gains in Year Ending

Each December, CBO predicts the amount of capital gains realized during the year ending with information about stock markets, economic output, and tax rates during the year. A regression equation is used that relates the growth of gains to growth in the explanatory variables using data from the early 1950s through the preceding year. The level of gains for the year ending is obtained by applying the growth rate predicted by the equation to the amount of gains realized in the previous year.

KMO developed alternative equations for predicting growth in tax-adjusted gains. Their equation that most directly parallels CBO's equations is the one developed for their two-step forecasting approach.

KMO's objective in developing the equation was to get the most accurate estimate of the growth rate of tax-adjusted gains in year t, given knowledge of the explanatory variables in that year. To meet their objective, they selected different combinations of explanatory variables from a pre-determined set and used each combination to predict the growth of tax-adjusted gains in years from 1971 to 2000. The predictions were made by recursively estimating an equation from 1949 through year t-1, predicting growth in year t, then re-estimating the equation through year t and predicting growth in year t+1, and so on. All estimation was done on the same data set with values of each variable from 1949 thorough 2000. After trying all possible combinations of their candidate variables, they identified the equation that produced the lowest root mean square error (RMSE) predicting the growth of tax-adjusted gains between 1971 and 2000.

KMO's method for selecting an equation parallels CBO's method for predicting gains in the year ending. Thus, it seems like a good method for identifying variables that would be useful to CBO. Furthermore, the equation that KMO identified as best over the full period predicts the growth rate of tax-adjusted gains more accurately than does a simplified version of CBO's current forecasting equation when it is fit to KMO's data. Over the 1971-2000 period, CBO's equation has a RMSE of 14.13 percentage points while KMO's equation has a RMSE of 10.94 percentage points.[3]

KMO developed other equations that predict the growth rate of tax-adjusted gains for their integrated quarterly models. Those equations use subsets of the variables in KMO's best two-step equation. Thus, the best two-step equation represents the full set of possibilities KMO identified.

In spite of the promise of KMO's equation, its usefulness for CBO's year-end prediction remains uncertain. The main source of uncertainty is that KMO's equation does not predict the same capital gains series that CBO does. Because KMO's focus was on forecasting gains in future years under current laws, they removed an estimate of the impact that tax rate changes have had on the historical

---

3. *Estmating and Forecasting.....* , Table 3. The RMSEs are computed from growth rates measured in percentages. An error of 14.13 percentage points would arise if gains had been forecast to grow 10.00 percent in a year but actually grew 24.13 percent (24.13-10.00=14.13).

gains data.[4] The result was a variable of tax-adjusted gains that tried to reflect what gains would have been between 1948 and 2000 if tax rates on gains had remained at their levels during 1998-2000. The equation developed to predict growth in tax-adjusted gains could be less proficient at predicting growth in actual gains because of limitations in the estimation of the tax effects or interactions between tax rate variables and other variables.

Another source of uncertainty about the KMO variables is that KMO selected them based on the use of their known values for the prediction year. CBO estimates gains in December for the year ending, and at that time it must rely on partial information supplemented with forecasts for the full year. A third source of uncertainty is the time-period over which KMO selected their variables. They started with 1949 and estimated through 2000, CBO starts from 1952 to avoid post-war adjustments, and now uses data through 2002.

KMO's equation explains the year-to-year growth rate of tax-adjusted gains as measured by the change in the natural logarithm of those gains. To be comparable, the explanatory variables that grow over time are also measured as the change in logarithms, and those that do not are measured as simple year-to-year changes. Table 2-1 lists the variables and their abbreviations. In the abbreviations, D indicates annual change and L indicates logarithm. A constant term was also included.

Among the variables, the growth rates of productivity, real GDP, and real investment should reflect growth in profitability of businesses, and therefore should positively impact business asset values and gains. The spread in interest rates between corporate bonds and T-bills tends to widen in recessions, shrink in recoveries, and possibly become negative in booms, which should make it negatively related to the growth of gains. The growth of real compensation per hour might positively impact gains as an indicator of economic growth, or negatively impact gains as an indicator of rising business costs.

Testing the Variables

I test the usefulness of the variables for predicting gains in three ways. The first is to estimate KMO's equation with the growth rate of actual gains as the dependent variable. The growth rate is measured as the change in logarithms. Because changes in tax rates change the actual amount of gains, I add CBO's two tax rate variables to KMO's five explanatory variables. The two tax rate variables, identified in Table 2-2, are the ones used to construct tax-adjusted gains. The re-estimated KMO equation will be compared to the current CBO estimating equation. Both equations are estimated over the 1952-2002 period using the data available in CBO's winter 2004 baseline, which I refer to as the November 2003 data set.

The second way I test the usefulness of the variables for predicting gain is to repeat the process

4. KMO estimated the tax effects using the version of the CBO equation whose RMSE is reported above.

through which KMO selected their equation: recursive estimation of regression coefficients and prediction of gains from 1971-2002. This way indicates how well the KMO and current CBO equation would have performed CBO's task over a wide variety of conditions, provided the conditions were fully known at the time. By fully known at the time, I mean that the November 2003 data set is used. It would be preferable to use the data sets as they were known at each prediction time, but those data are not available for so long a period.

The third way to test the variables is to repeat the annual predictions of the second step using the data that CBO had at the time it made its predictions (referred to as real-time data). In that data, values of explanatory variables for the year in which gains are to be predicted are partially forecast by CBO. For example, at the time CBO is making its forecast, the Bureau of Economic Analysis has usually released values of GDP through the third quarter of the year ending; CBO must forecast the fourth. Another difference in real-time data from the November 2003 data set is that some explanatory variables in the year ending and immediately preceding years will be revised. GDP, for example, is revised for three subsequent years as more information about a year becomes available. As noted in the introduction, these historical files are available for CBO's baselines constructed in the end of 1991 and later.

CBO Equation

CBO currently predicts gains for the year ending with an equation that combines variables from four equations developed by Miller and Ozanne (MO).[5] The four equations employ slightly different mixes of six explanatory variables. In tests with real-time data from 1990 to 1998, MO found that the average prediction of the four equations was more accurate than the prediction of any single equation. Based on their work, CBO began applying MO's four equations and averaging the predictions. Further thought suggested that the procedure should be improved by combining the six explanatory variables into a singe equation and allowing the historical data to determine how best to average their influences. Based on that reasoning and a few trials, CBO recently began predicting gains with what I refer to as CBO's combination equation.

CBO's combination equation uses a different dependent variable than I use with KMO's variables. Instead of the growth rate of gains, it uses the growth rate of the ratio of gains to potential output in the economy. Potential output, as computed by CBO, removes effects of the business cycle from actual GDP. MO found that scaling gains by potential output removes the major trend present in the level of gains, and facilitates the separation of business cycle effects from those of general growth in the economy. The variables in CBO's combination equation are listed in Table 2-2 (see MO for further discussion of how the dependent and explanatory variables were selected). A constant term is included but is not listed.

5. *Forecasting Capital Gains Realizations*, CBO Technical Paper 2000-5, August 2000, www.cbo.gov/tech.cfm.

Estimation 1952-2002

KMO's equation explains the growth rate of gains less well than CBO's combination equation when both are estimated over the years 1952-2002 with the November 2003 data set. The growth rate referred to here is not the change in the logarithm that is the dependent variable for the KMO re-estimate, but the percentage growth rate measured as the ratio of gains in a year to gains in the preceding year, less one, and then multiplied by 100. Because the two equations have different dependent variables, the standard R-squared statistics of the equations are not comparable. Instead, I follow KMO's practice of transforming the predicted dependent variable of each equation into a prediction of the percentage growth rate. The RMSE of KMO's re-estimated equation is 16.69 percentage points while that of CBO's is 11.10 percentage points.

The lower explanatory power of the KMO equation is reflected in the estimated coefficients of its explanatory variables (see Table 2-3). Aside from the added tax rate variables, only one of KMO's five variables has a coefficient that differs from zero at normal levels of statistical significance. That variable is the growth rate of real fixed investment. The growth rate of productivity has the expected positive coefficient and it is almost as large as its standard error. The three remaining variables are not promising. The coefficients of the spread and real compensation per hour are half or less of their standard errors and therefore are indistinguishable zero, statistically. The coefficient of the growth rate of real GDP has a negative sign when a positive sign is both expected and common in many other equations. In KMO's equation, it must be redundant with other variables picking up the growth of the economy.[6] In any event, the CBO combination includes a real GDP effect in its trend and cyclical terms. KMO's re-estimated equation, in addition to having less explanatory power than CBO's, has more serial correlation of its errors and those errors are less likely to be normally distributed.

Although KMO's equation as a whole is not promising for CBO's estimation of gains, the investment and productivity variables are. In addition to their statistical promise, both have plausible rationales for affecting the amount of gains realized.

On the strength of their theoretical and empirical support, I add both investment and productivity to CBO's combination equation. In keeping with the form of the dependent variable, I include investment as the growth rate of the ratio of nominal fixed investment to potential output (abbreviated as DLIFIXFE). In the re-estimated combination equation, investment has the expected positive coefficient and is statistically significant. Productivity is insignificant. In addition, both the GDP gap and the housing starts variables become statistically insignificant.[7] Apparently, investment is a better measure of business cycle and related influences on gains than is the GDP gap. Furthermore, because investment includes real estate investment, its supplanting of multifamily housing starts is plausible. A drawback of investment would seem to be its inclusion of investment

6. Real GDP growth is estimated to have a negative sign in one of KMO's forecasting equations as well. The next chapter investigates the origins of that negative sign.

7. The GDP gap is the ratio of GDP to potential GDP and is a measure of the extent to which the economy is in recession or boom.

in owner-occupied housing. Changes in investment in owner-occupied housing might well be related to the realization of gains on owner-occupied housing, but those gains are largely exempt from capital gains taxes.

I drop the three statistically insignificant variables and re-estimate the equation. This new combination explains more of the variation in the ratio of gains to potential GDP than CBO's combination equation. Its R-squared statistic is .845 compared to CBO's .822. When used to predict the percentage growth rate of gains, the new combination has a RMSE of 10.71, slightly better than CBO's combination. One disadvantage of the new combination is its slightly higher correlation of residuals, as measured by the Durbin-Watson statistic. Because of its superior fit, I carry the new combination equation along to the second test.[8] The regression statistics for CBO's combination equation appear in Table 2-4 and for the new combination equation in Table 2-5. The RMSEs for all three equations are as follows:

Root Mean Square Errors in Fitting the Growth Rate of Gains 1952-2000 (in percentage points)	
KMO Equation	16.69
CBO Combination	11.10
New Combination	10.71

Predicting Gains with November 2003 Data Set

The second step in testing the KMO variables is to predict the growth rate of gains for the year ending, as CBO does for its baseline revenue forecasts. The predictions are made for the years 1971 through 2002, with equations estimated through the prior year and then values of the explanatory variables used to predict gains for the year ending. All variables are as known in November 2003, so they reflect information that became available after CBO predicted gains for each year.

KMO's equation could predict better than CBO's combination over most of the 1971-2002 period even though it did not fit the 1952 to 2002 period as well. KMO selected their equation because it predicted tax-adjusted gains accurately over much of this period, and outperformed a simpler CBO equation fit to tax-adjusted gains. The new combination equation may predict gains more accurately near the end of the period than CBO's combination because it fits the full sample better. It could do less well in earlier decades, however.

As in the test of fit, I transform the prediction of each equation to a prediction of the percentage growth rate of gains. I then compute the error of each prediction and summarize the errors in the

8. Modifications of the investment variable could be considered in the future. For example, it could be expanded to include inventories so as to better reflect business cycles, or it could be narrowed to exclude investment in single-family homes. I did substitute non-residential fixed investment, but that worsened the fit.

RMSE for each equation.

I find that KMO's equation has the highest RMSE over the full 1971-2002 period, and that CBO's combination has the next highest. The RMSEs in percentage points are 26.95 for KMO's, 18.64 for CBO's, and 16.73 for the new combination (see Table 2-6). The RMSE for KMO's equation is considerably larger than the 10.94 percentage points they found when predicting tax-adjusted gains. The RMSE of CBO's combination is slightly larger than KMO report for the simpler CBO equation. Apparently, actual gains are harder to predict than tax-adjusted gains.[9]

None of the equations dominate in all sub-periods. Between 1971 and 1980, KMO's equation has a substantially lower RMSE than the other two. Between 1981 and 1990, and again between 1991 and 2002, the new combination equation has the lowest RMSE, followed closely by CBO's combination and distantly by KMO's equation. The new combination does dominate CBO's combination, however, having a lower RMSE in all subperiods.

The RMSE criterion gives extra weight to large errors. It is possible that CBO's combination predicts more accurately than the new combination most of the time, but when it makes a bigger error, it makes a much bigger error. I checked that possibility by tabulating the percentage of times the new combination is more accurate than the CBO combination. The new combination has the lower error more frequently. Between 1971 and 2002, it is more accurate 66% of the time. The accuracy percentage varies by time period, though. The new combination is closer 70% of the time in the first two decades and less than 60% of the time in the last 12 years.

Overall, these results keep the same relative ranking of the three alternatives as did the test of fit over the full period. The new combination is best, CBO's combination is a close second, and KMO's equation is a more distant third. The primary exception here is that KMO's is best in the decade from 1971 to 1980.

None of the equations show similar errors in all three subperiods. Thus, a fairly wide range of possible errors must be expected on future predictions. The two combination equations show declining errors in more recent subperiods, which could indicate improving estimates of stable coefficients on their explanatory variables. Unfortunately, it is impossible to distinguish this possibility from that of unusually small random or omitted influences in these equations since 1981.

A caveat to the findings on estimation and prediction with the November 2003 data is that some of its variables were revised after this analysis was completed in early 2004. A notable revision is an increase in the amount of gains in 2002 from $254 billion to $268 billion. While the impact of this change is unknown, it is likely to be small. As a tentative test, I omitted errors in predicting 2002 gains from the RMSE calculations reported in Table 2-6. The RMSEs overall and in the last period

---

9. Tax-adjusted gains probably are easier to predict than actual gains between 1971 and 2000 because the tax adjustments were made using an equation estimated with data from 1952 through 2000. Those years of experience provided better estimates of the equation's coefficients than can be obtained by equations estimated through fewer years, say through 1970. So some of the uncertainty faced by equations estimated with data only through 1970 when predicting actual gains has been removed from tax-adjusted gains.

changed little.

Predicting A Year's Gains with Data Available in December

CBO typically finalizes its prediction of gains for the year ending in early December. At that time, it has preliminary information on explanatory variables for part of the year and must predict their values for the full year before predicting gains. Variables that can predict gains when they are fully revised are likely to be less successful when they are only partially known. Some of these variables will be known more reliably than others as a year ends, and such differences could change the ranking of how well the equations predict gains. In particular, the new combination equation could lose its advantage over the CBO combination equation if by early December the year's investment is known less well than are the year's GDP and housing starts.[10] The availability of CBO's forecast data bases from December 1991 on allows us to test that possibility for a dozen years.

I begin by computing errors in the year-end predictions of the growth rate of multifamily housing starts (DLSTARTS), the growth rate of GDP relative to potential (DLGAP), and the growth rate of fixed investment relative to potential (DLIFIXFE); see Table 2-7. The predicted growth rate for a year is computed from the CBO macroeconomic forecast typically completed in December of the year. The actual growth rate is assumed to be that in data as revised through November 2003. The error is the difference between the November 2003 value and the year-end prediction.[11] The errors for each variable during the 1991-2002 period are summarized by the RMSE.

I find that the growth rate of the GDP gap has the lowest RMSE. The growth rate of fixed investment relative to potential has nearly three times as large a RMSE. The growth rate of multifamily housing starts is the hardest to predict, with a RMSE nearly twice as large as that of fixed investment (see Table 2-7). Investment is harder to predict than the rest of GDP most likely because it is more volatile from year to year. In addition, the information available on investment by early December is less complete than for that for other major components of GDP. Housing starts are even harder to predict than investment because they are even more volatile. The standard deviation of their annual growth rate is almost four times as large as that of the growth rate of fixed investment relative to potential GDP. The error in forecasting starts is twice instead of four times as large, because starts for a year are more completely known by December than is investment.

The explanatory variables with the largest and smallest prediction error turn out to be in CBO's combination equation. As a result, switching to real-time data could hurt that equation's predictions

10. The ranking of investment, GDP and housing starts could also be influenced by the need to forecast December values of stock prices and the dollar volume of trades. Errors in those forecasts might be better compensated for by one variable than another, a role that would not be apparent when predicting gains with data revised through November 2003.

11. Two characteristics of the data affect the errors in these three variables. One is that the baseline data in 3 years were completed in January and another was completed in February. See Table 2-7 for dates. Later baselines can use more complete information. The baseline for winter 1996 was not finalized until May of 1996, but an earlier version from February was available and used for this analysis. The second characteristic is that the some variables in the November 2003 data have been revised since then. As a result, not all errors are accurately captured here.

of gains as much or more than the predictions of the new combination equation. Furthermore, being able to predict GDP more accurately than fixed investment does not necessarily increase GDP's accuracy in forecasting gains relative to that of fixed investment. If the more predictable parts of GDP are not good indicators of gains, then GDP's greater predictability will not help it relative to fixed investment.

Errors in predicting the values of explanatory variables in the year ending are not the only differences between predictions made with real-time and November 2003 data. GDP, fixed investment, and potential GDP are revised in the three years after their initial release. Also, data on gains for the year before the year ending get revised in the year ahead. These revisions mean that the regression coefficients estimated from real-time data will differ from those estimated with November 2003 data. The different regression coefficients can add to differences in predictions of gains made with real-time and revised data.

The above considerations leave unanswered the question of whether the new combination will lose its advantage over CBO's when I switch to real-time data. The only way to answer that question is to predict gains with both equations and compare their errors using real-time data. I do not test KMO's full equation because it does so much worse than either combination equation from 1991 to 2002 with the data as of November 2003.

When I estimate both combination equations on the real-time data, the new combination equation continues to have a lower RMSE than CBO's combination equation, although the advantage is reduced. The RMSE of the new combination equation is 10.63 percentage points and the RMSE of CBO's combination equation is 12.17 percentage points (see Table 2-8). The difference of 1.54 percentage points is less than the difference of 1.91 percentage points using the November 2003 data. The smaller difference between the two equations suggests that in real-time data, CBO's equation is helped more by the relative ease of forecasting the GDP gap than it is hurt by the relative difficulty of forecasting starts. As expected, both equations have larger RMSEs with real-time data than with the November 2003 data.

The new combination equation also has a lower RMSE when 2002 is omitted. And it has a lower average absolute error with and without 2002 (see Appendix A, Table A-3). However, CBO's combination takes the lead in the number of years with more accurate forecasts. Using the November 2003 data, CBO's combination is more accurate in 5 of the 12 years from 1991 through 2002; using real-time data, the CBO combination is more accurate 7 of 12 times (see Table 2-8).

Conclusion

Of the variables that KMO found helpful in predicting gains, only fixed investment shows potential for improving CBO's prediction of gains for the year ending. The combination equation with fixed investment fits and predicts better than CBO's combination equation using the November 2003 data. The new combination equation also has a lower RMSE in predicting the growth of gains between 1991 and 2002 using real-time data, but that advantage is smaller than with the revised data. With real-time data, the new equation predicts gains more accurately in only 5 of 12 years, but three of those years were the most difficult to predict and the new combination does enough better in them to outweigh its larger errors in the other 7 years, as far as the RMSE is concerned. Of course, the new combination will not continue to have a lower RMSE than CBO's combination or even KMO's equation if the determinants of gains change from those in the last half century.

Appendix A documents three related conclusions about 1991-2002. Errors in forecasting stock prices in December of the year ending make the annual growth rate of the S&P 500 about as accurate a predictor of growth in gains as the fourth-quarter growth rate used in CBO's combination equation. Second, CBO's combination equation predicts gains more accurately than any of the four simpler equations MO developed, or than the average prediction of those four equations. Finally, the CBO and new combination equations could have predicted gains more accurately from December 1991 through 2002 than CBO did.

III. Further Evaluation of the Two-Step Model

Jan Kim and Preston Miller developed the two-step model for forecasting the growth rate of gains as described by KMO. KMO showed that under certain conditions, the model forecasts one year ahead more accurately than simpler methods, including the mean reversion method CBO uses, between 1971 and 2000.

I address two additional questions about the model. First, will the model's advantage over mean reversion remain in future years? This is a perpetual question; at the time this work is being done in 2004, only two additional years of information are available. Thus, I can offer no definitive answer. However, those two years indicate that the size of the two step model's advantage over mean reversion is unstable, as it has declined by 40 percent.

Second, will the two-step model's advantage over mean reversion remain when they are applied under CBO's actual forecasting conditions? I have recreated CBO's forecasting conditions for the annual winter baselines of 1992 through 2002, and applied both methods in each year. I find little deterioration in the two-step model's advantage under these real-time conditions.

In addition, I explore in greater detail than KMO did the second step equation they developed for forecasting gains. I find that their basic design has merit, but the estimated coefficients lack statistical significance and point to a shortcoming in the choice of explanatory variables.

The Two-Step Model

KMO designed the two step model to work in CBO's forecasting environment, but they had to develop it under somewhat simplified conditions. The forecast is assumed to be made as one year ends and the next begins. At that time, values of macroeconomic variables are assumed to be known through the year ending, but gains are known only through the previous year. Gains for the year ending are predicted with the KMO equation presented in Chapter 2. Tax changes are ignored; the gains being predicted and forecast are the tax-adjusted gains already discussed.

The first step of the forecasting model is a slimmed-down version of CBO's economic forecast. A BVAR model with a small set of macroeconomic variables is estimated with data through the year ending. That model is then used to forecast growth of the macroeconomic variables one year ahead.

The second step is to forecast growth of gains one year ahead using an equation that relates the growth rate of gains to the growth rates of the macroeconomic variables as forecast in the first step. One of KMO's innovations is to estimate the gains equation with forecasts rather than actual values of macroeconomic variables. To do so, KMO need forecasts of the macroeconomic variables for all years of estimation, through the year ending. They obtain these forecasts by using the BVAR model to forecast one year ahead for each of the years used to estimate the gains equation. The dependent variable in the gains equation is the growth rate of gains, and for the year ending a predicted value is used instead of the actual value.

More specifically, KMO began with a small set of macroeconomic variables they thought could be forecast reasonably well and could explain gains. KMO had quarterly data on the variables from 1948 through 2000, as known in early 2002. To estimate the second step equation, they needed year-ahead forecasts of the macroeconomic variable for an historical period. They obtained those forecasts by estimating the BVAR model from 1949 through 1959, and then forecasting each variable for 1960. Then they estimated the model through 1960 and forecast each variable for 1961. They continued the recursive forecasting until they had the year-ahead forecast of each variable in the set for 1960-1970. For the first step of the annual cycle to forecast tax-adjusted gains in 1971, they repeated the above estimating and forecasting one more time to get the year-ahead forecast of each macroeconomic variable for 1971.

In the second step, KMO regressed the actual growth rate of gains on the forecasts of the growth in the macroeconomic variables from 1960 through 1970. The actual value of gains for 1970 would be unknown as of the end of the end of 1970, so KMO substituted their predicted value in the data for the regression. Finally, they used the estimated regression and the forecasts of growth in the macroeconomic variables for 1971 to forecast the growth of gains in 1971.

KMO repeated the first and second steps to forecast gains in 1972, 1973, and so on through 2000. That procedure gave them 30 forecasts whose accuracy could be judged against the historical growth rate of gains. At the time they completed their work in 2002, preliminary information about the growth rate of gains in 2000 was available to judge the accuracy of forecasts in that year, but no reports on gains were available for later years.

Finally, KMO repeated the entire process using different sets of macroeconomic variables. They selected their best model as the combination of macroeconomic variables providing the lowest RMSE in forecasting the growth rate of tax-adjusted gains. Their best model used the forecasts of four macroeconomic variables:

growth rate of productivity
growth rate of real GDP
growth rate of real fixed investment
change in the ratio of real durable consumption to real disposable income

Mean Reversion Method

KMO also forecast one year ahead with a few simplified methods to observe the benefit of their two-step and integrated models. One of those simpler methods was the mean reversion method that CBO has been using. I replicate those mean-reversion forecasts here, to match the relative advantage of the two-step model that KMO reported.

The mean reversion method forecasts gains by assuming that gains gradually return to their normal size relative to the size of the economy, as measured by GDP. That assumption and forecasts of GDP are used to provide a specific future path for gains. The assumption that gains revert to a normal size relative to GDP is based on the historical pattern in which gains have frequently risen faster or slower than output from one year to the next, but have not risen or fallen relative to GDP over the long run (see Figure 1).

KMO and I use mean reversion to forecast tax-adjusted gains in the year ahead as follows. The ratio of tax-adjusted gains to GDP next year, called R(t+1) is determined by:

$$R(t+1) = R(t) + v*[\ A(t-1) - R(t)\ ]$$

where R(t) is the ratio of tax-adjusted gains to GDP in the year ending, A(t-1) is the average ratio of tax-adjusted gains to GDP over the years 1948 through year t-1, and v is the reversion rate.

Two of the terms on the right hand side of the equation need to be estimated. One is the ratio of tax-adjusted gains to GDP for the year ending (R(t)). KMO and I estimate that by predicting tax-adjusted gains for the year ending and then dividing it by the value of GDP in the data set. KMO and I predict tax-adjusted gains for the year ending with a version of the process described in the preceding chapter. The equation we use includes only the S&P500 and GAP variables from the CBO combination equation. The tax terms are not needed because tax-adjusted gains have estimates of the tax effects removed. The equation is estimated from 1949 through year t-1.

The second term to be estimated is the reversion rate.[12] KMO and I do that by estimating the amount of mean reversion that has occurred in the past. More specifically, we estimate:

$$R(s) - R(s-1) = v*[\ A(t-1) - R(s-1)\ ] + u(s)$$

where u is changes unrelated to reversion and s refers to years from 1949 through t-1.

When estimated with data ending in years 1969 through 1996, the estimated reversion rate fluctuates between about 26 percent and 30 percent. As additional years through 2000 are added, the estimated reversion rate falls steadily to 6 percent. That decline is caused by the steady rise of gains relative to GDP during the stock market boom. When data for 2001 and 2002 are included, the reversion rate jumps back to around 25 percent, reflecting the abrupt drop that occurred in the ratio of gains to GDP. The reversion rate used to forecast gains in year t+1 is estimated with data through year t-1. For example, the reversion rate used to forecast gains in 2000 is the rate estimated with data through 1998, and was 19 percent.

Replicating KMO Results

KMO reported the RMSE of their best-forecasting two-step model for the years over which they had developed it–1971-2000. They compared that RMSE to the RMSE from CBO's mean reversion

12. In another application, I fix the reversion rate at the 20 percent rate CBO has used most of the time since the winter 1999 baseline.

method to show the advantage of the two-step model. After they had completed this work, data on gains in 2001 became available. As a further test, KMO made year-ahead forecasts of 2001 with their best forecasting models. They did not, however, recompute their RMSEs to cover the full period.

I attempt to replicate KMO's forecasting record to learn more about the model and to make my extensions more credible. I use their BVAR forecasts of macroeconomic variables for 1960-2001, but differences are introduced from a few sources.

One source is that I use slightly different macroeconomic variables to predict gains for the year ending. KMO used data as of early 2002; my data are as of November 2003. Revisions in the interval cause small differences in variables starting in 1999. My replication extends through 2001, so only three years of data should differ on this account, although some earlier differences seem to arise.

Another source of difference is that I use final information on capital gains in 2000, whereas KMO used preliminary data. Their preliminary data showed gains were $638 billion; final data show gains were $644 billion. Finally, other differences seem to arise from differences in the many ways in which variables are transformed and used throughout the applications.

Differences in data keep my replication of their work from being exact, but I am close enough to be sure I follow the basic process correctly. KMO report a RMSE of 14.80 percentage points for the two-step model between 1971 and 2000 (see Table 3-1 for these and following RMSEs). My replication comes close, with a RMSE of 14.95 percentage points. My replication also captures the decade swings in KMO's errors. While my error is slightly larger on average, it is slightly smaller in the 1990s.[13]

I also replicate KMO's forecasts with the mean reversion method. KMO found larger errors with mean reversion than with the two-step model. Over the years 1971-2000 they found a RMSE of 18.57 percentage points. My forecasts with the mean reversion method have essentially the same RMSEs as KMO in the first two decades, and only a 0.12 percentage point difference in the third. Over the full 30 years, the difference in RMSEs is only 0.05 percentage points. The similarity of my replication and KMO's original is to be expected as I did both. The differences arise from revision of GDP and gains between early 2002 and November 2003.

---

13. The RMSEs from forecasting the growth rate of tax-adjusted gains can be evaluated better when viewed in relation to the typical variation in the growth rate of those gains. Between 1971 and 2000, tax-adjusted gains grew at an average annual rate of 12 percent. Their maximum growth rate was 44 percent and the minimum was a decline of 18 percent (a new low was set in 2001 when gains declined by 46 percent). The standard deviation of the growth rate is 16 percentage points over the 30 years, and it rises to 19 percentage points when 2001 is included.

The standard deviation is conceptually similar to the RMSE. If gains were forecast by assuming that they would grow in each year by the average rate over the period, then the standard deviation would be the RMSE of that forecasting method. The average growth rate over the period is not known until the period is over, but it can be approximated by the historical average up to the time of the forecast. This method's RMSE would exceed the standard deviation to the extent the mean changes over time. KMO applied this method and found its RMSE a little over 16 percent.

CBO stopped relying on an estimated reversion rate for the winter 2001 baseline, and in most situations since has somewhat arbitrarily relied on a fixed 20 percent rate.[14] Because 20 percent is the reversion rate CBO uses currently, I repeat the mean reversion forecasts of the replication with a fixed 20 percent reversion rate. Those forecasts have slightly lower errors over the 30 years, primarily because they do moderately better in the last decade. While the reduction in RMSE from using a fixed 20 percent rate is too small to be important, the closeness of the two error measures indicates that a constant 20 percent reversion rate is a reasonable approximation for the whole period.

Because I replicate the two-step and mean reversion methods fairly closely, the advantage for the two-step model over mean reversion is similar to that reported by KMO. They found that the two-step model lowered the RMSE in the first and third decades and over the full 30 years. I find a similar pattern. Over 30 years, they found the two-step model reduced the RMSE by 3.77 percentage points, and I find a reduction of 3.67 percentage points.

Turning to 2001, it is apparent that neither method anticipates the sharp decline that occurred in gains, although mean reversion comes closer. Nonetheless, I am able to replicate the forecasts of the two models fairly closely.

In 2001, actual gains declined almost 46 percent. Capital gains tax rates declined a very small amount, so the decline in tax adjusted gains is essentially the same. With the two-step model, KMO forecast that tax-adjusted gains would grow 13 percent and I forecast 11 percent in my replication. Their error is 58.82 percentage points and mine is 56.77 percentage points. With mean reversion, KMO and I forecast growth would be 0.7 percent with errors of about 46.5 percentage points. Mean reversion with the fixed reversion rate forecasts a decline of just over 3 percent, making an error of 42.49 percentage points.

When these errors for 2001 are combined with the errors from the previous 30 years, both RMSEs rise. The RMSE of the two-step model rises more, reducing its advantage over mean reversion substantially. According to my forecasts, the two-step model's RMSE from 1971-2001 is 17.89 percentage points compared to 20.12 percentage points for mean reversion. The advantage of the two-step model through 2001 is 2.22 percentage points, or 40 percent less than the advantage of 3.67 percentage points through 2000. Had KMO computed RMSEs through 2001, their results would be similar.

The Equation for Forecasting Gains in the Two-step Model

Replicating the forecasts of the two-step model allows me to explore the equation for forecasting

---

14. CBO used a 25 percent reversion rate in the winter 1995-1998 baselines and dropped to a 20 percent reversion rate for the winter 1999 and 2000 baselines to reflect the drop in the estimated reversion rate. CBO stopped using an estimated reversion rate in the winter 2001 baseline however, when it fell to nearly 10 percent and was statistically insignificant. Analysts at CBO felt a higher rate would reappear eventually. In place of the estimated rate, CBO continued using the 20 percent reversion rate (although 20 percent was not applied in the forecast of 2001 that year).

gains (step 2) in greater detail than KMO did. The equation explains the growth rate in tax-adjusted gains with forecasts of the four variables listed earlier. Three of those are the growth rates of productivity, real GDP, and real fixed investment; the fourth is the change in the ratio of real consumption of durables to real disposable income.

Actual values of the four variables can explain some of the movement in capital gains, but it is their forecasts that KMO used in the equation for forecasting gains. As KMO explain, that is because only forecasts of the explanatory variables will be available to forecast gains in the year ahead, and differences in the ability of the explanatory variables to be forecast will alter their relative importance in forecasting gains.

I examine their premise by comparing coefficients of the forecasting equation estimated with actual data to those estimated with forecasts (see Table 3-2). The equation estimated with forecasts is the one I use to forecast the growth of tax-adjusted gains in 2001. Both regressions are estimated over the years 1960-2000. Both equations have the growth rate of tax-adjusted gains as the dependent variable, however, the equation estimated with forecasts uses a prediction of gains in 2000 instead of the actual value.[15]

Three of the four estimated coefficients differ substantially between the two regression equations: those for real GDP, real investment, and the ratio of real durable consumption to real disposable income. The coefficient of real GDP jumps from -3.1 when estimated with actual data to -9.2 when estimated with forecast data. The coefficient on real investment makes a similar percentage change, going from 1.1 when estimated with actual data to 3.2 when estimated with forecast values. The coefficient on the ratio of durables to disposable income increases by a large amount but a smaller percentage than the other two: from 14 with actual data to 19 with forecast data. In contrast, the coefficient on productivity remains near 3.8 with actual or forecast values. Clearly, substituting forecast variables for actual ones substantially changes the coefficients that can best explain variation in the growth of gains. Had the coefficients estimated from actual data been used with the forecasts of the explanatory variables, the forecasts of gains would seem to have had much larger errors.

Although three coefficients are numerically different in the two equations, the differences are not statistically significant. Standard confidence intervals around the coefficients estimated with forecast values include the coefficients estimated with actual values. Nonetheless, the large numerical differences would lead to large differences in forecasts of gains, whether due to chance or not.

The equation estimated with forecasts of the macroeconomic variables explains much less of the growth rate of gains than does the equation estimated with historical values of the explanatory variables. Using forecasts of the macroeconomic variables reduces the fraction of the variation in growth rates explained from 38 percent to 23 percent (see the R-squared statistics in Table 3-2).

15. My replication of KMO's equation for predicting gains in the year ending predicts tax-adjusted gains in 2000 to be $593 billion. Tax-adjusted gains should have equaled actual gains which were $644 billion.

This difference is an indication of the difficulty of forecasting the explanatory variables.[16]

The difficulty of forecasting the explanatory variables also makes it much more difficult to identify the impacts of the forecast variables on the growth rate of gains. The probability that all the coefficients of the forecast variables could actually be zero is about 1 in 20 while the probablity that all the coefficients of the historical variables are zero is about 1 in 1,000. See the significance levels of the F-statistics in Table 3-2.

In reviewing the 30 recursive equations used to forecast gains from 1971 to 2000, I am surprised to see how poorly the forecast values of explanatory variables explain the growth of gains. Consider the forecasts between 1971 and 1980. The forecast of growth in tax-adjusted gains in 1971 is based on a regression equation estimated with forecasts of the explanatory variables in years 1960-1970. The hypothesis that all coefficients in the equation are zero comes nowhere near to being rejected by the F-test (p=.72). The forecast of growth in tax-adjusted gains in 1980 is based on the equation estimated with 10 more observations. The hypothesis that all of its coefficients are zero still cannot be rejected at normal levels of significance, although at p=.20 the coefficients are less plausibly all zero than the ones estimated 1960-1970. Nonetheless, all the forecasts made by the two-step model between 1971 and 1980 are made with equations whose coefficients are statistically indistinguishable from zero. Yet as shown in Table 3-1, the equation's forecasts of growth in tax-adjusted gains during the decade are more accurate than the forecast made with the mean reversion method.

As the number of years used to estimate the forecasting equation increases, the likelihood that all coefficients are zero decreases. However, the hypothesis that all coefficients are zero cannot be rejected at the normal 5 percent probability level until the equation is estimated through 2000.

The gradual decline in the likelihood that all explanatory variables are zero suggests that the forecasting equation will continue to improve its fit as more years become available for estimating the equation. For this improvement to occur, however, the general relationship between the explanatory variables and gains must remain unchanged. As I discuss below, the relationship is not what was expected, which decreases the odds it will remain stable.

The gradual decline in the likelihood that all coefficients in the estimating equation are zero is partly due to the decline in the likelihood that the coefficient on the forecasted real growth rate of GDP is zero. That probability falls below 5 percent first in 1987, and becomes less probable with additional years. No other variable becomes statistically significant until 2000 when the forecasted growth rate of real investment passes the 5 percent standard (see Table 3-2).

The statistical importance of forecasted growth in real GDP contrasts with the unimportance of the actual growth rate of real GDP. In the equation with historical values of all explanatory variables estimated 1960-2000, the coefficient of the growth rate of real GDP cannot be reliably distinguished

---

16. The use of actual tax-adjusted gains for 2000 in one equation and predicted gains in the other, as noted in preceding footnote, will also contribute to differences in R-squared statistics between the two equations. That contribution is likely to be a small part of the observed difference.

from zero (see Table 3-2). Furthermore, the coefficients on the three other variables are less likely to be zero. (None, however, have probabilities of being zero as low as 5 percent, the normal standard for rejecting that hypothesis.) The greater statistical importance of forecasted real GDP must be due in part to a greater ability to forecast real GDP than some of the other variables. KMO's RMSE from forecasting the growth rate of real GDP between 1960 and 2001 is one-third the size of their RMSE from forecasting the growth rate of real investment. Their RMSE from forecasting the growth rate of productivity, however, is slightly lower than that from forecasting real GDP growth.[17]

A related surprise in the estimated equations is that real GDP has a negative coefficient in both equations. Real GDP and gains grow over time, so they have a positive simple correlation. Furthermore, real GDP is a primary reflection of business cycles, and capital gains fluctuate with business cycles. For these reasons, variables based on GDP have frequently been used in equations to explain variation in gains, and their coefficients are typically significantly greater than zero. Hence, the negative sign in the forecasting equation is surprising.

The normal positive effect of GDP on gains appears to be better reflected by the other variables in the equation, which are themselves based on components of real GDP. The negative coefficient on real GDP could be caused by omitted components of GDP which are negatively related to gains. When I explore that hypothesis in Appendix B, I find theory and evidence suggesting that net exports and changes in private inventories are negatively related to gains. They could cause the negative coefficient on real GDP in the equations of Table 3-2. Those variables, therefore, seem like better candidates than real GDP for future forecasting equations.

Forecasting under Actual Conditions

The advantage of the two-step model over mean reversion that KMO found could change when CBO applies the methods under actual forecasting conditions. Applying the methods under actual conditions would require some adjustments, most notably, the use of CBO's macroeconomic forecasts. To investigate, I adjust both methods for real-time conditions to determine how their RMSEs change from those in my replication reported above.

Adjustments

In addition to using CBO's macroeconomic forecasts, CBO would need to forecast actual gains rather than tax-adjusted gains. The long-run target in the mean reversion method would have to be adjusted for tax rates expected at the time of the forecast. Finally, the equation used to predict gains for the current year would be CBO's equation rather than KMOs.

1. CBO Would Use Own Forecasts. The major difference between KMO's testing and CBO's application seems likely to arise in the forecasts of the four explanatory variables: productivity, real

---

17. RMSEs are shown in Table 3-3. Errors in measuring the change in the fraction of disposable income spent on durables are not comparable to those of the other macroeconomic variables because they are measured in different units.

GDP, real investment, and the ratio of real consumer durable consumption to real disposable income. KMO produced their own forecasts of these variables with a simplified BVAR model using data fully revised as of 2002. CBO uses a more extensive method, but each forecast used the data that was available at the time the forecast was made.

Differences in the data available to forecasters is likely to be important. When forecasting macroeconomic variables and gains for a particular year, say 1995, KMO used data for 1994 and earlier years as revised through early 2002. CBO's forecasts of 1995 were made late in 1994 when values for the explanatory variables for 1994 were incomplete, and CBO had to forecast values for the end of 1994 as well 1995 and beyond. Even the data that statistical agencies had released for 1994 and to a lesser extent the previous few years, were based on preliminary information and were revised in the next few years. These regular revisions often lead to significant changes. In addition, the data underwent several important conceptual revisions between 1992 and 2002, among which were major changes in the accounting for the computer sector that sharply altered both the level and growth rate of productivity and GDP. KMO's forecasts assume that those conceptual revisions are known at the time of the forecast; CBO and other forecasters who forecast in real time have to use pre-revision data.

The differences in data available and methods used are likely to cause CBO's real-time forecasts to differ from KMO's I do find that differences exist. As a way of identifying differences, I compute the errors each forecast method makes predicting the four macroeconomic variables used in the two-step method. KMO forecast growth rates of productivity, real GDP, and fixed investment, measured as changes in logarithms. Consequently, I measure errors for those variables as the difference between the actual (as reported in November 2003) and the forecasted change in logarithms. KMO forecast simple changes in the ratio of real consumption of consumer durables to real disposable income, so I measure errors as the difference between the actual and forecasted change in the ratio.

CBO's year-ahead forecasts of the four macroeconomic variables have a higher RMSE than KMO's forecasts for 1992-2002 (see Table 3-3). CBO's errors tend to be larger in the late 1990s when the economy repeatedly grew faster than most forecasters, expected. CBO's RMSEs are about 30 percent larger than KMO's for all four variables. For example, KMO's year-ahead forecasts of growth in real GDP for 1992-2001 have a RMSE of .0111 while CBO's have a RMSE of .0143. The uniformity of errors probably reflects the close interrelations between growth in real GDP, productivity, investment, and consumption of durables in the economy. If one is forecast to be higher, internal consistency leads to all being forecast to be higher.

CBO's larger errors seem to be primarily due to the differences in data available at the time the forecasts were made. CBO's forecast errors were not significantly different from those of other real-time forecasters for the same period.[18] All real-time forecasters underestimated the length and strength of the economic expansion and, later, missed the recession when it finally came. What matters for my analysis is mainly that the two forecasts differ. The difference raises the possibility

18. Congressional Budget Office, *CBO's Economic Forecasting Record*, September 2004.

that the accuracy of the two-step method may change significantly when it is applied with CBO forecasts.

To apply the two-step method with CBO's forecasts, I need to work those forecasts into the second step forecasting equation, where they become extensions of KMO's explanatory variables. The procedure is as follows. From CBO's December 1991 forecast, I take the forecasts for 1992 of the four explanatory variables. Those values are used to forecast gains in 1992 with the equation that I previously estimated from 1960 through 1991 with KMO's forecasts of explanatory variables for 1960-1991. In the next round, I add CBO's 1992 macroeconomic forecasts to KMO's forecasts of earlier years to estimate the gains forecasting equation through 1992. That equation is used to forecast gains in 1993 with the 1993 forecasts of the macroeconomic variables taken from CBO's next annual forecast completed in December 1992. The process continues until I estimate the gains forecasting equation with data through 2001. The explanatory variables are KMO's forecasts of 1960-1991 and CBO's forecasts of 1992-2001. Then, I use CBO's forecasts of the explanatory variables for 2002 from the winter 2002 baseline in the equation to forecast the growth of gains in 2002.

2. CBO Would Forecasts Actual Gains. KMO forecast tax-adjusted gains, but CBO would forecast actual gains. Thus, KMO's forecasts need to be adjusted to reflect actual gains. I do this by changing KMO's step two equation for forecasting gains. I substitute the growth rate of actual gains for the dependent variable and include as explanatory variables the tax rate terms CBO uses in its equations for predicting gains in the year ending. Those changes have their main impact on the forecast of gains growth in 1998. The tax rate on gains was lowered mid-way through 1997, so 1998 was the first full year at the lower rates. The tax terms allow a forecast of gains that reflect the change in tax rates from the partial year of 1997 to the full year of 1998. Tax changes to ordinary rates in 1993 and 2001 caused small indirect changes in the net tax rate on capital gains, but the estimated impacts were inconsequential.

I ignore the tax rate change in 1997 when replicating CBO's forecasting circumstances of late 1996. The forecasts are intended to reflect the law as it was known at the beginning of 1997 before the legislation was introduced. (CBO did not anticipate the law change in its actual forecast either). To judge the accuracy of forecasts for 1997, therefore, I remove an estimate of how much the tax rate reduction in 1997 added to the growth rate of gains in that year. The estimate of the tax effect was made using CBO's combination equation for predicting gains in the year ending. That equation, estimated through 2002, predicts that the legislation in 1997 accounted for 17.0 percentage points of the total growth of 39.9 percent in that year, implying that gains would have grown about 22.9 percent had taxes not changed. Therefore, I measure errors in forecasts of 1997 against a target of 22.9 percent growth.

3. CBO Would Adjust the Mean Reversion Target. Mean reversion as applied by KMO assumed that the future ratio of tax-adjusted gains to GDP would revert to its historical average (A(t-1). When actual gains are being forecast instead of tax-adjusted gains, the target to which the ratio will revert depends on the tax rates expected to be in place in the coming year. To account for those rates, the historical average ratio would be replaced by an estimate of what the ratio had tended to be at the tax rates expected to be in place during the year forecast. I follow CBO's current practice for obtaining that estimate by using an equation that explains the historical ratios in terms of a constant term, the year's top tax rate on gains, and a dummy variable taking the value of one in 1986. The equation is estimated from 1954 through the year before the year ending. The ratio predicted by the equation at the tax rates anticipated to be in place during the year ahead replaces A(t-1) as the target to which gains revert.

4. CBO Would Use Its Equation to Predict Gains. CBO predicts gains for the year ending with a different set of variables than those KMO found to work best with tax-adjusted gains. My testing of those two equations in the previous chapter finds that CBO's combination equation predicts actual gains more accurately than KMO's alternative. Consequently, when I apply the two-step model under actual conditions, I predict gains for the year ending with CBO's equation. The predicted value will differ from the value KMO predicted both because the equations differ and because the actual data CBO had for each forecast differs from that KMO used. The two-step model uses predicted gains for the year ending in the dependent variable of the second step's forecasting equation and for computing the dollar level of gains in the year ahead. Mean reversion uses the predicted value for the year ending to obtain the ratio of gains to GDP for the year ending (R(t-1)).

Findings

When I apply the two-step model under actual CBO forecasting conditions, I find slightly larger errors than I find in my replication of KMO's forecasts. My replication of the two-step forecasts for the growth rate of gains for 1992-2001 has a RMSE of 21.75 percentage points and my application under actual forecasting conditions has a RMSE of 22.24 percentage points (see Table 3-4). An increase in the RMSE is to be expected because CBO's macroeconomic forecasts had larger errors than KMOs, but the small size of the increase surprises me. Whether surprising or not, the small increase indicates that no fundamental inconsistency is introduced by applying the two-step model under CBO's actual forecasting conditions.

When I apply mean reversion, I use the 20 percent reversion rate that CBO has been using in recent years for all forecasts from 1992 through 2001. With the real-time data, the mean reversion method also has a larger RMSE than it has when applied with the revised data set and KMO's forecast of GDP. The RMSE for the growth rate of gains during 1992-2001 increases from 23.44 to 24.43 percentage points, a slightly larger increase than for the two-step model Thus, over those 10 years, forecasting under actual conditions increases the advantage of the two-step model modestly.

KMO did not forecast macroeconomic variables for 2002. When I substitute CBO's macroeconomic forecasts for KMO's in the real-time analysis, I can forecast 2002 gains with both methods. The two-step model forecasts 2002 slightly less accurately than mean reversion does. The RMSE of the two-step model increases slightly and that of mean reversion declines slightly. The advantage of the two-step model over mean reversion during 1992-2002 is 1.95 percentage points, moderately smaller than I find through 2001.

The small advantage of the two-step model over mean reversion in RMSE covers up large differences in specific years (see Figure 2). The two-step model has much smaller errors than mean reversion during the boom years from 1996 through 2000, but it has a substantially larger error in 1994 and 2001. The errors of the two methods are closer in the years 1992, 1993, 1995 and 2002.

It is apparent from Figure 2 that both methods have difficulty capturing the boom and bust of the last decade. The two-step model captures the upswing better, but mean reversion captures the downswing better. Overall, the errors of both methods are substantial (22.35 percentage points for two-step and 24.26 percentage points for mean reversion). The advantage of the two-step model amounts to an 8 percent reduction in the RMSE of the mean reversion method between 1992 and 2002.

Finally, CBO has been applying variants of the mean reversion method during the 1992-2002 period. CBO's forecasts were slightly more accurate than mean reversion as I applied it to the same period, and slightly less accurate than the two-step model (see Table 3-4). The slight improvement in CBO's historical forecasts over my mean reversion forecasts occurs because CBO supplemented mean reversion in three years. In the forecasts for 1995 and 1996, CBO raised its forecast of gains growth in the coming year because it anticipated taxpayers would shift gains to those years in anticipation of tax legislation reducing tax rates on capital gains. Those adjustments brought the forecasts closer to the actual growth of gains. In the third year, 2001, CBO abandoned mean reversion, assuming instead that gains would remain at the level predicted for 2000. Gains did fall relative to GDP that year, making CBO's actual forecast less accurate than strict application of mean reversion with a 20 percent reversion rate. The larger error CBO made forecasting 2001 does not entirely offset the smaller errors it made forecasting of 1995 and 1996.

Conclusion

Neither the two-step model or the mean reversion method forecast well during the last decade. The two-step model captures the boom better; mean reversion captures the bust better. About 40 percent of the advantage KMO report for the two-step model appears to arise because their analysis ends at the peak of the boom in 2000 instead of after the bust is under way in 2001. The modest advantage that remains after 2001 data are included is not diminished by switching to the real-time data available to CBO nor by extending the comparison to 2002. CBO's actual forecasts of growth in gains for 1992-2002 were about as accurate as those of the two-step and mean-reversion methods.

If the decline in gains relative to GDP in 2001 and 2002 signals a continuing gravitation of gains toward the overall size of the economy, then the advantage of the two-step model may remain small. If gains drift above their historical size relative to GDP in coming years, the advantage of the two-step model may return to its size over the 1971-2000 period.

My examination of the forecasting equation in the two-step model found support for their use of forecasts of explanatory variables. But it also found evidence that real GDP does not belong in the equation and might be replaced by net exports and the change in private inventories.

The specification problem could be addressed by retesting the KMO equation with the additional components of GDP. That approach would require re-estimating the BVAR model on all the new possible combinations of variables, which is beyond the scope of this paper. An alternative approach is undertaken in the next chapter.

IV: Alternatives for Forecasting Capital Gains

KMO identified several variables that appear to be useful for forecasting gains. Yet, I have found that the forecasting equation in their two-step model has a specification problem: forecasts of the growth rate of real GDP are negatively related to the growth of gains when a positive relation is expected. In Appendix B, I report evidence that components of GDP, such as net exports and changes in private inventories may account for the specification problem with real GDP. Thus, it may be possible to rectify the specification problem of the forecasting equation by changing the number variables used to forecast gains.

CBO's mean reversion method also has specification problems. It would be helpful if a revision of any forecasting equation would also solve those problems. One specification problem is that the reversion rate CBO uses is arbitrary. It comes from an equation that CBO used in the 1990s to predict gains for the year ending. That equation also estimated a reversion rate and when estimated with data through the later 1990s the reversion rate was around 20 percent. However, when the equation was estimated through the early 1990s the reversion rate was closer to 25 percent and when estimated through 2000 was well below 20 percent. Furthermore, the equation specified that reversion was in addition to whatever changes occurred in stock prices, but it is now applied without any information about stock prices.

Another problem with CBO's mean reversion method is that it incorrectly specifies the impact of future tax changes. An example illustrates the nature of this problem. The top tax rate on capital gains is scheduled to rise from its current 15 percent rate to 20 percent in 2009. Mean reversion reflects the increase by reducing the targeted ratio of gains to GDP for 2009, and then forecasting that the ratio expected for 2008 moves 20 percent of the way to that target in 2009. The remainder of the adjustment to the 2009 target is spread out over many future years. Econometric studies of taxpayer response to realizations, however, find that taxpayers respond more rapidly. They are likely to shift realizations from 2009 into 2008 to beat the increase in the tax rate, and they are likely to have largely adjusted by 2010 or 2011. Thus, mean reversion's forecast for 2008 and beyond is inconsistent with existing evidence. It is also inconsistent with CBO's prediction for the year ending because the equation CBO uses to predict those gains does reflect the evidence of faster responses.

Strategy

I try to resolve these problems by adapting CBO's equation for predicting gains in the year ending to forecasting the year ahead. A conceptual version of the combination equation CBO uses to predict gains for the year ending is presented in the top of Table 4-1; an estimated version appears in Table 2-4 of Chapter 2.

Strategy for Adapting the Combination Equation

I start from the combination equation for several reasons. One is that the equation can incorporate the variables that KMO and the previous chapter found could be important for forecasting. Productivity trends, though not cyclical fluctuations in productivity, are incorporated in the equation through the potential GDP variable. Fixed investment can be incorporated, as already done in Chapter 2. The other components of GDP, durables consumption, net exports, and changes in private inventories, can be added in similar ratio form to test whether they help explain gains historically.

A second advantage is that the equation should be able to address the limitations listed above in the current mean reversion method. The equation already includes CBO's best formulation of tax rate variables. Therefore, it should be able to better forecast how tax changes currently legislated to take effect in future years will affect our gains forecast in those years. Furthermore, its structure can be modified to estimate a rate at which the ratio of gains to potential GDP (hereafter, the gains ratio) reverts to its size expected under current tax rates. Because CBO's macroeconomic forecast predicts that GDP itself reverts to potential GDP, reversion of the gains ratio to its expected size would be similar to CBO's current method of assuming that the ratio of gains to GDP reverts to its expected size under current tax rates.

The equation has other advantages explained in Miller and Ozanne. One advantage relevant here is that it allows GDP and other macroeconomic variables to have different impacts on gains during business cycles than during steady economic expansions. Gains tend to decline more rapidly than output going into a recession and recover more rapidly coming out. But gains and output tend to grow at similar rates over longer intervals as shown by the absence of any trend in the ratio of gains to GDP. The equation keeps gains growing at the same rate as GDP when GDP is expanding at the rate of potential GDP, but allows gains to grow at a different rate when GDP is expanding faster or slower than potential, as happens during the business cycle. The forecasting equations in the two-step model do not allow economic growth to have such differential effects.

The equation also has disadvantages for forecasting, some of which are addressed here.[19] First, the equation includes stock prices and dollar volume of stock trades. CBO and KMO both developed forecasting approaches without stock prices because stock prices are extremely difficult to forecast. My approach will follow theirs by dropping the terms from the equation.

In addition to removing both stock market variables, I drop multifamily housing starts. The variable has limited significance in the combination equation and also is difficult to forecast, as reported in Chapter 2. Recall from Chapter 2 that fixed investment apparently can replace both the GDP and housing starts variables in the combination equation and slightly improve the statistical fit. Thus, fixed investment may replace both variables in the equations developed here.

Another disadvantage is that the specification in logarithms precludes the use of variables that have negative values. Both the change in inventories and net exports are negative in some years. The accommodation is to experiment with the equation as simply the change in the ratio of variables

19. Other disadvantages are discussed at the end of the chapter and left for future work.

relative to potential GDP. Miller and Ozanne experimented with that formulation and switched to the logarithm specification for a small improvement in fit.

The result of all the modifications is to change the combination equation from its original form shown in the top of Table 4-1 to the simpler equation shown in the middle of the table.

The simpler equation can test for the importance of net exports and changes in inventories by substituting them for GDP. Reversion in the gains ratio can arise partly through reversion of GDP to potential, or in whatever components of GDP are found to be useful. Reversion in omitted factors that show up in the residuals of the simplified equation can be modeled statistically, depending on the time-series properties of the gains ratio and the residuals of the simplified equations.[20]

A final adjustment to be considered is estimating the simplified equation in the level of the ratio instead of the annual change in the ratio. That version is shown in the bottom of Table 4-1. Miller and Ozanne used the change specification because over the period they developed the equation, 1952-1998, they could not strongly reject the hypothesis of a unit root in the gains ratio. The presence of a unit root means that surprise changes in one year are fully carried over to the next year, as occurs in a random walk. If the gains ratio has a unit root, then equations explaining the gains ratio are likely to lead to spurious regression results when explanatory variables also have unit roots.[21] Estimating the equation in first differences avoids potentially spurious findings.

Estimating the equation in first differences, however, restricts what can be forecast. A difference equation can only forecast how much the ratio changes from one year to the next. That forecast must be combined with a value of the ratio for the prior year to forecast the level of the gains ratio (and gains) for the year being forecast. That limitation is particularly problematic for forecasting gains a year ahead because gains and the gains ratio for the year ending are unknown. Gains for the year ending must be predicted and that prediction is subject to its own errors.

The period over which Miller and Ozanne developed their equations included part of the stock market bubble, which could have distorted their tests for a unit root. Tests for shorter or longer periods could lead to the rejection of the unit root, which would then allow estimation of the equation in annual levels instead of changes. Thus, I will test for the presence of unit roots in the gains ratio, the ratio of GDP to potential GDP (the gap) and the ratio of other potential explanatory variables to potential GDP.

While testing for unit roots is useful, the tests are imperfect. Hence, some thinking about the likelihood of unit roots is also useful. For instance, the gap seems unlikely to contain a unit root. The Federal Reserve Board continually works to keep actual output from wandering far above or below potential output as it tries to keep inflation and unemployment low. In addition, the President

---

20. Diane Nhu Pham applied these ARIMA procedures to a similar gains forecasting equation at CBO in the summer of 2004.

21. Estimating with the ratios directly can be appropriate even when the ratios have unit roots if the ratios move together, or are cointegrated. In these cases, equations also can be estimated in first differences and error correction terms from level equations can be included.

and Congress occasionally act to counter unemployment and inflation. To the extent these institutions are successful, they should keep the gap from wandering randomly, as is required for a unit root. To the some extent, these institutions probably keep the ratios of consumption and investment to potential output from wondering randomly, merely as a derivative of their efforts to keep the gap from wander too high or too low.

The gains ratio may not contain a unit root either. No powerful institution tries to keep gains at any particular size relative to potential output, which means a unit root is more likely than with the gap. Still, it is unclear that a surprisingly large amount of realizations in one year should become embedded in a base level of realizations the next year. A surprise would become embedded if it reflects an unobserved increase in the pool of accrued gains that will continue into the next year as well. It would not become embedded, however, if the surprise means that fewer gains will remain to be realized in the next year. The frequent return of the ratio to its historical average is consistent with there being no unit root.

Strategy for Testing Forecasts

The equations which fit the historical data best will not necessarily forecast the best. One reason is that coincidental relationships between the gains ratio and other variables tend to be found when experimenting with a variety of alternative variables. To test forecast ability, I develop a few promising equations with the historical data available at the end of 1991. Then, I use the equations to forecast gains one year ahead in 1992 and in the following years though 2002 in a replication of the actual process CBO would use. The forecast accuracy of the equations can then be compared. The process is a repeat of the real-time tests used in the previous chapters. A full test of forecasting ability would require more years of experience, but none are readily available.[22] Nonetheless, the limited evidence on forecasting accuracy can be combined with the evidence on statistical fit to historical data to evaluate the alternative equations.

The particular years used for model development and forecasting tests are determined by the availability of historical CBO forecasts of macroeconomic variables. As discussed in the earlier

---

22. KMO identified an additional reason that the equations that explain the historical pattern of gains the best may not forecast the best. That is because the variables that are most successful at explaining gains once they are known may be harder to forecast than other variables. As a result, the forecasts of those other variables may forecast gains more accurately than the forecasts of the variables that best explain gains. KMO cited stock prices as a variable that explains gains well once it is known, but may be useless for forecasting because of the difficulty of forecasting it. KMO further noted that differences in the ability to forecast variables means that coefficients estimated from historical data need not be the best ones for forecasting. The previous chapter reports support for this point with the forecasting equation they developed in their two-step approach.

To overcome these problems, KMO developed macroeconomic models to test the ability to forecast the potential explanatory variables. They also estimated the equations to be used for forecasting with their forecasts of the explanatory variables. Finally, they selected among equations based on their accuracy in forecasting gains, not in explaining gains with historical data.

I have decided against implementing their full procedures. Revising their macroeconomic models or developing alternatives would take me a long time. Furthermore, the forecasts of those models would not necessarily be consistent with CBO's more complex forecasts which would be used to forecast gains. Such inconsistency appeared with KMO's integrated quarterly approach, though not with KMO's two-step approach. Finally, the limited evidence I have through 2002 finds modest improvements in forecasting with their method, and no improvement over the two years they did not use for model development. Their model development may suffer from over-fitting.

chapters, CBO's earliest forecast is the one for the winter 1992 baseline, which was completed in December 1991. At that time, preliminary data on gains were available for 1990, but not for 1991. Consequently, I use data from 1952-1990, that were available to researchers at the end of 1991, to develop potential forecasting equations. I begin the test period with 1952 because that was the first year used by Miller and Ozanne. Once the equations are developed, I use them to forecast the ratio of gains to potential output in 1992, or its change from 1991, based on CBO's winter 1992 forecast of the explanatory variables.

Note that equations forecasting the change in the gains ratio in 1992 must be combined with an estimate of the gains ratio in 1991 to forecast the ratio for 1992. An estimate of gains for 1991 is also necessary to compute a growth rate for gains from 1991 to 1992. I get an estimate of gains in 1991 by using CBO's combination equation estimated on the data through 1990 and the values of its explanatory variables for 1991 as known in the end of that year.

To forecast the gains ratio for 1993, I re-estimate the selected equations with data available at the end of 1992 and then use CBO's forecasts of the explanatory variables from the winter 1993 baseline to forecast the gains ratio, or the change in the gains ratio, for 1993. The estimate of gains for 1992 is constructed by re-estimating the combination equation with the available historical data through 1991 and using it with CBO's forecasts of the explanatory variables for 1992. By repeating the 1993 process for annual winter baselines through the winter 2002 baseline, I obtain 11 forecasts of growth rates whose accuracy I can compute using the RMSE.

Development of Equations

The first step is to test for unit roots in the selected variables. The second is to develop alternative specifications consistent with those tests. All work uses data through 1990 as known at the end of 1991.

Unit Root Tests

Augmented Dicky-Fuller tests for unit roots are conducted on the gains ratio, the gap, and the ratios of the alternative explanatory variables to potential GDP.[23] The tests reject the hypothesis of a unit root in most ratios under consideration. Unit roots were rejected at the 5 percent probability level for gains, fixed investment, consumer durables, and the change in private inventories, all measured relative to potential output. Although there are sound reasons for not expecting a unit root in the gap, slightly different specifications of the test narrowly rejected or narrowly failed to reject the hypothesis of a unit root for that ratio.

An augmented Dickey-Fuller test could not reject the hypothesis of a unit root in the export or import ratios. Nor could it reject that hypothesis for the difference between the two ratios (the net export ratio).

23. The tests are done with data starting in 1951 so that change equations can be estimated with data from 1952 to 1990.

Based on those results, an equation in first differences of ratios may be needed to avoid spurious regression results across all variables. Depending on the ones that turn out to be promising, though, an equation in ratios could well be acceptable. Therefore, initial testing is done with changes in ratios (referred to as change-in-ratio equations).

Tests of Variables 1952-1990

The simplified equation in the middle of Table 4-1 adequately explains variation in the gains ratio when estimated over the years 1952-1990 (see column 1 of Table 4-2). Its R-squared statistic is 0.80, and the estimated coefficients on the tax terms and the gap have the expected signs and are statistically different from zero at the 1 percent probability level.[24] Furthermore, the residuals appear to be well behaved. The Durbin-Watson statistic cannot reject the hypothesis that the residuals are serially uncorrelated, and the Jarque-Bera statistic cannot reject the hypothesis that they are normally distributed (see Table 4-2).

Replacing GDP with either fixed investment or consumer durables does not improve the explanation of the gains ratio. The R-squared statistic for the equation with the investment ratio is also 0.80 and for the equation with the consumer durables ratio it is slightly higher at 0.83. Investment and durables seem to be explaining the same acts of realization because when both are included in an equation the durables ratio remains significant but the investment ratio becomes completely insignificant. Nor does the sum of the two components improve on the fit of either individually. The equations with either explanatory variable have well behaved residuals.

Adding other components of GDP to equations with either the investment or durables ratios does not improve the explanation of the gains ratio. That finding contrasts with the results reported in Appendix B. In the equation with the investment ratio, for example, the change-in-private-inventories ratio has a positive coefficient instead of the negative one found in Appendix B and is completely insignificant. Imports do have the negative sign found in Appendix B, and the t-statistic is above 1.0, but it is not significant at even the 10 percent level. The net export ratio comes the closest to being statistically significant; it has the expected negative sign and is almost statistically significant at the 10 percent level.

Adding the same variables to the equation with the durables ratio leads to similar results. The failure of the change in private inventories and net exports to contribute significantly in the current specifications when they did in the previous chapter probably results from the different years covered and the use of the ratio specification instead of the growth rate specification. In a later test of the variables with data through 2002, I found results very close to those reported here.

The results leave as an alternative to the simplified equation one with either the investment ratio or the durables ratio. I think investment is likely to forecast gains better than consumer durables will. The relationship between gains and consumer durables could well reflect two-way causation. People

24. These regression statistics cannot be compared to those of the combination equation presented in Table 2-4 because the dependent variables have different specifications.

may sometimes decide to buy durables after they decide to realize gains, and other times decide to realize gains to enable them to purchase durables. The influence of realizations on the purchase of consumer durables would not be well forecast in the macroeconomic forecasts of durable consumption. For that reason, the primary equation I consider as an alternative to the simplified equation is the one replacing the gap with the investment ratio. Regression statistics for the alternative are shown along side those of the simplified equation in Table 4-2.

The tests for unit roots reject the hypothesis of such a root in the investment ratio and the gains ratio. Those rejections indicate that the equation with those variables can be estimated in the levels of those ratios with little risk of spurious regression results. The tests do not reject the possibility of a unit root decisively for the gap, but it is unlikely on theoretical grounds to have one. I therefore estimate ratio equations as well as change-in-ratio equations.

Estimating the equations in ratios reveals strong evidence of serially correlated errors. In response, I re-estimate both equations with a first-order serial correlation adjustment. The resulting regression with the gap estimates the serial correlation coefficient to be 0.77, which is a little over two standard errors below 1.0. The serial correlation coefficient in the regression with fixed investment is estimated to be 0.69, which is also just over two standard errors below 1.0. The serial correlation coefficient being more than two standard errors below 1.0 in both equations indicates that the first difference specification most likely over simplifies both equations. Table 4-3 shows regression statistics for both equations with the gains ratio.

A serial correlation coefficient of 0.77 means that 77 percent of the unadjusted residual in one year carries over to the predicted gains ratio in the next year. Equivalently, 23 percent of the unadjusted residual is not continued, which can be interpreted as a reversion rate for the unexplained divergence of the actual gains ratio from its predicted ratio. This rate of reversion is similar numerically to the 20 percent rate CBO has been using recently in its mean reversion forecasts, but it is not conceptually the same. It is reversion beyond that occurring in the gap, while CBO's rate includes any effect of reversion in the gap.

When the serial correlation adjustments are incorporated, both equations in ratios estimate coefficients on explanatory variables very similar to those in the change-in-ratio specifications. Both ratio equations have R-squared statistics near 0.83, and the adjusted residuals are well behaved.

Because I have only four equations, all of which fit the data well, I test the ability of all four to forecast the gains ratio. The two ratio equations should forecast better than the two change-in-ratio equations because they include more information about the ratio, but they could fail if that extra information is unstable over the test period. The two equations with the gap may do better than the two with the investment ratio because GDP is forecast more accurately than fixed investment over the test period, as shown in Table 2-3 of Chapter 2.

Re-Estimating the Equations in Subsequent Years

The equations reported in Table 4-2 and Table 4-3 are used to forecast the gains ratio for 1992 with data from CBO's winter 1992 baseline. To forecast the gains ratio in subsequent years, the equations are re-estimated with data current at the time of the new baseline. Those data include an additional year of historical information and any revisions of the previous historical data that have become available since the last baseline. Before reporting on the forecasts made with those equations, I review how the additional and revised data affect the estimated equations. I find that adding in the years with the stock market bubble and the companion economic expansion and recession force changes in the estimated equations.

One noticeable change occurs in the serial correlation coefficient, which indicates how long surprises in the gains ratio persist. The gains ratio had been lower than predicted in the early 1990s and gradually recovered through 1995. That recovery was roughly consistent with previous experience as shown by the relative constancy of the estimated serial correlation coefficient when additional years of experience are added through 1996 (see Table 4-4). The one exception is the dip in the coefficient when data for 1994 are added. The gains ratio rose above its predicted level in 1996 with the beginning of the stock market and economic booms, and continued to increase in each year through 2000. That absence of reversion causes the estimated serial correlation coefficient to rise with each year of additional data. With data through 2000, the estimated reversion rate is 0.95, less than a standard error below 1.0.[25] The gains ratio fell well below its predicted level in 2001, and when data for that year are included, the estimated serial correlation coefficient returns abruptly to its level in 1996 and is again more than two standard errors below 1.0.[26]

A second change occurs in the coefficient of the gap variable. When the preliminary value for gains in 2001 became available in late 2002, it showed that the gains ratio had fallen from 6.7 percent in 2000 to 3.2 percent. Data at the same time showed that the gap fell from 1.03 percent to 1.00 percent that year, and essentially no change occurred in the tax rates on gains. In an attempt to explain the unusually large drop in the gains ratio with the smaller drop in the gap, the regression equation estimated with that data raises the estimated coefficient on the gap by about one standard error from its value with data through 2000.

The increase in the estimated coefficient of the gap and the big drop in the estimated serial correlation coefficient are insufficient to explain much of the drop in the gains ratio in 2001. As a result, an exceptionally large residual remains, and causes the Jarque-Bera test to reject for the first year the hypothesis that the residuals are normally distributed. Table 4-5 shows how the regression statistics change from those estimated with data through 1990 when data through 2000 and then 2001

---

25. When the serial correlation coefficient is within a standard error of one, as it is when estimated with data through 1999 and 2000, the hypothesis that surprises persist indefinitely cannot be rejected. That is the hypothesis of a unit root. Direct tests for the possibility of a unit root are needed, because the normal standards of statistical significance are incorrect if in fact a unit root is present. It turns out that unit roots in the gains ratio cannot be rejected for 1999 or 2000, but they can be in other years. The gap and the investment ratio continue to reject the possibility of a unit root in those years, so spurious regression results are unlikely. Furthermore, the estimated coefficients in the ratio equations remain close to those in the first difference equations, suggesting that no spurious effects are distorting the estimated coefficients in the ratio equations.

26. The pattern in the serial correlation coefficient is similar to that found for a reversion rate estimated as part of an application of CBO's mean reversion method described in Chapter 3. One difference is that the serial correlation coefficient adjusts less because the gap in the equation explains some of the variation in the gains ratio.

are added.

Similar patterns occur in the equation that uses the investment ratio in place of the gap. Its estimated serial correlation coefficient rises steadily to near one as data from 1997 to 2000 are included, and then plunges back to its early 1990s level when data for 2001 are included. Data from 2001 cause a jump in the estimated coefficient on the investment ratio and still create a residual large enough that the hypothesis of normally distributed residuals is rejected.

The equations estimated in first differences change less as data through 2000 are added. Because their structure assumes that any unexplained change in the gains ratio in one year is carried over to the next year, the experience in 1997-2000 only confirms their structure. Nonetheless, the experience of 2001 causes about as large a shock to the change-in-ratio equations as to the ratio equations. In the equation with the change in the gap, the drop in the gains ratio causes the coefficient estimated for the gap to rise by about a standard error. That adjustment still leaves the residuals failing the test for normality. The impact on the equation with fixed investment is similar. Table 4-6 shows regression statistics for the change-in-ratio equation with output when it is estimated through 1990, 2000, and 2001.

Test of Forecasting

The four equations are estimated with data available at the end of each year from 1991 through 2001. At the end of each year, available data go through the year before because gains for the year ending are unavailable. Gains for the year ending are predicted with CBO's combination equation. Then the four equations developed here are used to forecast the gains ratio for the year ahead. Those year-ahead forecasts are made for 1992-2002. The winter 2003 baseline can be used to estimate the four equations through 2001 (those estimates have just been discussed), and the equations used to forecast the gains ratio in 2003. Errors for 2003 cannot be computed at this time, though, because gains for 2003 are still unknown.

I measure errors in forecasting the gains ratio. Errors in forecasting the change in the gains ratio would be an alternative, but the ratio is the preferred item to know.

Accuracy is measured by the root mean squared error (RMSE), as in previous chapters. The RMSEs presented here cannot be compared to the ones in earlier chapters, however, because the earlier ones measure errors in forecasting the growth rate of gains. I present comparable RMSEs after discussing RMSEs from forecasting the gains ratio.

Forecasts of the Gains Ratio

The RMSEs of the four equations are similar. The largest is less than half a percent above the smallest (see Table 4-7). While I am unaware of a formal statistical test for differences in RMSEs, it seems clear that the small differences in RMSEs generated by the four equations are not statistically significant. All four equations forecast about equally well over the entire 11 years.

The small differences in RMSEs that the equations do have are unexpected. The equations with the investment ratio as an explanatory variable have smaller RMSEs than the equations with the gap as an explanatory variable. That result is unexpected because the equations with the gap fit better when estimated through 1990 and because output is forecast more accurately than investment over the period (see Table 3-3).

A second surprise is that the change-in-ratio equations forecast gains ratio more accurately than the ratio equations. Usually, a level equation will forecast the level more accurately than the same equation in differences because it contains more historical information about the level.

The explanations for both surprises arise from the way the stock market bubble and economic boom affected gains. Those effects can be seen in the yearly data. The yearly data also show that the biggest differences among the forecasts are between the ratio and change-in-ratio equations, not between those with the gap and the investment ratio. Those bigger differences favor the ratio equations in some years and the change-in-ratio equations in other years, so the difference in RMSEs over 11 years turns out to be small. In addition, the yearly data show which years were the hardest to forecast.

When the stock market bubble swells in the latter half of the 1990s, the gains ratio starts rising in large steps (see Figure 3).[27] Then in 2001, after the bubble starts deflating, the ratio takes a giant step down to a level about equal to that in 1996. It steps down again in 2002 to about its level in 1995.

The forecasts of the change-in-ratio equation with the gap, shown in Figure 3, illustrate the main forecasting problems of all four equations. The equation forecasts fairly accurately until the actual gains ratio takes off in 1996. From then through 2000, its forecasts are consistently below the actual but follow along rather than falling farther behind. The forecasts follow along largely because they build on gains for the year ending that are predicted by CBO's combination equation. That prediction accepts as permanent all of the surprise from the year before and adds information about stock market and other economic activity in the year ending. As long as the boom continues through the year ending, the change-in-ratio equation builds its forecast from a higher base than it did for the year before. The change-in-ratio equation uses CBO's forecast of actual and potential output in the year ahead to adjust its forecast for the year ahead from what the combination equation predicts for the year ending. During most of the boom, CBO was forecasting that actual output would grow less rapidly than potential in the year ahead, which by itself would forecast a decline in the gains ratio from what CBO predicted for the year ending. However, the increase in the amount predicted for

27. The gains ratio for 1997 shown in Figure 3 is 0.03882, which is not the actual ratio of 0.04422. As explained in the previous chapter, forecasts made in the end of 1996 could not anticipate the reduction in the capital gains tax rate legislated in 1997 and becoming effective for part of the year. To make a more relevant standard against which to judge the accuracy of forecasts using information available at that time, I have reduce gains for 1997. Consequently, the gains ratio actually increased less between 1997 and 1998 than shown in Figure 3. Appendix Table C-1 has values for bars shown in Figure 3.

I determined the amount to reduce gains in 1997 by estimating with the combination equation the amount of the growth in gains in 1997 that was due to the tax change. Gains actually grew 40 percent in 1997, and the combination equation predicts that 17 percentage points of that were caused by the tax change. Thus, I calculated adjusted gains to be 23 percent above gains in 1996. This correction is similar to the construction of tax-adjusted gains described in Chapter II. Very small changes in taxes occurred in 1993 and 2001, and the combination equation shows these changes had an imperceptible impact on the gains ratio in those years.

the year ending was enough in those years to push the forecast for the year ahead higher than the comparable forecast made in the preceding year.[28]

With the data available at the end of 2000, the change-in-ratio equation builds on the boom through 2000 to forecast an even higher gains ratio for 2001 than it did in 2000. CBO's forecast for 2001 of slower growth in GDP than potential only tempered that forecast slightly. As a result, the sharp drop in the actual gains ratio is totally missed by the change-in-ratio forecast, and that leads to the largest forecast error by far. The change-in-ratio equation could foresee only a slowing in the economic boom. The equation's forecast of 2002 made with data available at the end of 2001, builds on the decline in stock prices and the recession that the combination equation uses to predict a decline in gains for 2001. The decline predicted for 2001 turned out to be too small, once actual data were finalized. That error and CBO's slightly optimistic forecast for output in 2002, lead to the second largest error in forecasting the gains ratio.

In comparison to the change-in-ratio equation, the ratio equation also using the gap does not predict that all of the surprise in the gains ratio from the preceding years will carry over to the year ahead. The serial correlation coefficient indicates the fraction that is carried over. During the boom years from 1996 to 2000, when the surprises kept building, the ratio equation carried over only part of the surprise and therefore consistently forecast lower ratios for the year ahead than the change-in-ratio equation (see Figure 3). That caused larger errors than the change-in-ratio equation had. The succession of surprising jumps in the gains ratio does cause the serial correlation coefficient to rise towards 1.0, as discussed previously. With the data available at the end of 2000, which include gains through 1999, the serial correlation coefficient is estimated to be over 0.9, so the ratio equation is accepting over 90 percent of the surprise predicted for 2000. For that reason, the ratio equation predicts for 2001 that most of the surprise in earlier years continues, and therefore it has an error nearly as large as that of the change-in-ratio equation. The serial correlation coefficient that hurt the forecasts of the ratio equation during the boom is too weak to help much in the bust. As a result, the ratio equation has larger RMSEs 1992-2002 than the change-in-ratio equation.

Each equation that use the investment ratio consistently forecasts nearly the same gains ratio as its counterpart with the gap. The small differences tend to favor the forecasts made with the investment ratio; both the ratio and change-in-ratio equation with investment have a lower RMSE than the counterpart with the gap. Their lower RMSEs probably arise because investment moved more closely with the stock price bubble than did the rest of GDP.

Forecasts of the Growth Rate of Gains

I convert the forecasts of gains ratios for the year ahead to forecasts of growth rates using CBO's macroeconomic forecast of potential GDP and the prediction of gains for the year ending. Then, I compare these forecasts to actual growth rates to measure errors in forecasting growth rates for each year and compute RMSEs.

---

28. The forecasting equations also incorporate the tax change in their forecasts of 1998.

The differences among these RMSEs are larger than the differences among RMSEs from forecasting the gains ratio. The largest RMSE is almost 9 percent higher than the smallest RMSE. In addition to greater variation, the relative ranking of the equations changes. The equations using the investment ratio no longer have lower RMSEs than the equations using the gap. However, the change-in-ratio equations continue to have lower RMSEs than the ratio equations (see Table 4-8).

The advantage of measuring RMSEs for forecasts of growth rates is that the errors of the four equations developed in this chapter can be compared to the errors of the methods tested in the previous chapter on the same data. In that chapter, I reported that the two-step method had the lowest RMSE at 22.03 percentage points, followed by CBO's historical forecasts with a RMSE of 23.19 percentage points, and the mean reversion method with a RMSE of 23.98 percentage points.[29] The two change-in-ratio equations developed in this chapter have identical RMSEs that are just below the RMSE of the mean reversion method; the two ratio equations have RMSEs above that of the mean reversion method. The RMSE of the ratio equation using the gap is slightly higher than mean reversion's RMSE, and the RMSE of the ratio equation using the investment ratio is higher yet (see Table 4-8). Thus, the four equations developed here fail to improve on the mean reversion method and two-step model that they were designed to improve.

The annual pattern of errors offers some understanding of why the new equations do not forecast better than the earlier methods. (Of the four new equations, I pick the ratio equation with the gap to compare to the mean reversion and the two-step approaches because it incorporates all of the conceptual improvements.) When the stock market bubble is inflating from 1996 into 2000, the ratio equation is slightly more accurate than mean reversion because its estimated serial correlation coefficient begins rising during the period while the mean reversion method has a fixed 20 percent reversion rate (see Figure 4).[30] But the ratio equation has larger errors forecasting 2001 and 2002. In those years, the ratio equation predicts very little reversion. The mean reversion method, with its reversion rate arbitrarily set at 20 percent, becomes more appropriately specified and does enough better then to offset its larger errors on the upswing.

The two-step model predicts more accurately than mean reversion or the ratio equation primarily because it picks up the tendency for gains to grow fast while the stock market bubble is inflating. It predicts much more accurately in all of the five boom years from 1996-2000, and actually predicts more growth than occurred in two of those years. Offsetting most of that advantage, however, is the tendency to predict too much growth in the other years as well. It predicts too much growth in 1992-1994 as well as in 2001 and too little decline in 2002. As a result, it has larger errors than mean reversion and the ratio equation in most of those years.

Measuring errors in the growth rate of gains instead of the gains ratio makes 1996 nearly as unusual

---

29. The RMSEs are reported first in Table 3-4 and are repeated in Table 4-8 for comparison to the RMSEs of the four equations developed in the current chapter.

30. Numerical values for bars in Figure 4 appear in Appendix Table C-2.

a year as 2001. The percentage decline in gains in 2001 is 46 percent which is essentially the same as the 45 percent increase in 1996 (see Figure 4). Recall from Figure 3 that the decline in the gains ratio between 2000 and 2001 is much larger than the rise from 1995 to 1996. The percentage declines are essentially equal because the prior year is always used as the base for computing growth rates, and the 2000 base is much larger than the 1995 base. Percentage growth rate seems to be an undesirable standard with such large changes, because a hypothetical reversal in 2002 of the decline occurring in 2001 would not appear as the same growth in the opposite direction but as a growth of 84 percent.

The ratio equation using the gap forecasts low growth in both 1996 and 2001, so its error in 1996 is only a little smaller than in 2001. The 1996 error is also larger than the errors in all other years of the boom and bust. Thus, when growth rates are being forecast, the error in 1996 makes nearly as large a contribution to the RMSE as the error in 2001. That is not the case when the gains ratio is being forecast. The similar absolute sizes of the growth rates in 1996 and 2001 probably means that both years have similar effects on equations estimated to explain growth rates, whereas I find that 2001 has a uniquely large impact when the gains ratio is being explained.

Conclusion

The findings of this chapter lead to some conclusions about forecasting with the alternative approaches considered in this paper. They also suggest new approaches.

Choosing Among the Alternatives Tested

The goal of this chapter has been to improve upon the two-step and mean reversion methods. By some measures, that goal has not been reached. Most obviously, the four equations developed do not forecast more accurately over the test period of 1992-2002. In addition, no variables were found to expand upon the specification of the two-step forecasting equation. In fact, only one ratio variable reflecting economic activity could usefully be included at a time. Nothing has been found to replace the stock market variables of CBO's combination equation. Finally, the equations are unstable as new information becomes available during the test period.

The failure to find improved forecasts is disappointing, but the efforts of the entire paper show that CBO's original forecasts could not be improved upon easily. The new equations, in spite of their shortcomings, have achieved some success. As a whole, the new equations forecast as well as the mean reversion method (at a 20 percent rate), and are more credible. The treatment of future tax changes in the new equations is clearly more credible. In addition, the estimated serial correlation coefficient is more credible than the arbitrary 20 percent reversion rate because it is based more completely on historical experience. Two examples illustrate the credibility problems of the arbitrary rate. At the end of 2000, CBO staff decided no reversion in 2001 was more credible than 20 percent and forecast that gains in 2001 would remain at the 2000 level. Thus, CBO abandoned the mean reversion method in the very year that mean reversion achieved its greatest forecasting advantage over the other methods. The other example occurred in the next year when analysts debated using a 40 percent reversion rate instead of a 20 percent rate. Because the 20 percent rate had been chosen arbitrarily, other rates were as plausible.

The failure to achieve lower forecast errors than the two-step method may not be as serious a shortcoming as it appears at first. The two-step specification was selected in part because it could predict gains well during the stock market boom through 2000. That could have led to the specification of an equation that will forecast less accurately in periods of slower growth. The two-step method does forecast less accurately than the other approaches in 1992-1994 and 2001-2002.[31] As a result, the two-step method could have larger errors going forward than it had in the period tested here. The development of the current equations has been less influenced by developments after 1990, and therefore their forecasting errors should be less likely to worsen in future years, as long as the stock market and economic booms of the late 1990s do not return. Finally, the two-step method's finding of a negative impact from real GDP growth is worrisome for future forecasting.

The failure of the new equations to achieve a lower RMSE than CBO's actual forecasts can be explained largely by CBO's incorporation of additional information in its forecasts of 1995 and 1996. Those forecasts included an assumption that taxpayers would defer some realizations from the year ending to the year coming in anticipation of a reduction in the tax rate on gains in the year coming. Similar information could have been added to forecasts made with one of the new equations.

In terms of picking among the four new equations, it seems that none have a clear forecasting advantage. The differences in the test period are small compared to the overall errors of the equations. Alternative test periods are almost certain to be different, and those differences could reverse the ordering of the RMSEs. An alternative way of choosing among the equations would be to select the ones with more credible fits to the historical data. Here, it seems that the ratio equations have an advantage over the change-in-ratio equations because some degree of mean reversion is supported within most periods tested. Furthermore, mean reversion is likely to be important for forecasting several years ahead, and the ratio equations, because they incorporate reversion, could be used to forecasts all ten years of CBO's budget period.

Thinking about New Alternatives

A major drawback to the equations developed in this chapter for forecasting ahead is that their explanation of what causes variation in gains changes over the period tested. All four equations change when the experience of 2001 is included. The estimated impact of the business cycle changes by as much as its standard error and the residuals no longer can be assumed to be normally distributed. Furthermore, the degree to which surprises in gains can be assumed to persist changes during the stock market boom and bust.

The lack of stability and the failure to find additional variables to replace stock prices suggest that the original innovation of dropping stock prices from the forecasting equation was not useful. CBO

31. KMO found that the two-step model forecast tax-adjusted gains more accurately than mean reversion did during 1971-1980, which was a slow-growth decade. The model was fit to that period, and may not do as well in a future period of slow growth.

staff stopped using an equation with stock prices to forecast gains in the winter 1992 baseline because they felt that stock prices could not be forecast accurately one or more years ahead.[32] More recently, KMO agreed that stock prices would not be useful for forecasting because the best year-ahead forecast of stock prices would be a random walk. Instead KMO concentrated on developing specifications with variables that could both predict gains and be forecast themselves.

So far, though, I have been unable to find other variables to replace the role of stock market variables. Furthermore, I have not found stable relationships in equations without stock prices. It may be that inclusion of stock prices is useful for insuring that the effects of other variables are estimated more accurately. Even if stock market variables are always forecast to grow at their average historical rate, their inclusion could allow better estimates of the impacts on gains of the variables that can be forecast.

To see whether inclusion of stock market variables would improve the stability of the estimated equations, I examine CBO's combination equation estimated with data from 1952 through 1990, then through 2000, 2001, and 2002. I also examine the new combination equation developed in Chapter I which uses fixed investment. I find that the estimated coefficients of both combination equations change less than the coefficients of the four equations developed in this chapter. Furthermore, the residuals of the combination equations can be considered normally distributed in all years. Regression statistics for CBO's combination equation are shown in Table 4-9).[33]

Forecasting with an equation that includes stock market prices and trading volume requires that those variables be forecast. KMO observe that stock prices one year ahead tend to follow a random walk with drift, which is to say that the best forecast is the historical average growth rate. Including such a forecast would not harm the forecast of gains and could help. It would not harm the forecast because the constant term in the equation without stock prices would reflect the effect of average stock price growth on the growth of gains. It could help if including stock prices allowed more accurate estimates of the effects of other variables that can be forecast more accurately than stock prices. For example, the coefficients on the gap variable would have been more stable over the last decade if stock market variables had been included in the forecasting equation.

Including stock prices could help the forecasts directly. At the end of each year, many financial analysts forecast stock prices, and some organizations survey those forecasts. The central tendency of those forecasts may contain more information than a random walk forecast of stock prices, just as the average of the forecasts of other economic variables by the "Blue Chip" economists carries some useful information. If so, including stock prices in the equation would help the forecasts directly.

Over long horizons, it seems likely that stock prices will tend to move jointly with real output and

32. The equation CBO used to predict gains in 1992 was not the combination equation.

33. While the combination equations are more stable than the four forecasting equations, they do not explain the decline of gains in 2001 well. Their error in that year causes an increase in the standard errors of the equations and a reduction in the likelihood that all equation errors are normally distributed. See Table 4-9.

price levels, rather than randomly. Stock prices are one measure of the value of the corporate capital stock, and that capital stock, along with labor and the level of knowledge, determine real output. Inflation in prices of output and capital are also likely to move together over long intervals. The tendency for the values of the capital stock and output to move together may explain much of the tendency observed in the historical data for gains to return to its average percentage of output. The likely co-movement, or cointegration, of stock valuations and output should provide some stability to long-run forecasts of stock prices, and keep those forecasts consistent with CBO's forecasts of output.

Unfortunately, the tendency for stock prices to move together with output is not always apparent within a decade. Consequently, forecasts of stock prices over CBO's forecasting horizon cannot be pinned down as well as the forecasts of output. Hence, including stock prices will likely lead to greater variation in CBO's forecasts of gains from one baseline to the next than occurs under mean-reversion forecasting.

Inclusion of stock prices can help with related forecasting tasks. CBO would be able to identify, after the fact, how much of the error in the gains forecast could plausibly be attributable to the error in the forecast of stock prices. In addition, a stock market forecast could help achieve greater internal consistency in forecasts of revenues from alternative sources. Capital gains, withdrawals from tax-deferred retirement accounts, and values of estates subject to estate taxes all depend on stock prices. A third benefit is that historical ranges of variation in stock prices could measure uncertainty in forecasts and help construct alternative revenue paths.

If stock market variables are to be included in a forecasting equation, some modifications to the combination equations should be considered. The work in this chapter has pointed out that such an equation might carry more information if the dependent variable were the gains ratio instead of the change in the logarithm of that ratio. Changes in the stock market variables would be needed as well if they were to enter as ratios. The S&P 500 Index now used could enter as a ratio to potential output or to a more closely related concept like profits of C-corporations. The dollar volume of trades might be converted to a turnover rate by dividing dollar volume by the total value of stocks outstanding in each year.

Other shortcomings of the combination equations could be addressed at the same time. The current equation assumes that a year's appreciation in stock prices is the only appreciation that affects gains in that year. However, appreciation in all years since an asset was purchased adds to the amount of gains held. Also, past realizations and the exclusion from tax of gains held when a person dies subtract from the amount of gains held. A combination of these influences into a measure of outstanding gains might be helpful in explaining extreme changes in gains such as in 2001. Another shortcoming is that the current equation estimates the amount of gains taxpayers will shift into a year when tax rates are temporarily low, but it does not reduce the amount of gains that will be realized in any other year. Allowing some of the gains shifted into one year to reduce gains in adjacent years would be more realistic. A third shortcoming is that the current equations specify that taxpayers respond only to enacted legislation. They probably also respond to presidential and congressional proposals for changes that do not become law. For example, when the Republican Party won control of the House and Senate in 1994 after running on a platform of cutting the capital gains tax, taxpayers may have held off on realizing gains in the end of 1994 and possibly in 1995 as legislation to fulfill the promise was under consideration.

Table 2-1. Explanatory Variables in KMO's Equation	
Abbreviation	Description
DLPRDCTVT	Growth rate of productivity, measured as output per hour worked in non-farm businesses.
DLQGDP	Growth rate of real gross domestic product
DSPREADC	Annual change in the spread between interest rates on Moody's seasoned corporate bonds and the 3-month T-bills in secondary market.
DLQWAGE	Growth rate of real compensation per hour
DLQIFIX	Growth rate of real private fixed investment

Table 2-2. CBO's Combination Equation	
Abbreviation	Description
Dependent variable: DLRATIOFE	Growth rate of the ratio of net positive gains to potential output
Explanatory variables:	
DMTRNEXT	Change in the expected top tax rate on capital gains. Expectation is determined by the tax rate legislated to be in effect in the coming year.
DMTRTRANS	Change in the transitory tax rate on capital gains. Transitory is measured as the difference between the current top rate and MTRNEXT, with a sign-preserving square.
DLGAP	Growth rate of the GDP gap. The GDP gap is measured as the ratio of GDP to potential GDP, and indicates the extend to which the economy is in recession or boom.
DLSP500Q4	Growth rate of the S&P 500 from the 4th quarter of one year to the 4th quarter of the next
DLDOLVOL	Growth rate of dollar volume of share traded on the NYSE, NASDAQ, and AMEX
DLSTARTS	Growth rate of multifamily housing starts, where multifamily includes all starts other than single family.

Table 2-3. Estimation of KMO's Equation

Linear Regression - Estimation by Least Squares
Dependent Variable DLGAIN
Annual Data From 1952 To 2002
Usable Observations 51
Degrees of Freedom 43
Centered R**2 0.5654
R Bar **2 0.4947
Mean of Dependent Variable 0.07218
Std Error of Dependent Variable 0.2575
Standard Error of Estimate 0.1830
Sum of Squared Residuals 1.4407
Regression F(7,43) 7.9932
Significance Level of F 0.0000
Durbin-Watson Statistic 1.8264
Jarque-Bera Statistic on residuals 19.719
Significance level of J-B statistic 0.000

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	0.0320	0.0675	0.4734	0.6383
2. DMTRNEXT	-2.9518	1.1380	-2.5937	0.0129
3. DMTRTRANS	-125.6199	23.4743	-5.3514	0.0000
4. DLPRDCTVT	2.5595	2.5916	0.9876	0.3289
5. DLQGDP	-3.2432	3.0810	-1.0527	0.2984
6. DSPREADC	0.0203	0.0377	0.5392	0.5925
7. DLQWAGE	-0.7653	2.2103	-0.3463	0.7308
8. DLQIFIX	2.5947	0.8628	3.0074	0.0044

Table 2-4. Estimation of CBO's Combination Equation

Linear Regression - Estimation by Least Squares
Dependent Variable DLRATIOFE
Annual Data From 1952 To 2002
Usable Observations 51
Degrees of Freedom 44
Centered R**2 0.8220
R Bar **2 0.7977
Mean of Dependent Variable 0.0041
Std Error of Dependent Variable 0.2600
Standard Error of Estimate 0.1170
Sum of Squared Residuals 0.6019
Regression F(6,44) 33.8637
Significance Level of F 0.0000
Durbin-Watson Statistic 2.0695
Jarque-Bera Statistic on residuals 3.969
Significance level of J-B statistic 0.137

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	-0.0733	0.0188	-3.8932	0.0003
2. DMTRNEXT	-2.4376	0.6844	-3.5615	0.0009
3. DMTRTRANS	-116.9045	13.8256	-8.4557	0.0000
4. DLGAP	2.7489	0.8639	3.1821	0.0027
5. DLSP500Q4	0.5208	0.1621	3.2142	0.0025
6. DLDOLVOL	0.2986	0.1011	2.9536	0.0050
7. DLSTARTS	0.1456	0.0808	1.8012	0.0785

Table 2-5. Estimation of New Combination Equation

Linear Regression - Estimation by Least Squares
Dependent Variable: DLRATIOFE
Annual Data From 1952:01 To 2002:01
Usable Observations 51
Degrees of Freedom 45
Centered R**2 0.8452
R Bar **2 0.8280
Mean of Dependent Variable 0.0041
Std Error of Dependent Variable 0.2600
Standard Error of Estimate 0.1078
Sum of Squared Residuals 0.5233
Regression F(5,45) 49.1516
Significance Level of F 0.0000
Durbin-Watson Statistic 2.1814

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	-0.0757	0.0173	-4.3804	0.0001
2. DMTRNEXT	-2.0422	0.6333	-3.2248	0.0023
3. DMTRTRANS	-113.427	12.7438	-8.9005	0.0000
4. DLSP500Q4	0.5711	0.1492	3.8285	0.0004
5. DLDOLVOL	0.2825	0.0933	3.0272	0.0041
6. DLIFIXFE	1.4289	0.2529	5.6496	0.0000

Table 2-6. Prediction Errors with Revised Data (percentage points)

Years	Root Mean Squared Errors CBO Combo	KMO	New Combo	New's Accuracy Rate
1971-1980	22.72	11.93	19.67	70
1981-1990	20.69	39.43	19.77	70
1991-2002	11.84	22.85	9.93	58
1971-2002	18.64	26.95	16.73	66

File: current year\forecasts6.xls

Notes: Data as available in November 2003.

New's Accuracy Rate is the percent of times the new combination equation predicts gains more accurately than the CBO combination.

Table 2-7. Errors Predicting Explanatory Variables in year ending with Real-Time Data

Date Predicted	Year Predicted	DLSTARTS	DLGAP	DLIFIXFE
Dec 16, 1991	1991	-0.0168	0.0024	0.0164
Dec 7, 1992	1992	-0.0423	0.0042	0.0186
Dec 8, 1993	1993	-0.0009	-0.0086	-0.0079
Dec 8, 1994	1994	0.0448	-0.0041	-0.0251
Feb 20, 1996	1995	-0.0378	-0.0019	-0.0093
Dec 12, 1996	1996	0.0287	0.0048	0.0128
Dec 9, 1997	1997	0.0403	-0.0021	0.0062
Dec 7, 1998	1998	-0.0063	-0.0005	0.0018
Dec 29, 1999	1999	0.0200	-0.0014	-0.0111
Jan 10, 2001	2000	-0.0380	-0.0118	-0.0314
Jan 9, 2002	2001	0.0243	-0.0082	-0.0121
Jan 14, 2003	2002	0.0163	-0.0023	0.0008
RMSE	1991-2002	0.0293	0.0055	0.0154

Note: Variables are measured as the annual change in the logarithm of their actual value because that is the value used in CBO's and the new combination equations. Errors are measured as difference between predicted amounts from real-time data and revised amounts in November 2003 data. Errors in percentage points can be approximated by multiplying reported errors by 100.

Table 2-8. Errors Predicting the Growth Rate of Capital Gains (in percentage points)

	November 2003 Data		Real-time Data	
Year	CBO	New	CBO	New
1991	-6.47	-10.61	-5.93	-7.90
1992	4.98	3.42	5.81	7.02
1993	10.10	1.79	7.68	0.34
1994	-14.54	-7.20	-14.58	-11.19
1995	-5.57	-9.92	-5.69	-10.74
1996	24.15	20.68	26.93	23.43
1997	-6.77	-8.68	-4.52	-6.00
1998	0.58	-5.51	1.94	-3.57
1999	-0.87	-4.74	0.66	-5.09
2000	3.80	3.59	-0.54	-1.96
2001	-23.49	-18.30	-24.28	-19.33
2002	-8.58	1.09	-8.09	1.63
RMSE	11.84	9.93	12.17	10.63

Table 3-1. RMSEs from year ahead forecasts of tax-adjusted gains (in percentage points)

Method	1971-80	1981-90	1991-00	1971-00	2001	1971-01
Two-step						
KMO	15.47	15.76	13.03	14.80	58.82[a]	[b]
Replication	15.88	15.83	12.94	14.95	56.77	17.89
Mean Reversion						
KMO	21.13	13.28	20.29	18.57	46.50[a]	[b]
Replication	21.14	13.28	20.41	18.62	46.37	20.12
Replication (20% reversion)	20.34	13.30	19.76	18.08	42.49	19.36
Two-step less mean reversion						
KMO	-5.66	2.48	-7.26	-3.77	12.37[a]	[b]
Replication	-5.25	2.55	-7.47	-3.67	10.40	-2.22
Addendum on growth rate of tax-adjusted capital gains						
Mean	13.69	6.42	16.50	12.21	-45.75	10.34
Standard deviation	18.18	14.80	13.38	16.15	NA	18.90

Notes: KMO refers to results in Jan Kim, Preston Miller, and Larry Ozanne, *Estimating and Forecasting Capital Gains with Quarterly Models,* CBO Technical Paper 2004-14.

[a] KMO errors for 2001 are computed from growth rates in their Table 8.

[b] KMO did not report RMSEs for 1971-2001.

NA Standard deviation is not informative for a single year.

Source: Forecasts RMSEs.xls

Table 3-2: Regression of the Growth Rate of Gains on Actual and Forecast Growth Rates of Explanatory Variables: Annual Data From 1960 to 2000

	Actual Explanatory Variables and gains	Forecasts of explanatory variables in all years and predicted gains in 2000
Dependent Variable	DLG	DLGF
R-square	0.3821	0.2267
Mean: dep. var.	0.0926	0.0905
SE of dep var	0.1645	0.1643
SE of estimate	0.1363	0.1523
SS Residual	0.6692	0.8348
Reg. F(4,36)	5.5650	2.6389
Sig. Level of F	0.0014	0.0497
Durbin-Watson	2.1782	1.8698

variable	coefficient	variable	coefficient
constant	0.050	constant F.	0.184**
(SE)	(0.054)	(SE)	(0.077)
DLP	3.815*	DLPF	3.792
(SE)	(2.035)	(SE)	(4.371)
DLO	-3.125	DLOF	-9.214**
(SE)	(2.521)	(SE)	(3.753)
DR	13.780	DRF	18.663
(SE)	(8.429)	(SE)	(15.312)
DLI	1.0949	DLIF	3.221**
(SE)	(0.870)	(SE)	(1.532)

Notes: *, **, *** indicates 10%, 5%, and 1% significance level respectively.
DLG is change in logarithm of tax-adjusted gains.
DLGF is DLG with a predicted value of gains for year 2000.
DLP is change in logarithm of productivity in nonfarm business sector
DLO is the change in the logarithm of real GDP
DR is the change in the ratio of real consumption of durables to real disposable income.
DLI is the change in the logarithm of real fixed investment.
The letter F added to the end of the explanatory variables indicates a forecast of the variable.

Table 3-3: RMSEs from forecasting macroeconomic variables by KMO and by CBO in year ahead

	KMO			CBO
Variables	1960-01	1960-91	1992-01	1992-01
Growth rate of productivity	0.0124	0.0134	0.0088	0.0115
Growth rate of real GDP	0.0153	0.0163	0.0111	0.0143
Change in share of income spent on durables	0.0033	0.0034	0.0032	0.0042
Growth rate of fixed investment	0.0413	0.0439	0.0318	0.0406

Notes: Growth rates computed as year to year change in logarithms.

Income spent on durables measured as real durable spending over real disposable income.
Errors are measured relative to published data as of November 2003.
Source: forecasts RMSEs.xls

Table 3-4: RMSEs of year-ahead forecasts of growth of tax-adjusted gains using fixed-time or actual gains using real-time data (in percentage points)

	1992-01		1992-02
	fixed	real	real
Two-step model	21.75	22.24	22.03
Mean reversion (20%)	23.44	24.43	23.98
CBO baselines		23.61	23.19

Notes: Fixed-time data includes KMO's forecast of macroeconomic variables based on data in early 2002 and my data as of November 2003.
RMSEs for real-time data are computed using final release of gains in 2002 at $268 billion. The preliminary value of $254 billion was used in Chapter 2.
Source: forecasts RMSEs.xls

Table 4-1: Combination Equation and Simpler Equations For Forecasting

Combination Equation

Δlog(gains/potential) =
Constant#1
B1 * Δ(permanent tax rate)
B2 * Δ(transitory tax rate)
B3 * Δlog(GDP/potential)
B4 * Δlog(SP500)
B5 * Δlog($ volume of stock trades)
B6 * Δlog(multifamily housing starts)
Error term #1

Simplified Change-in-Ratio Equation:

Δ(gains/potential) =
Constant #2
C1 * Δ(permanent tax rate)
C2 * Δ(transitory tax rate)
C3 * Δ(GDP/potential)
Error term #2

Simplified Ratio Equation:

(gains/potential) =
Constant #3
C1 * (permanent tax rate)
C2 * (transitory tax rate)
C3 * (GDP/potential)
Error term #3

Table 4-2. Equations Explaining Change in the Gains Ratio with Changes in Tax Rates and Change in Either the Gap or Investment Ratio. Estimated from 1952-1990.

	Gap	Investment Ratio
Regression Statistics		
R-squared	0.805	0.797
Mean of dependent variable	0.000	0.000
Standard error of dep. var.	0.010	0.010
Standard error of estimate	0.005	0.005
Sum of squared residuals	0.001	0.001
Regression F(6,32)	48.010	45.839
Significance level of F	0.000	0.000
Durbin-Watson statistic	2.158	2.077
Jarque-Bera statistic on residuals	1.797	0.781
Significance level of J-B	0.407	0.677
Variables		
Constant	0.000	0.000
	(0.001)	(0.001)
DMTRNEXT	-0.083***	-0.070**
	(0.290)	(0.030)
DMTRTRANS	-6.533***	-6.323***
	(0.555)	(0.561)
DGAP in column 1	0.106***	0.218***
DIFIXFE in column 2	(0.034)	(0.077)

Notes: *, **, and *** indicate probabilities that coefficients are zero of less than 10%, 5%, and 1%. Standard errors are shown in parentheses below the coefficient estimates. Appendix Table C-3 defines variable names in this table.

Table 4-3. Equations Explaining Gains Ratio with Tax Rates and Either Gap or Investment Ratio. Estimated from 1952 through 1990

	Gap	Investment Ratio
Regression Statistics		
R-squared	0.831	0.830
Mean of dependent variable	0.277	0.277
Standard error of dep. var.	0.010	0.010
Standard error of estimate	0.004	0.004
Sum of squared residuals	0.001	0.001
Durbin-Watson statistic	2.003	1.856
Q-Statistic	12.806	10.911
Significance of Q-Statistic	0.119	0.207
Jarque-Bera statistic on residuals	1.314	1.033
Significance level of J-B	0.519	0.597
Variables		
Constant	-0.056	0.010
	(0.035)	(0.014)
MTRNEXT	-0.076***	-0.063**
	(0.027)	(0.026)
MTRTRANS	-6.529***	-6.342***
	(0.576)	(0.588)
GAP in column 1	0.105***	0.223***
IFIXFE in column 2	(0.035)	(0.079)
Serial Correlation Coefficient	0.767***	0.692***
	(0.101)	(0.132)

Notes: *, **, and *** indicate probabilities that coefficients are zero of less than 10%, 5%, and 1%.
Standard errors are shown in parentheses below the coefficient estimates. Appendix Table C-3 defines variable names in this table.

Table 4-4. Serial Correlation Coefficients and Standard Errors from Equation Explaining Gains Ratio in terms of gap and tax rates. Estimated from 1952 through year indicated.

Last Year of Data	Serial Correlation Coefficient	Standard Error
1990	0.767	0.101
1991	0.754	0.104
1992	0.755	0.102
1993	0.753	0.101
1994	0.697	0.116
1995	0.743	0.102
1996	0.737	0.108
1997	0.792	0.105
1998	0.830	0.097
1999	0.916	0.093
2000	0.951	0.083
2001	0.736	0.095

Table 4-5. Equation explaining ratio of gains to potential GDP with the gap and Tax Rates. Estimated from 1952 through year indicated

	1990	2000	2001
Regression Statistics			
R-squared	0.831	0.868	0.766
Mean of dependent variable	0.277	0.289	0.029
Standard error of dep. variable	0.010	0.012	0.012
Standard error of estimate	0.004	0.005	0.006
Sum of squared residuals	0.001	0.001	0.002
Durbin-Watson statistic	2.003	1.847	1.664
Q-Statistic	12.806	10.910	5.728
Significance of Q-Statistic	0.119	0.451	0.891
Jarque-Bera statistic on residuals	1.314	2.301	38.699
Significance level of J-B	0.519	0.317	0.000
Variables			
Constant	-0.056	-0.024	-0.085*
	(0.035)	(0.036)	(0.047)
MTRNEXT	-0.076***	-0.092***	-0.089**
	(0.027)	(0.028)	(0.035)
MTRTRANS	-6.529***	-6.358***	-6.401***
	(0.576)	(0.557)	(0.814)
GAP	0.105***	0.094***	0.139***
	(0.035)	(0.033)	(0.046)
Serial Correlation Coefficient	0.767***	0.951***	0.736***
	(0.101)	(0.083)	(0.095)

Notes: *, **, and *** indicate probabilities that coefficients are zero of less than 10%, 5%, and 1%.
Standard errors are shown in parentheses below the coefficient estimates. Appendix Table C-3 defines variable names in this table.

Table 4-6. Equation explaining change in the gains ratio with changes in the gap and tax rates. Estimated from 1952 through year indicated.

	1990	2000	2001
Regression Statistics			
R-squared	0.805	0.764	0.619
Mean of dependent variable	0.000	0.001	0.000
Standard error of dep. variable	0.010	0.009	0.010
Standard error of estimate	0.005	0.005	0.007
Sum of squared residuals	0.001	0.001	0.002
Regression F(6,32)	48.010	45.516	24.957
Significance level of F	0.000	0.000	0.000
Durbin-Watson statistic	2.158	1.907	1.696
Jarque-Bera statistic on residuals	1.797	2.062	289.647
Significance level of J-B	0.407	0.357	0.000
Variables			
Constant	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)
DMTRNEXT	-0.083	-0.089***	-0.088**
	(0.290)	(0.027)	(0.039)
DMTRTRANS	-6.533	-6.336***	-6.352***
	(0.555)	(0.540)	(0.781)
DGAP	0.106	0.092***	0.136***
	(0.034)	(0.032)	(0.045)

Notes. *, **, and *** indicate probabilities that coefficients are zero of less than 10%, 5%, and 1%.
Standard errors are shown in parentheses below the coefficient estimates. Appendix Table C-3 defines variable names in this table.

Table 4-7. Root Mean Squared Errors from Forecasting the Gains Ratio in 1992-2002 with Four Alternative Equations. (Errors are from gains ratios measured in decimals.)

Ratio equation with fixed investment	0.01268
Ratio equation with gap	0.01281
Change in ratio equation with fixed investment	0.01263
Change in ratio equation with gap	0.01275

Note: Errors in 1997 used in the RMSEs are measured relative to an estimate of what gains would have been without the tax change in 1997.

Table 4-8. Root Mean Squared Errors from Forecasting the Growth Rate of Gain in 1992-2002 with Seven Alternative Methods. (Errors are from growth rates measured in percentages.)

Ratio equation with Gap	24.81
Ratio equation with Investment	25.52
Change in ratio equation with Change in Gap	23.49
Change in ratio equation with Change in Investment	23.49
Two-Step Approach	22.03
CBO Baselines	23.19
Mean Reversion Method with 20% reversion rate	23.98

Note. Errors in 1997 used in the RMSEs are relative to an estimate of what gains would have been without the tax change in 1997

Table 4-9. CBO Combination Equation Estimated from 1952 through Year Indicated

	1990	2000	2001	2002
Regression Statistics				
R-squared	0.841	0.826	0.818	0.822
Mean of dependent variable	0.001	0.024	0.011	0.005
Standard error of dep. variable	0.256	0.241	0.258	0.259
Standard error of estimate	0.111	0.107	0.117	0.116
Sum of squared residuals	0.396	0.483	0.592	0.595
Regression F(6,32)	28.310	33.282	32.183	33.931
Significance level of F	0.000	0.000	0.000	0.000
Durbin-Watson statistic	1.930	2.074	2.013	2.114
Jarque-Bera statistic on residuals	0.815	0.696	3.262	4.021
Significance level of J-B	0.665	0.706	0.196	0.134
Variables				
Constant	-0.051**	-0.053***	-0.069***	-0.072***
	(0.202)	(0.019)	(0.019)	(0.019)
DMTRNEXT	-2.734***	-2.519***	-2.445***	-2.440***
	(0.696)	(0.628)	(0.687)	(0.681)
DMTRTRANS	-120.9***	-117.4***	-117.2***	-117.0***
	(13.39)	(12.68)	(13.88)	(13.75)
DLGAP	2.612**	2.166**	2.700***	2.728***
	(0.889)	(0.815)	(0.868)	(0.858)
DLSP500Q5	0.464**	0.462***	0.491***	0.508***
	(0.176)	(0.153)	(0.166)	(0.161)
DLDOLVOL	0.233**	0.241**	0.300***	0.299***
	(0.107)	(0.094)	(0.101)	(0.101)
DLSTARTS	0.179*	0.191**	0.153*	0.149*
	(0.090)	(0.076)	(0.082)	(0.080)

Notes: *, **, and *** indicate probabilities that coefficients are zero of less than 10%, 5%, and 1%. Standard Errors are in parentheses.
Appendix Table C-3 and Table 2-2 define variable names in this table.
Regression statistics here differ slightly from those presented in Table 2-4 because that equations is estimated with the preliminary value of gains for 2002 of around $254 billion, whereas the one here is estimated with the final value of around $269 billion.

Figure 1: Ratio of Gains to GDP

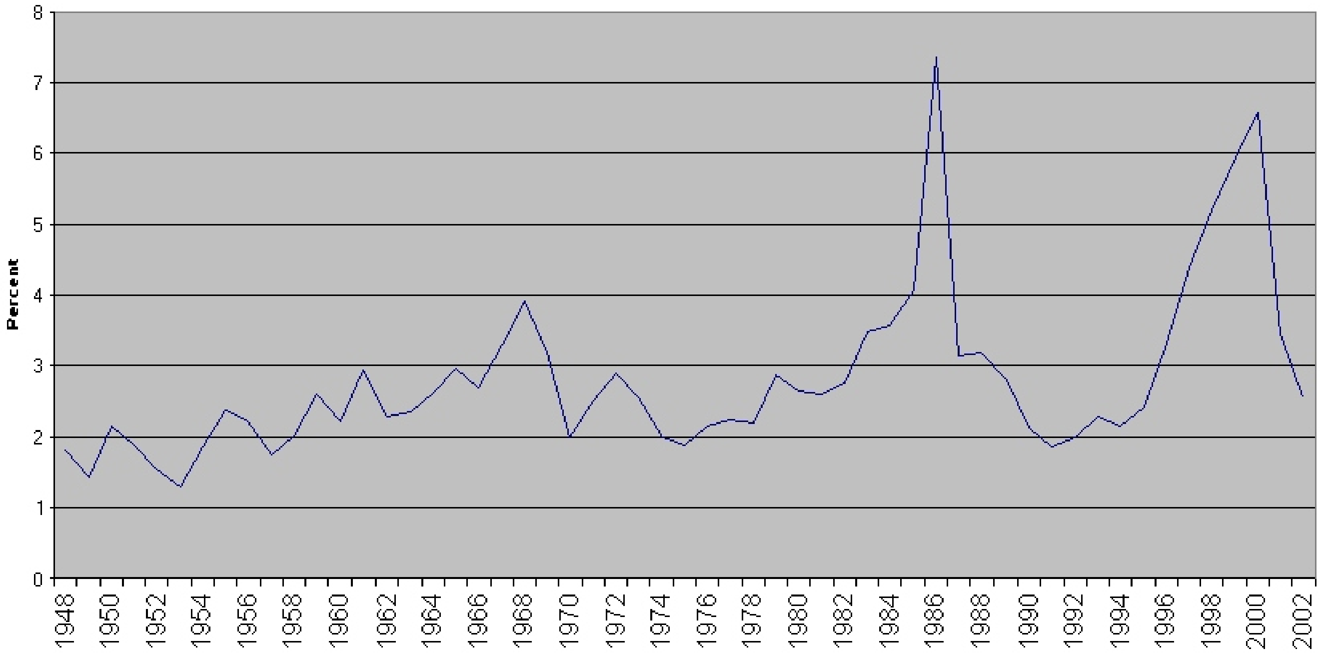


Figure 2: Errors Forecasting Growth Rate of Gains with Two-Step and Mean Reversion Methods (mean reversion uses 20% reversion rate, both methods use real-time data)

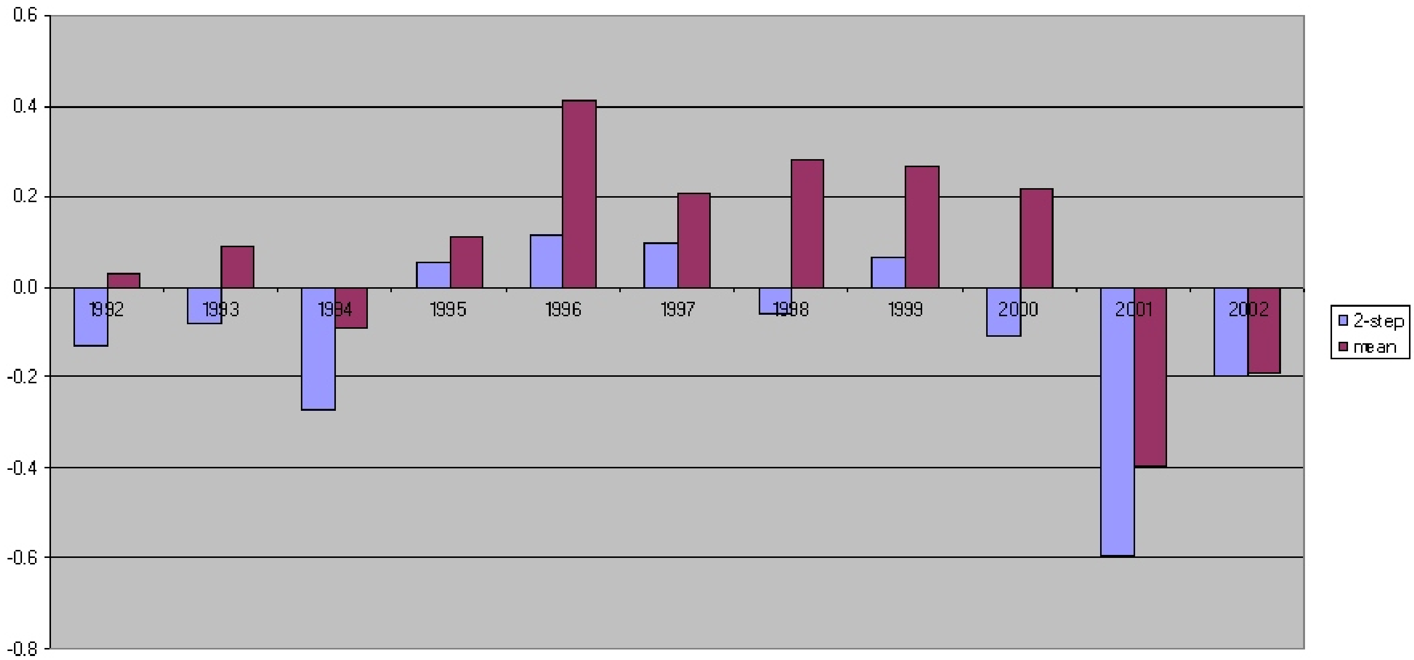


Figure 3: Actual and Forecast Gains Ratios. Forecasts are from ratio and change-in-ratio equations with gap variable.

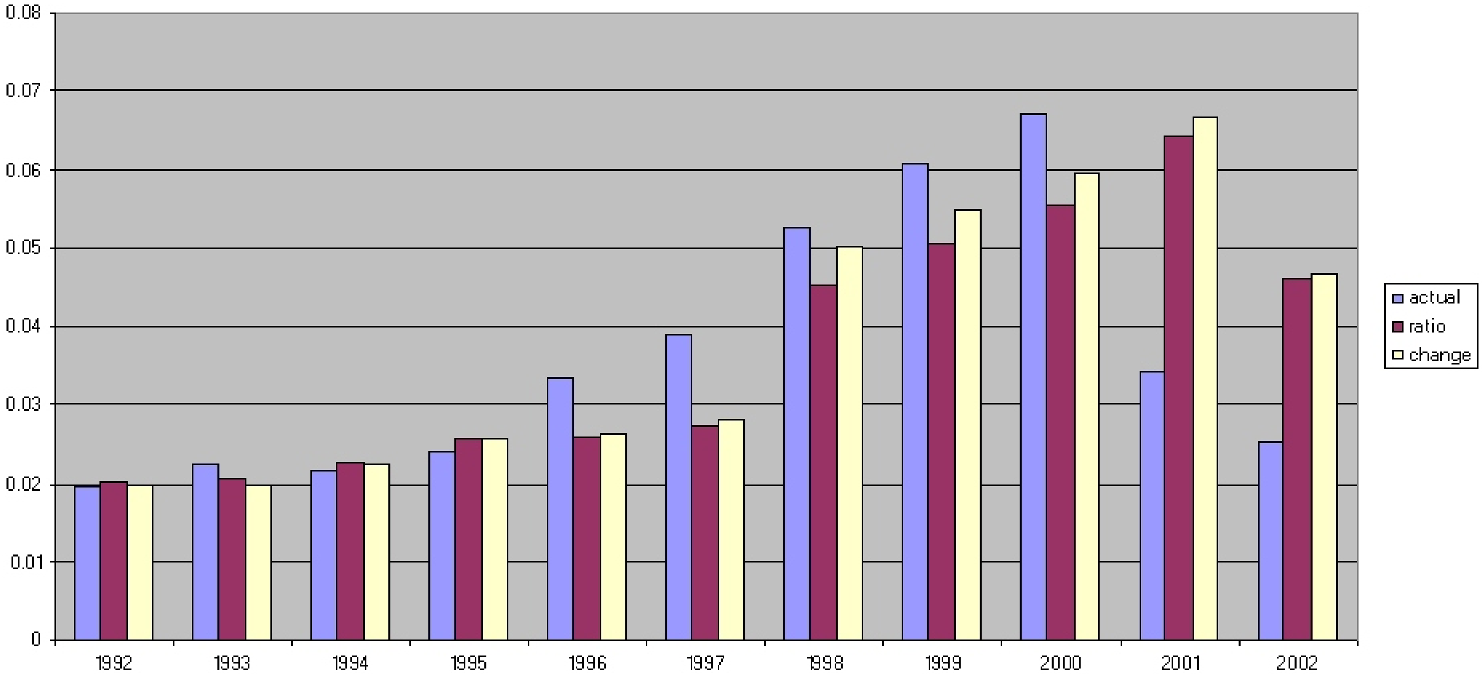
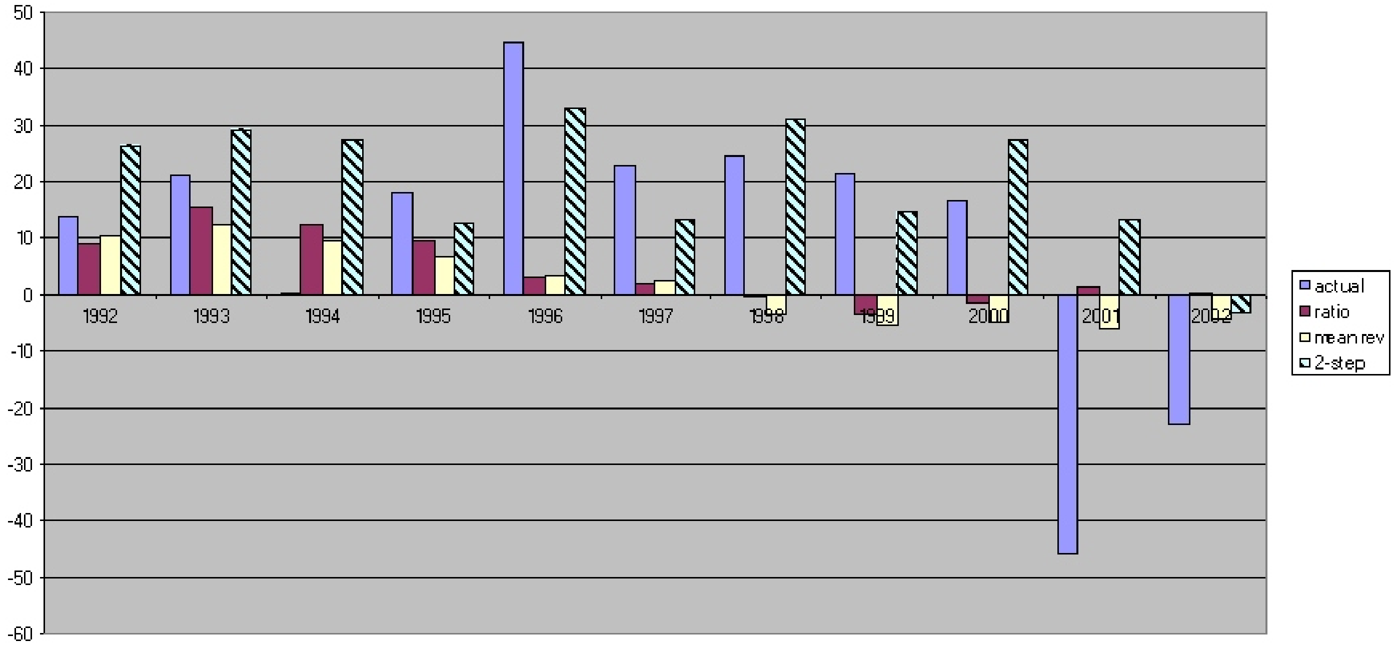


Figure 4: Actual (ex 1997) and Forecasted Growth Rates of Gains. Forecasts made by ratio equation, mean reversion, and 2-step approach. (Growth rates in percents)



APPENDIX A: Additional Results for Equations Predicting Gains

Three additional comparisons of real-time equations are informative. Two of them involve alternative specifications of equations, and one compares equation forecasts to CBO's actual predictions from 1991 through 2002.

Annual vs. Quarterly Stock Prices

CBO must predict the December values of the S&P 500 and the dollar volume of shares traded before predicting gains for the year ending. CBO predicts those December values by using the historical average monthly growth rate. As both series are highly volatile, the predictions are frequently off.

When reconstructing stock prices and dollar volume for the real-time analysis, I noticed that errors in the December S&P caused bigger errors in the fourth quarter average S&P than errors in the December dollar volume caused in the annual average dollar volume. The primary reason is that errors in December cause larger errors in a three month average than in a 12 month average. That observation led me to wonder whether the annual growth rate of the S&P 500 would predict gains better in real time than does the fourth quarter average.

Miller and Ozanne settled on the fourth-quarter rather than the annual average based on the higher R-squared statistic obtained with regressions fit over the 1952-1998 years, using known stock prices and dollar volumes in every year. Subsequent estimates of the CBO combination equation between 1952 and 2002 with one and then the other stock price measure continue to show a slight advantage for measuring stock price growth from fourth quarter to fourth quarter.

In this appendix, I verify that having to approximate December stock prices does lead to larger errors in the fourth quarter growth rate than the annual average growth rate, at least for 1991-2002. I then substitute the annual growth rate for the quarterly growth rate in CBO's combination equation and find mixed success. The RMSE is still lower with the quarterly S&P growth rate, but the average absolute error is smaller with the annual average S&P growth rate. Substituting the annual S&P growth rate into the new combination leaves the fourth quarter growth rate with a smaller error on both measures. See Appendix Table A-1.

The differences between using the fourth quarter and annual growth rates are small. Using partial data for the month of December might improve the quality of the quarterly estimate enough to retain with real-time data the advantage it has with full information.

The CBO Combination and Its Four Predecessors

Miller and Ozanne found that the average prediction of their four equations had a lower RMSE than did any one of them. We subsequently reasoned that an equation combining the explanatory variables of the four equations should predict at least as accurately as averaging. Using real-time data, I find that the average prediction of MO's four equations has a lower RMSE than any of the individual equations, as MO found for their shorter sample. In addition, I find that CBO's combination equation has a lower RMSE than the simple average of the four equations does. The combination also has the lowest average absolute error of any of the four individual equations. Finally, it has a lower absolute error than the average prediction over the 1991-2002 period but not over the shorter period 1991-2001. In the shorter period, the average absolute error obtained by averaging is slightly lower than that from the CBO combination.

The four Miller and Ozanne equations are identified here as MO3, MO4, MO3v, and MO4v. All equations include the tax rates and GDP gap variables specified in Table 2, as well as a constant term. In addition, equations numbered 4 include housing starts, and equations with the letter v substitute the growth of dollar volume for the growth of the S&P 500. See Appendix Table A-2 for RMSEs.

CBO's Actual Predictions

CBO used the average of MO's four equations for its forecasts of 2000, 2001 and 2002. Consequently, its errors in those years are similar to those found with real-time data using CBO's combination equation because it uses the same variables. In earlier years, CBO used equations with alternative measures of stock prices, GDP, and tax rates. It also included an inflation rate and, in some versions, a mean reversion term. Finally, the predictions of the equations were modified in some years by judgments that the equations omitted unique influences, such as anticipated tax changes.

Some of those differences between CBO's historical methods and the current equations appear to be important. CBO's predictions of the growth rate have a RMSE of 15.86 percentage points between 1991 and 2002. All of the alternatives tested in this appendix have lower RMSEs. The next highest is 14.09 percentage points for MO4v and the lowest is 10.63 percentage points for the new combination. CBO's actual forecast errors remain higher if 2002 is omitted or the average absolute error is compared. See Appendix Tables A-1, A-2, and A-3.

The advantage of the new equations will not necessarily hold in future years. Obviously the future frequently differs from the past. An example has been provided by the different successes of KMO's equation and CBO's combination equation between 1971-1980 and 1991-2002 (Table 6). A key difference between the equations CBO used in the 1990s and the ones tested here is that the ones tested here were built in part on what variables could explain the 1991-2002 period. Thus, the newer equations have gained from the experience of 1991-2002. If the future differs from the conditions of that period, the new equations could lose their advantage over CBO's older equations.

Table A-1. Errors with Quarterly and Annual S&P 500 (in percentage points)

Error Type	Period	CBO Combo	New Combo	CBO Combo-A	New Combo-A
Annual	1991	-5.93	-7.90	-2.17	-6.35
Annual	1992	5.81	7.02	5.32	6.15
Annual	1993	7.68	0.34	6.69	-0.64
Annual	1994	-14.58	-11.19	-16.20	-10.93
Annual	1995	-5.69	-10.74	-1.45	-3.92
Annual	1996	26.93	23.43	26.75	24.08
Annual	1997	-4.52	-6.00	-1.84	-2.98
Annual	1998	1.94	-3.57	0.38	-4.60
Annual	1999	0.66	-5.09	0.32	-5.53
Annual	2000	-0.54	-1.96	-6.65	-10.97
Annual	2001	-24.28	-19.33	-25.78	-20.96
Annual	2002	-8.09	1.63	-10.38	-2.32
RMSE	1991-01	12.48	11.09	12.69	11.34
RMSE	1991-02	12.17	10.63	12.51	10.87
Avg. absolute	1991-01	8.96	8.78	8.50	8.83
Avg. absolute	1991-02	8.89	8.18	8.66	8.29

Notes: Combo-A equations substitute growth in the annual S&P 500 for fourth quarter growth. Specifications of equations are otherwise as specified in body of paper.
From "fcsts errors.xls"

Table A-2. Prediction Errors in MO and CBO Combination Equations

Error Type	Period	MO3	MO4	MO3v	MO4v	Avg	Combo
Annual	1991	-15.40	-5.20	-10.12	-1.01	-7.93	-5.93
Annual	1992	6.71	6.51	5.46	5.47	6.04	5.81
Annual	1993	9.78	12.64	0.18	4.58	6.79	7.68
Annual	1994	-5.66	-15.06	-8.83	-17.84	-11.85	-14.58
Annual	1995	-4.77	-5.90	0.56	-1.01	-2.78	-5.69
Annual	1996	28.12	27.43	30.05	29.25	28.71	26.93
Annual	1997	-9.97	-6.76	2.02	3.73	-2.75	-4.52
Annual	1998	0.70	1.85	3.07	3.97	2.40	1.94
Annual	1999	1.48	4.53	-2.95	0.76	0.96	0.66
Annual	2000	10.49	10.43	-15.84	-12.61	-1.88	-0.54
Annual	2001	-24.90	-27.77	-24.26	-27.29	-26.06	-24.28
Annual	2002	-5.30	-7.55	-12.96	-14.85	-10.16	-8.09
RMSE	1991-01	13.67	14.12	13.40	14.02	12.84	12.48
RMSE	1991-02	13.18	13.69	13.36	14.09	12.64	12.17
Avg. absolute	1991-01	10.73	11.28	9.39	9.77	8.92	8.96
Avg. absolute	1991-02	10.27	10.97	9.69	10.20	9.03	8.89

Notes: See this paper for specification of combo equation. See appendix text and or Miller and Ozanne for specification of MO equations. Avg is average of errors in MO equations.

Table A-3. Errors for Historical CBO Predictions and Current Equations (in percentage points)

Error Type	Period	CBO	CBO Combo	New Combo
Annual	1991	-19.85	-5.93	-7.90
Annual	1992	-8.69	5.81	7.02
Annual	1993	14.11	7.68	0.34
Annual	1994	-3.14	-14.58	-11.19
Annual	1995	3.01	-5.69	-10.74
Annual	1996	35.84	26.93	23.43
Annual	1997	-6.43	-4.52	-6.00
Annual	1998	11.81	1.94	-3.57
Annual	1999	7.75	0.66	-5.09
Annual	2000	-0.89	-0.54	-1.96
Annual	2001	-26.28	-24.28	-19.33
Annual	2002	-10.69	-8.09	1.63
RMSE	1991-01	16.25	12.48	11.09
RMSE	1991-02	15.86	12.17	10.63
Avg. absolute	1991-01	12.53	8.96	8.78
Avg. absolute	1991-02	12.37	8.89	8.18

Notes:

CBO refers to CBO's historical predictions for growth rate of gains.
Other equations are as specified in this paper.

From "fcsts errors.xls", and "CGBASELINE.xls"

APPENDIX B: Explaining the Estimated Negative Effect of Real GDP on Capital Gains[34]

Capital gains have tended to follow the upward trend of GDP since 1950. Gains fluctuate much more than GDP from year to year, but some of that fluctuation also follows the business cycle. Consequently, GDP has often been used as a variable to explain the path of capital gains, and its estimated coefficient is normally statistically greater than zero.

Estimates of the equation developed by KMO for forecasting gains, however, produce a coefficient that is statistically less than zero (Table 3-2). In KMO's specification, the dependent variable is the growth rate of tax-adjusted capital gains, and the explanatory variables are forecasts of four variables. Three of them are the growth rates of real GDP, real fixed investment, and productivity. The fourth is the change in the ratio of real consumption of consumer durables to real disposable income. The equation is estimated with annual data from 1960 to 2000.

Using forecasts instead of actual data for the explanatory variables accentuates the negativeness of the estimated coefficient on real GDP. When actual data are substituted for the forecasts, the coefficient of real GDP remains negative but becomes smaller and statistically insignificant.

Including the three other explanatory variables also obscures the normal relationship of GDP to gains. When the growth rate of real GDP, measured with actual data, is the only variable used to explain the growth rate of tax-adjusted gains, its coefficient is positive and statistically significant at the 5 percent level. When any of the other three variables is included in the equation, though, the coefficient of real GDP growth becomes statistically insignificant. When only the growth rates of real GDP and real fixed investment are included, the coefficient of real GDP becomes negative. When all three are included, the coefficient falls farther below zero, but still is not statistically different from zero, as shown in Table 3-2.

Thus, the negative coefficient on GDP depends on the inclusion of the three other variables. All of them contain components of GDP, and therefore are correlated with GDP. When included in the equations, their coefficients are positive (though mostly not significant at normal levels), suggesting that they are picking up the normal positive relationship between GDP and gains. In fact, they may be the source of the positive relationship between GDP and gains.

The negative coefficient of GDP when those other variables are included could indicate that other components of GDP reduce the amount of gains realized. Furthermore, the significance of real GDP's negative coefficient in the equation with forecast values could indicate that this component is forecast relatively more accurately than the components with positive impacts on gains.

The omitted components of GDP are consumption of nondurables and services, changes in private inventories, net exports, and government consumption and investment. Preston Miller pointed out that the change in inventories is a leading indicator of the business cycle, accelerating as sales turn down and then decelerating when sales turn up. For this reason, he suggested that changes in inventories could be negatively related to gains. Net exports might also be negatively related to

34. Diane Nhu Pham assisted with this appendix.

gains if higher net exports reflect a low value of the dollar and weak demand by foreigners for assets in the U.S. On the other hand, consumption of nondurables and services is less volatile than other components over the business cycle, and therefore seems likely to reflect the long-run trend in GDP better than other components. Thus, these components of consumption should be positively related to gains. Finally, government spending might be somewhat countercyclical, growing faster in recessions than in recoveries, which would could give it a negative coefficient. At the same time, much government growth from 1960 to 2000 was related to changes defense and entitlement spending which could well be unrelated to gains.

To test whether any of those four components accounts for GDP's negative coefficient, I switch to data that break the growth rate of real GDP into contributions from its major components—Table 1.1.2 from the National Income and Product Accounts. My first step is to approximate the specification giving GDP its negative coefficient. Annual growth in tax-adjusted gains is the dependent variable, and the explanatory variables are the growth rate of productivity, the growth rate of real GDP, and the contributions to the growth rate of real GDP from fixed investment and consumption of durables. All growth rates are measured as the normal percentage change from the previous year, in keeping with the variables in Table 1.1.2. (In equations of Table 3-2, growth rates of all variables are measured by the annual change in the logarithm.)

The equation, estimated from 1960 through 2000, is grossly consistent with the equation estimated with actual data in Table 3-2. It finds the negative and statistically insignificant coefficient on real GDP, and the other variables all have positive coefficients (see column a of Table B-1). The coefficients on the growth rate of real GDP (abbreviated as GDP) and productivity (DPL1) are within standard confidence intervals of their values in Table 3-2. The coefficients on the contributions of fixed investment (Fixi) and consumption of durables (DG) are positive, but their magnitudes are not comparable to those of Table 3-2. Real fixed investment is measured as a contribution to real GDP growth in Table B-1, and a growth rate in Table 3-2, and durables is a contribution to growth in Table B-1and a change in a ratio in Table 3-2.

To determine why GDP has a negative coefficient in column a, I substitute for GDP its components other than durable goods and fixed investment: the sum of non-durable goods and services (Sumnds), net exports (NE), government spending (G), and the change in inventories (CIPI). These results are shown in Table B-1, columns b though e. The coefficients of those variables are estimated to be negative. Because each one contributes to real GDP growth, the negative signs imply they all contribute to the negative relationship between GDP and gains. None of the variables are statistically significant on its own, but the change in inventories and net exports have t-statistics greater than one. When the contributions of the four variables are summed into a single variable (Sumof4), the group has a negative effect that is significant at the 10 percent level (Table B-1, column f). When just the sum of net exports and the change in inventories (Sumneci) is included, the estimated coefficient is negative and significant at the 5 percent level (Table B-1, column g). These results suggest that the two components of GDP that best explain its negative coefficient are net exports and changes in inventories.

The negative impacts of net exports and changes in inventories are consistent with the speculation above. The estimated coefficient on government is also consistent with my speculation, but because it is so much smaller than its standard error, it seems to have no significant effect. The negative coefficient on consumption of nondurables and services is not what I speculated it would be, but it also is too insignificant to be considered to have any independent effect. The positive trend in GDP as well as its cyclical fluctuations must be better reflected by fixed investment, durables consumption, and possibly productivity. The main lesson seems to be that GDP does not have a monolithic impact on gains. In fact, individual components of GDP may tell more about the movement of gains than does their sum.

Why the coefficient of real GDP growth should be more statistically significant when forecasts of the explanatory variables are used instead of actual values remains unclear. Changes in inventories and net exports would seem to be as difficult to forecast as fixed investment and durable consumption.

Table B-1. Explaining the Annual Growth Rate of Tax-Adjusted Gains in Equations with GDP and Its Components, 1960-2000

Regression Statistics	a	b	c	d	e	f	g
R-square	0.383	0.320	0.347	0.316	0.346	0.378	0.393
Mean: depend. variable	11.133	11.133	11.133	11.133	11.133	11.133	11.133
SE of depend. variable	17.923	17.923	17.923	17.923	17.923	17.923	17.923
SE of estimate	14.846	15.583	15.265	15.628	15.278	14.904	14.716
SS Residual	7934.8	8741.5	8388.9	8793.0	8403.4	7996.6	7795.8
Reg. F(4,36)	5.575	4.230	4.786	4.153	4.762	5.462	5.835
Sig. Level of F	0.001	0.007	0.003	0.007	0.003	0.002	0.001
Durbin-Watson	2.222	2.301	2.265	2.279	2.262	2.230	2.217
Variables							
Constant	7.594	2.952	0.746	-1.024	-3.622	7.149	-1.663
	(6.013)	(9.670)	(4.397)	(4.529)	(4.505)	(5.989)	(3.980)
DG	13.536	11.769	3.148	9.569	6.564	7.910	-1.780
	(10.722)	(12.147)	(11.755)	(11.203)	(11.014)	(10.572)	(11.616)
Fixi	10.084*	2.634	1.953	2.090	5.755	4.310	6.217
	(5.896)	(4.656)	(4.503)	(4.732)	(5.243)	(4.522)	(4.710)
DLP1	4.833**	3.487	3.506	3.256	3.925	4.791*	4.461*
	(2.389)	(2.423)	(2.319)	(2.379)	(2.374)	(2.402)	(2.300)
GDP	-5.668*	______	______	______	______	______	______
	(2.863)						
Sumnds	______	-3.329	______	______	______	______	______
		(6.857)					
NE	______	______	-6.514	______	______	______	______
			(4.912)				
G	______	______	______	-0.756	______	______	______
				(4.941)			
CIPI	______	______	______	______	-6.564	______	______
					(5.044)		
Sumof4	______	______	______	______	______	-5.424*	______
						(2.854)	
Sumneci	______	______	______	______	______	______	-8.15**
							(3.785)

Notes: *, **, *** indicates the 10%, 5%, 1% level of significant respectively.
Variables are defined in text of Appendix B.

APPENDIX C. Tables of Annual Forecasts Used in Figures

Table C-1: Actual (ex 1997) and Forecast Gains Ratios from Equations in Ratio or Change-in-Ratio Using Gap (see Figure 3)

	Actual-x	Ratio	Change
1992	0.01957	0.02024	0.01992
1993	0.02241	0.02058	0.01981
1994	0.02148	0.02276	0.02231
1995	0.02413	0.02569	0.02566
1996	0.03325	0.02581	0.02622
1997	0.03882	0.02735	0.02825
1998	0.05258	0.04520	0.05009
1999	0.06075	0.05054	0.05489
2000	0.06696	0.05547	0.05965
2001	0.03432	0.06417	0.06690
2002	0.02525	0.04594	0.04672

Note: The actual value of the gains ratio for 1997 is 0.04422. The amount shown above is lowered by an estimate of how the capital gains tax change in 1997 raised the gains ratio.

Table C-2. Actual (ex. 1997) and Forecast Growth Rates of Gains in Figure 4 (Growth Rates in percentages.)

Year	Actual-X	Ratio	Mean Reversion	Two-Step
1992	13.53	9.09	10.65	26.55
1993	20.94	15.28	12.18	29.11
1994	0.31	12.22	9.48	27.58
1995	17.94	9.58	6.65	12.59
1996	44.73	3.08	3.50	33.06
1997	22.87	2.04	2.28	13.29
1998	24.79	-0.38	-3.45	30.80
1999	21.39	-3.49	-5.37	14.88
2000	16.59	-1.34	-5.10	27.43
2001	-45.78	1.22	-6.19	13.51
2002	-23.14	0.43	-4.33	-3.37

Notes:
Forecasts are made by ratio equation with the gap, mean reversion method, and two-step method.
Actual growth rate for 1997 is 39.93. The figure above is reduced to remove effect of capital gains tax reduction enacted in 1997. Actual figures in 1993 and 2001 differ from the above figures by small amounts because of small tax changes enacted in those years.

TABLE C-3. Variable Definitions	
Root Name	**Description**
MTRNEXT	Expected top tax rate on capital gains. Expectation is determined by the tax rate legislated to be in effect in the coming year.
MTRTRANS	The transitory tax rate on capital gains. Transitory is measured as the difference between the current top rate and MTRNEXT, with a sign-preserving square.
GAP	The ratio of GDP to potential GDP
IFIXFE	The ratio of fixed private investment to potential GDP
SP500Q4	Average value of the S&P 500 during the 4th quarter of one year.
DOLVOL	Dollar volume of shares traded on the NYSE, NASDAQ, and AMEX
STARTS	Starts of dwelling units other than single family dwellings.
Prefixes	
D------	Indicates annual change in variable ------
DL------	Indicates annual change in logarithm of variable ------