Projecting Tax Return Filings by Major Workload Categories

By Russell Geiman
National Headquarters Office of Research
Internal Revenue Service

This paper was presented at the 2002 IRS Research Conference.

Introduction

In calendar year 2001, individuals, businesses, governmental units and nonprofit organizations filed nearly 229.6 million tax returns and related forms with the Internal Revenue Service (IRS). Among the many IRS administrative duties associated with the regular filing of returns are two core responsibilities. One is to timely process these submissions to record the relevant information reported. Another is to systematically examine the information reported on a segment of these filings to promote compliance with reporting requirements under the tax code. These two core responsibilities are clearly embedded in the IRS mission to help taxpayers meet their tax responsibilities and to apply the tax law with fairness to all. To help senior management and staff throughout the IRS plan and budget for these submission processing and reporting compliance activities, members of the IRS National Headquarters Office of Research (NHQ Research) prepare updated projections of return filings on a regular basis. These projections cover a myriad of IRS workload categorizations including forecasts by form type, medium of filing, "examination class," and geographical location.

Tax return projections by type of filing medium (i.e., paper versus electronic), type of form and examination class are important budgetary and planning matters to IRS since processing costs, as well as compliance concerns, vary by type of workload category. For example, in terms of processing traditional paper forms, the more lines of data on the return that must be transcribed by IRS employees, the more it costs to process. In the case of individual income tax returns, data developed for IRS by a consulting firm for fiscal year 1999 in the individual returns area indicated that the average direct labor cost to process a Form 1040 return filed on paper was $1.93, compared to $1.50 for a paper Form 1040A, and $1.01 for a paper Form 1040EZ. As another illustration, the number (and skill level) of revenue agents needed to audit the income tax returns for a given number of large corporations with assets over $50 million is quite different from those required to examine the same number of income tax returns for individuals with no business income or expenses. In short, major parts of the strategic planning, resource allocation and related analytical decision making processes of the IRS rely upon workload projections developed by Research staff.

The following paper summarizes the basic methodologies we in NHQ Research use to generate forecasts of the number returns to be filed by major workload groupings. While Research staff project well in excess of 100 unique data series at the U.S. level, alone, we focus particularly on the methods for a few major workload categories/form types which we felt were interesting as well as indicative of the core statistical forecasting approaches we commonly use. We also limit our discussion to just the U.S. level methodologies to keep the paper manageable, although most IRS workload categories projected by NHQ

Research are ultimately driven down to much finer levels of projections such as states. We consider, in particular, the forecasting approaches for projecting the individual Form 1040 return series and its breakout by electronic filings versus paper Forms 1040, 1040A and 1040EZ; the Forms 1120, *U.S. Corporation Income Tax Return*, 1120A, U.S. *Corporation Short-Form Income Tax Return* and 1120S, *U.S. Income Tax Return for an S Corporation*, and; the breakouts of corporation and individual income tax returns by IRS-defined "examination classes." In the process, we also highlight a few of tax law changes, form changes, administrative developments and other unique factors (called "interventions") that have produced major disruptions to the underlying historical trends in question. Finally, we provide a bit more detail on how NHQ Research projections are used by staff in the IRS operating divisions.

One other caveat, readers should note that while the projections cited in this article are indicative of those prepared by NHQ Research staff, the forecasts presented here are only for illustrative purposes and do not reflect official IRS/Office of Research projections--which go through a more formal management review process and which are updated at different times of the year.

Baseline Forecasting Models And Consideration of Off-Model Adjustments

In general, we look to use econometric-based models (i.e., regression models based on economic or demographic information) when there exists a reasonably logical causal relationship between independent and dependent variables, and when long-term projections are needed. Time series (extrapolation) models are preferred when the estimation trend horizons are short or when historical values contain all the information. We also turn to extrapolation models when limited data or other factors prevent us from establishing a credible regression model. We refer to the output of these models as our "baseline" forecasts. These projections essentially capture the expected future volumes based on existing historical trends.

Many of the tax return data series we project contain major disruptions in their recorded "actual" (historical) trends. When this is the case, we often apply the standard "intervention" adjustment techniques of statistical forecasting, such as the use of "dummy variables" as part of the baseline models to account for these unique events. The statistical nature of these impacts (interventions) varies, but most effects are instantaneous and step-based with the trends continuing at the new level. An example discussed below is the 1993 drop in total individual return filings which was associated with the IRS' "Reduce Unnecessary Filing" initiative. However, some of the interventions have temporary effects on the data series, which then tends to return to "steady state." For short term forecasting purposes, these types of decayed response interventions must be addressed. However in the longer run, we can often ignore them in our forecasting approaches—as newer actual data become part of base period and the effects of the intervention fade away.

Once the baseline projections are produced, we must next consider whether there are any future interventions, such as tax law changes and/or significant administrative

developments, that will significantly disrupt the projected trend in question, but which are not adequately captured in the historical data. (An example would be an enacted piece of legislation that phases out the requirement to file a certain type return at some future date, such as the Economic Growth and Tax Relief Reconciliation Act of 2001 that gradually phases out the tax on estates, filed on Form 706, in the coming years.) If so, we then make further "off-model" (subjective) adjustments to our baseline forecasts to account for these future interventions. Such off-model adjustments are subjective in that they must be done outside of the statistical model that produced the original baseline forecasts. However, these off-model adjustments are often based on other empirical data, such as experience from a similar event in the past, in addition to applied assumptions drawn from persons with domain knowledge about the intervention factor in question. As a general rule, NHQ Research projections only incorporate the effects of enacted legislation and confirmed future developments—although alternative forecast scenarios are also occasionally developed to help with "what if" planning.

One other overall approach embedded in most NHQ Research projections is that we will generally forecast the largest aggregate total for a return series in question, first, before proceeding to the subordinate pieces of that workload category. For example, we will usually project the national ("U.S.") level total volume for the form type in question, before proceeding to projections by state, or by some finer U.S. level breakouts such as examination classes. These more aggregate totals, then serve as "controls" for the projections for the subordinate components. To achieve this consistency, the initial baseline projections from models for the subordinate pieces are adjusted, e.g., on a proportionate basis, to maintain consistency with the larger aggregate controls.

Individual Income Tax Returns: Projecting The Total Form 1040 Series Based on the U.S. Economy

In terms of projecting individual income tax returns, our first step is to set (forecast) the total individual Form 1040 series—defined as the sum of paper Forms 1040, 1040A, 1040EZ and 1040PC (for the years 1991 through 2000), plus electronically filed (e-file) returns. E-file returns include those filed electronically through an authorized third party tax professional, on-line filed returns using commercial tax preparation software, and "TeleFile" returns. In the case of the total individual return series, prior IRS modeling has established a clear statistical relationship between it and various indicators of the U.S. economy such as total employment and personal income (lagged one year). Personal income is a logical predictor variable since the returns filed are in fact **income** tax returns. Total employment is an equally logical predictor variable, not only because employment generates income potentially subject to tax, but also because "employment" generally entails a situation where the individual's wages are subject to withholding. And in an employment situation involving withholding, one might need to file a tax return to claim a refund even in instances where there is no income tax liability, per se.

The historical time series data for the total individual return series since 1980 is presented in Table 1, along with the corresponding year-to-year percentage change. Overall, this time series reveals a fairly "smooth" trend line that tracks the overall performance of the U.S.

economy, lagged one year (e.g., a decline or anemic growth in return filings during periods of recession). This is generally to be expected since the individual return series reflects an aggregate total that is unaffected by shifts among the subordinate pieces such as medium of filing or particular paper form type. Also, most tax law changes over the years have had de facto effects on tax rates and/or credits, but rarely on basic requirements for filing.

In terms of discernible interventions, we have generally identified two worth note. One intervention occurred in 1988 and 1989. Filings in these two years were somewhat higher than expected (as indicated by the relatively high recorded growth rates of 3.7 percent and 2.7 percent, respectively) as a result of the impact of the Tax Reform Act of 1986 (TRA86). While certain provisions of TRA86 eliminated filing requirements, other provisions (particularly the repeal of the personal exemption for those who could be claimed as a dependent on another's return) actually lead to an upward spike in total individual return filings. A second intervention in the total individual return series occurred in 1993. Total individual return filings in that year actually dropped by nearly one percent. While we suspect that this drop somewhat reflects a delayed effect from the recession in the early 1990's (and associated developments such as an unprecedented drop in interest rates which could have reduced income earned from savings), there were also other key factors at work. One of those factors was the culmination in 1993 of a major IRS initiative to reduce the number of individuals filing returns unnecessarily.

Projecting the Total Form 1040 Series with an Ordinary Least Squares Model

In terms of the specific forecasting model we pursue for the total individual return series, we test various econometric models incorporating economic variables like personal income, employment and gross domestic product, along with a "dummy variable" (i.e. step function) for the 1993 intervention. (In more recent years we have ignored the TRA86 impact since the intervention appears to follow a decayed response where the impact has slowly faded away.) We consider various model combinations, along with their comparative statistics (such as coefficients of determinations, F-tests, T-tests, P-values, Durbin-Watson values, etc.) and other characteristics such as the out-of-sample trend "nowcasts" (estimated values at the origin of the forecasts). We also examine the forecast results obtained from averaging the projections from two or more models.

An illustration of our approach is a recent effort where the methodology for individual return series trend entailed an average of two time series multiple regression models, both with the base period 1973-2001. The models were the same except that the primary predictor variable in one was total employment, while the second model substituted personal income (adjusted for inflation) instead. Presented below are some statistical details on the former model.

An ordinary least squares (OLS) model was estimated as follows:

Individual Return Series = 35,462,022 + 514,721(x1) – 1,190,671(x2) + 847,492(x3)

where

x1 = total employment in the previous year (measured in millions)
x2 = a dummy (indicator) step variable to adjust for the effects of the 1993 drop in the total return series
x3 = a time trend

Model Statistics: Adjusted R squared = 0.994 Significance F = 0.0001
Parameter t-statistic P-value for x1 = 0.0019
x2 = 0.1748
x3 = 0.0083
Mean Absolute Percentage Error (MAPE) = 0.73%

Both of the multiple regression models for the individual return series contain variables that have significant T statistics, i.e., P-values of less than 0.05 with confidence intervals of 95% and F-test less than 0.005. The one exception is in the model summarize above where there is a P-value of 0.17 for the dummy variable. However, we were comfortable with this relaxation of the 0.05 rule of thumb because of the intuitive logic of the "intervention" this variable represented, and because of the proper (i.e., negative) sign on the coefficient parameter (indicative of the observed drop in the series). In addition, we observed that the residual values fell within their horizontal bands on their relative correlogram (pass white noise/autocorrelation test) and the both models had mean absolute percentage errors (MAPE) less than 1%.

Both models also recorded adjusted R squares above 0.99, although we note that these values are based on nominal data that were not detrended. However, we also note that the time variable in both models served as a de facto method for detrending the data. In fact, in our experimentation we regressed the annual percentage changes in the return series data against the percentage changes in personal income, and in total employment, and got very comparable results/forecasts (albeit with lower Adjusted R Squares of approximately 0.7 and 0.4, respectively). However, we preferred the initial OLS models (with the time variable) since their resulting forecasts tended to be a slightly more conservative (lower) than the detrended models based on annual percentage change. We also elected to use an average of the two OLS models since it was a simple approach that gave us a set of forecasts that seemed intuitively sensible, and that also got us around the problem of multicollinearity—given that personal income and total employment are so highly correlated.

In the case of the total individual return series, there are presently no future interventions that we are aware of that would significantly increase or decrease total filings. Hence our

total individual return series is set after averaging the output of our two OLS models, and now serves as an overall "control" on all the other subordinate forecasts by filing medium and by form type--to be described in more detail below. These total individual series forecast controls are presented in Table 1, as well as Figure 1. For example, for filing year 2002 and 2003, the total number of US individual returns is estimated to be 131,270,800 and 132,465,600, respectively (reflecting annual growth rates of 1.41 percent and 0.91 percent). They are projected to reach nearly 137 million returns by CY 2005.

Projecting Corporation Income Tax Returns—Observed Trends and Speculation on Causes

The historical time series since 1980 on corporation income tax Forms 1120, 1120A and 1120S are also presented in Table 1, along with associated graphs in Figures 2 and 3. The Forms 1120, 1120A and 1120S make up the vast majority of all corporation income tax returns. (Other corporation forms include such returns as Form 1120F, *U.S. Income Tax Return of a Foreign Corporation*, and Form 1120H, *U.S. Income Tax Return for Homeowners Associations*, in addition to others.) Forms 1120 and 1120A are used by the traditional C (or "regular") corporations to report income, expenses, and associated corporate income tax liabilities. The Form 1120S is used by S corporations, who retain attributes of the regular corporations such as limited liability and freely transferable ownership, but in exchange for certain limitations, receive the benefits of a flow-through entity for income tax purposes. In effect, for S corporations, income and expenses pass through the corporation to the shareholders, who are then responsible for reporting any resulting tax liability on their individual income tax returns.

A review of the historical corporation return filing counts in Table 1 reveals a much different picture than the smooth, gradual upward trend observed for the total individual Form 1040 series. Over the years 1980 through 2001, regular C corporation (i.e., Forms 1120 and 1120A) filings vacillated between periods of growth and periods of decline, and when considered together (i.e., summed), actually "peaked" back in 1987—as indicated in Figure 2. In contrast, S corporation filings have rose continuously, including some periods with exceptionally strong bursts of growth such as 1988-1990, and 1997-1999. Not surprisingly, we tend to attribute much of these observed patterns to our good old friend "Mr. Intervention", along with some subtle economic dynamics.

In terms of the economic relationships that might be in play, on one hand, we generally hypothesize that the overall growth in the U.S. economy should translate into the formation of more corporations and thus a growth in corporation income tax return filings. However, mergers and acquisitions in the corporate world might actually serve to push down the number of Forms 1120 filed. Both these economic factors are assumed to be running through the historical data series.

In addition, Research staff over the years have also pointed to tax law changes that could be altering the trends in question. For example, considering the two major C corporations as one (to eliminate the distortions arising simply from the introduction and use of the alternative "short" Form 1120A), we see a period of general growth in Form 1120/1120A

filings from 1980 to 1987. This is then followed by a downward trend for the next six years starting in 1988. We generally speculate that this decline was at least partially attributable to the Tax Reform Act of 1986 (TRA86), which contained provisions that caused corporate tax rates to be higher than individual tax rates. In order to take advantage of the lower tax rates for individuals, some regular corporations likely took measures to shift over to Subchapter S corporation status. Such an intervention effect would also help explain the exceptionally strong growth in S corporation filings in the 1988 through 1990 period noted earlier.

This period of decline in regular C corporation filings during the late 1980's through early 1990's ended in 1993, at which point it resumed an upward trend for 1994 through 1996. And again, we suspect this may have been in part a reaction to further tax law changes, this time the Omnibus Budget Reconciliation Act of 1993. This act raised the marginal income tax rates for higher income individuals and we suspect may have resulted in more individuals incorporating to save on their taxes. Finally, the decline in C corporation filings (Forms 1120 plus 1120A) starts again in 1997 and has continued to date. And here again we suspect an intervention effect, this time arising from the Small Business Job Protection Act of 1996. This act contained provisions which relaxed previous requirements and enabled certain C corporations to now qualify for Subchapter S corporation status—which would serve to shift some filing volumes from Forms 1120 and 1120A and into Form 1120S.

Time Series Extrapolation Models for Forms 1120, 1120A and 1120S

While the discussion above is largely speculative about the causes for the trends in filings of Forms 1120, 1120A and 1120S, it does provide an overview of the many unique interventions and subtly changing economic dynamics that could logically be in play. In wrestling with these somewhat erratic trends, NHQ Research staff has tried various approaches over the years to project corporation income tax returns. This includes efforts that consolidated the various Form 1120 series return types into a single series and attempted econometric models tied to the U.S. economy. However, econometric models generally proved inferior to time-series extrapolation models, both in terms of model fit and the intuitive appeal of the forecasted trends. Hence we currently rely on a set of time series models tied to the particular form type to prepare our projections in the corporation income tax area. The extrapolation models explored include various autoregressive (AR) and moving average (MA) techniques, with a prime emphasis on the root mean square error as the selection criterion for the final model (but also with a watchful eye on the intuitive reasonableness of the resulting forecasts).

A recently developed model that illustrates our approach to the projection of Form 1120 filings is a damped trend exponential smoothing model with the base period 1980-2001.

Key parameters for this Form 1120 model and associated statistics are as follows:

Level smoothing weight =	0.55537	with t-statistic P-value =	0.0156
Trend smoothing weight =	0.99900	with t-statistic P-value =	0.2247
Damping smoothing weight =	0.83380	with t-statistic P-value =	0.0001
Smoothed Level =	2,136,400		
Smoothed Trend =	- 39,586		

Root Mean Squared Errors =	71,762
Mean Absolute Percent Error =	2.75%
R squared =	0.731

Our review of the actual data left us intuitively comfortable with an overall projected downward trend in this return series, although the P-value on the trend smoothing weight was above the 0.05 rule of thumb (and no doubt indicative of the up-and-down nature of the actual data series). The projections for this model are presented in Table 1. Also, a very similar damped trend exponential smoothing model was developed for the Form 1120A series and its forecasts are also show in Table 1. The combined Form 1120 and 1120A series for regular C corporation filings is presented in Figure 2.

In projecting the filings Form 1120S, the current model we like uses is an autoregressive process with a lag of one (1), and a step dummy variable starting in 1988. The base period is 1980 to 2001.

Key parameters for this Form 1120S model and associated statistics are as follows:

Autoregressive parameter, lag 1 =	0.96223	with t-statistic P-value =	0.0001
Dummy variable parameter =	146,585	with t-statistic P-value =	0.0001

Root Mean Squared Errors =	30,268
Mean Absolute Percent Error =	1.93%
R squared =	0.999

The projections from this Form 1120S model are contained in Table 1. They are illustrated graphically in Figure 3. They reflect a relatively strong upward trend.

Projecting Electronically Filed Individual Returns Versus Paper Returns—Use of the Diffusion Curve

Our approach to dividing the total individual Form 1040 series into electronic (e-file) versus paper filings focuses on the projections of e-file. In summary, e-file volumes are established, including their breakouts by underlying form type (i.e., Form 1040 versus 1040A versus 1040EZ). These e-file projections are then subtracted from corresponding

total return controls to derive the remaining volumes. These remaining volumes are the associated projections of paper returns. (This is also the same basic approach we use for most other form types that require a split between e-file and paper.)

The IRS introduced individual electronic filing in 1986 as an alternative method to filing paper individual income tax returns. Since then, individual return electronic filing has grown and evolved into three distinct markets (or product lines) that require separate consideration for forecasting purposes, viz., practitioner e-file, on-line filing, and TeleFile. However, a key and distinct forecasting methodology we use in the e-file area can be illustrated by focusing on our projection methodology for on-line filing. That distinct forecasting method is the application of the cumulative form of the innovation diffusion curve.

On-line filing refers to electronic filing of self-prepared returns (i.e., returns not professionally prepared by a tax-practitioner). To file on-line, the taxpayer must have a computer, modem, and tax preparation software from an IRS certified private vendor. On-line filing has grown with the proliferation of personal computers and the popularity of tax preparation software. In fact, it has experienced truly explosive growth. In calendar year 1996, in its second year of existence, 158,000 taxpayers participated in the on-line filing program. By 1998, 942,000 were filing using this method. And results to date from the current 2002 filing season indicate a total of around 9.4 million. This growth pattern for on-line filing is following that of a typical product innovation diffusion, or "S" curve, as depicted in Chart 4.

The "S" curve growth pattern typifies the historical usage/purchase pattern of many innovative consumer products, such as the automobile, the television, and in more recent decades the personal computer. Adoption is slow at first, explodes, and finally slows as the market's saturation point is reached. The marketing industry uses a variety of labels to describe the different segments of the curve, and the distinct groups of consumers and overall consumer behavior it represents. For example, the beginning of the curve shows initial adoption by a few "innovators." Moving right, along the curve, follows the progression to "early adapters," then to an "early majority," and finally to a "late majority" then "laggards." Chart 5 shows that the yearly volumes of on-line filing through 2002 strongly mirror the earlier stages of the "S" curve pattern.

A practical way to model this on-line filing market growth in the form of the diffusion curve is to express the volumes in terms of participation rates. Considering participants as a percentage of the total number of potential, or eligible, filers, we defined the following on-line filing participation rate ratio:

PR = [number of on-line filed returns] / [population of self prepared returns belonging to taxpayers that own a personal computer and that have internet access]

Since there is no ready source of information for the unique market reflected in the denominator of the above ratio, we have to estimate that component. To do so, we use U.S. Census, Forrester Research, Inc., and other sources of historical and projected data on the number of U.S. households with internet access. We combine this external data with internal tax return data on the number of filers that self prepare, and make certain other assumptions, to arrive at the denominators for the years in the projection horizon.

To model and forecast this participation rate at the U.S. level, we used the following two-parameter-bounded logistic growth function.

$$PR(t) = u / [1 + e^{(a - b*t)}]$$

where

t = time (in calendar years)
PR(t) = participation rate at time t
u = participation rate ceiling (predetermined)
e = 2.7182 (power series expansion)
a = scale parameter
b = shape parameter

Our first step in applying this function, and one that is particularly critical, is to set the "ceiling" participation rate *u* that will be achieved at some future point in the long run. By definition this value for *u* cannot exceed 100 percent (i.e., 1.0). Unfortunately this is a parameter that cannot be readily discerned from existing IRS information, and must be set by us at some assumed level. However, we have found some relevant data that we feel gives us a reasonable starting point for setting this ceiling. Information from a 1999 Council for Electronic Revenue Communication Advancement (CERCA) survey suggests that over 60 percent of the population of eligible e-filers stated they would file electronic returns provided the removal of all perceived barriers (including costs). We also found some similar empirical data from a not dissimilar "electronic convenience" product from the private sector that seemed in sync with the CERCA survey. It concerned the use of automatic teller machines (in existence since the 1970s) by individuals with bank accounts. According to a 1996 American Bankers Association and Gallup Consumer survey, 66 percent of bank customers in the U.S. had an ATM card. Similar surveys in 1993 and 1994 showed a 60 to 66 percent rate, suggesting a plateau (i.e., a "natural" ceiling) had been reached after participation growth through the 1970s and 1980s.

Based on this survey data information, we thus set this upper bound, *u*, at 0.66 in what we term our "likely" scenario for on-line filing. Once the assumed e-file participation rate ceiling u is set at 0.66, we can then use the historical values of PR(t) from 1996 through the latest available year to estimate the scale (*a*) and shape (*b*) parameters. We do this by selecting values for *a* (the scale parameter that moves the curve up, down, left, and right) and *b* (the shape parameter that determines the steepness of the curve) such that the root mean squared error of the fitted values is minimized. This selection process is accomplished by using a grid search routine programmed in SAS software. In recent

estimation efforts, we found the best fit at a = -4.0 and b = 0.6. These parameters produced fitted historical and projected on-line participation rate PR(t) values. We then multiplied the forecasted participation rates by our previously forecasted eligible pool of on line filers to produce the nominal value return volume forecasts.

The projection methodology for practitioner e-file also makes use of the diffusion curve model. The prime difference from the on-line filing modeling effort is simply the eligibility pool. With practitioner e-file, the participation rate modeling is focused on the population of individual income tax returns prepared by private sector tax professionals. For TeleFile, however, the diffusion model is not used, since the historical filing pattern for TeleFile does not follow the classic product innovation curve. As a result, we have relied upon trend extrapolations models to project TeleFile.

Alternative Individual e-File Forecast Scenarios to Deal With Uncertainty

We view the diffusion model described above as the best conceptual approach to modeling on-line filing and practitioner e-file. However, the reality is that our total e-file projection accuracy suffers relative to that of more traditional return types, as a consequence of the novelty of e-file and the larger variance in its recorded historical data series. For example, the mean absolute percent error for the total individual e-file (i.e., the sum of practitioner, on-line, and TeleFile) forecasts over the last three years for the one-year-out projection is 3.8 percent. This compares to a 0.3 mean absolute percent error for our total individual tax return (i.e., paper and e-file combined) forecasts. Though the "S" curve modeling provides a reasonable pattern for future growth, the speed of market maturation and the ultimate saturation point are not certain. Some of this uncertainty comes from less obvious factors (i.e., interventions) mostly outside the IRS's control. For example, to what extent will the practitioner community bundle their services (including e-file) under one package and one price? How many free on-line filing packages will be made available through the internet; by whom; and accessible to what segments of taxpayers? What new tax law provisions might Congress enact to further promote e-filing?

To handle this uncertainty, we now regularly provide senior IRS management with alternative e-file scenarios. Though most of the resource planning (particularly short-term staffing) tends to be based on the forecasts developed using our "likely" scenario, we also produce forecasts to reflect "optimistic" and "cautious" scenarios. These alternative scenarios provide IRS management with a fuller range of possible outcomes, in recognition of the uncertainty, and thereby enables them to tailor contingency plans. The basic approach for the scenario building for the practitioner and on-line filing projections is to consider different participation rate ceilings (i.e., the u parameters) in their respective two-parameter-bounded logistic growth functions. Again drawing from what limited relevant data we could discern, we have generally varied those ceilings for the cautious and optimistic scenarios from 50 percent and 80 percent, respectively.

Projecting Forms 1040, 1040A and 1040EZ—at an Adjusted Level

The volumes of individual income tax returns filed as paper Forms 1040, 1040A and 1040EZ have generally declined over the past decade as the e-file options have grown. However, the respective historical trend lines for these three major paper return types have shown everything but smooth, gradual declines. The data at the bottom of Table 2 provide a limited view of this fact by presenting the historical filings counts for paper Forms 1040, 1040A and 1040EZ, respectively, since 1996. The basic reasons for this situation are two fold: a) alternative ways of filing (i.e., mainly e-file) have reduced paper filings in all three form types in an unequal fashion, and; (b) there have been a number of other "interventions," such as new tax laws and administrative changes to the line items included on the respective forms, that have caused taxpayers to shift from one paper form type to another.

Alternative ways of filing (AWF) generally reflects the introduction and growth of the various e-file methods. However, from the early 1990's through 2000, AWF also included a highly condensed paper return, produced by special IRS-approved software, called the Form 1040PC. Still, whether an e-file return or a Form 1040PC, the effects of these AWF options were the same, viz., to reduce the volumes of the traditional paper Forms 1040, 1040A and 1040EZ. However, analyses of the AWF that determine the simplest type of traditional form the taxpayer could have filed (had they filed on paper) show that, to date, AWF have draw relatively higher shares of the simpler Forms 1040A and 1040EZ, and than from the "longer" Form 1040. This can be discerned from the data presented in the middle section of Table 2, which presents data on AWF sorted by the simplest form type the taxpayer could have used, had they elected paper. For example, in calendar year (CY) 2001, nearly 11.1 million AWF returns had the characteristics of Form 1040EZ, a volume that nearly equaled the total paper Form 1040EZ filings for that same year (11.6 million). In contrast, 13.7 million AWF had the characteristics of Form 1040 in CY 2001, where as nearly 62.4 million Forms 1040 were filed on paper.

To help get at the dynamics involved in this interplay between form types and AWF options, we employed another basic practice common in statistical forecasting, viz., data transformation. In effect, we "transform" the data into a revised series in an attempt to reduce the variations being observed. We refer to these transformed data as the "adjusted level" format. This format simply adds the AWF returns by type to the corresponding paper counts to derive "adjusted level" figures. Adjusted level data essentially serve to negate the impacts of AWF options and helps better reveal the true historical trends in the Form 1040 type returns, versus Form 1040A type returns, versus Form 1040EZ type returns. The adjusted level data by form type is presented at the top of Table 2. Also Figure 6 provides a graphical illustration of this transformation for the Form 1040A—contrasting the "paper only" trend line versus the "adjusted level" trend. As Figure 6 depicts, while there are still discontinuities in the adjusted series, these adjusted data provide a bit clearer picture of the underlying upward direction of the trend involved in Form 1040A type returns.

It is the adjusted level data that we focus on in projecting the individual income tax returns by form type. In effect, the three adjusted level data series (i.e., adjusted Form 1040 type, adjusted Form 1040A type and adjusted Form 1040EZ type) are projected and controlled to the total individual return series. Later, we then complete breakouts of the total e-file volumes by form type, and then subtract these e-file components from the adjusted levels to arrive at the final paper only forecasts of Form 1040, 1040A and 1040EZ.

Models for Adjusted Level Forms 1040, 1040A and 1040EZ to Deal with Other Interventions

The adjusted level data help stabilize the variances in the trends. Unfortunately, other legislative and administrative interventions are still embedded in the data series so some problems with non-stationary remain. In fact the number of interventions in the adjusted level data are quit extensive. As illustrative sampling of these includes the following:

- The 1988 through 1990 filing volumes for all three form types were significantly impacted (changed) by the sweeping provisions of the Tax Reform Act of 1986. Provisions such as the repeal of the personal exemption for those taxpayers (primarily young people) who could be claimed as a dependent on another's (primarily parent's) return initially increased filings of the shorter Forms 1040A and 1040EZ. In addition, major changes to the rules on itemized deductions, including the gradual phase-out of the deduction for state sales taxes paid, further added to a shift from Form 1040 filings to the simpler Forms 1040A and 1040EZ. Later, the Technical and Miscellaneous Revenue Act of 1988 allowed parents to claim the unearned income of certain children on their return, starting with 1990 filings, distorting yet again the nature of the underlying trends in the Forms 1040, 1040A and 1040EZ in the wake of TRA86.

- Data for 1994 reflect a major shift from Form 1040A filings, to Form 1040EZ, as a result of a form change to the 1040EZ enabling it to accept the "married, filing joint" filing status.

- IRS's 1995 "Revenue Protection Strategy" instituted a series of measures to combat refund fraud, particularly with respect to electronic filings, and contributed to a dramatic drop in the volume of e-file returns—particularly among those submitted through tax preparation professionals. As these former e-file returns were switched to paper, however, many were submitted on Form 1040 (the default paper return type for most tax practitioners)—even though a large share of these returns had characteristics of the simpler Forms 1040A or 1040EZ. This, in turn, tended to distort the recorded adjusted level results for all three forms in that year.

- Among disruptions embedded in the filings for 1996 is a shift of returns from Form 1040A to 1040EZ, as a result of a form change to the latter to accept the reporting of income from unemployment compensation.

- Finally, in the more recent years, 1999 showed a shift from Form 1040EZ filings to Form 1040 and 1040A, as a result of tax law changes that introduced new education credits, and ability to deduct interest paid on certain student loans. And these law changes continued to contribute to a shift from Form 1040EZ to 1040A in the year 2000.

In point of fact, most of the recorded historical data at the adjusted level from 1983 to 2001 reflect major interventions. Not surprisingly, such a limited data situation significantly restricts the number of traditional statistical forecasting methodologies we can apply, and requires us to employ more judgment in those processes we ultimately select. The adjusted level Forms 1040, 1040A and 1040EZ projection approach we currently favor uses weighted moving average models of order 3—starting with the three most recent historical years without an intervention. The these MA(3) models use the year-to-year percent changes as the detrending technique to attempt some level of stationary in the data. The weights were set at 0.7, 0.2 and 0.1, with the most recent data point attached the heaviest weight. These weights simply reflected our judgment but were based on an iterative process in which we varied them and examined the reasonableness of the resulting forecasts. The "force" to the total Form 1040 series control was handled by leaving the Form 1040EZ forecasted trend exactly as projected by the MA(3) model given its slight downward slope, and proportionately adjusting the results from the Form 1040 and Form 1040A 3 MA models to absorb the difference needed to match the control figure for the total individual return series.

Deriving the Final Projected Volumes of Paper Forms 1040, 1040A and Form 1040EZ

As previously mentioned, the baseline forecasts of paper Forms 1040, 1040A and 1040EZ are derived by simply subtracting expected AWF/e-file return volumes by form type, by year, from the corresponding adjusted level volumes. In general, the breakouts of AWF/e-file by form type are projected by first transforming these data into shares of the corresponding adjusted level volumes. These shares are then projecting by using simple extrapolation techniques not unlike those summarized above for adjusted level data by form type. The initial AWF/e-file forecasts by type are then adjusted to the to total e-file controls, before being subtracted from the adjusted volumes to arrive at the final paper counts of Form 1040, Form 1040A and Form 1040EZ. The data in Table 2 present an illustrative set of resulting projections for the adjusted level volumes by type; the AWF (i.e., e-file) volumes by type, and; the resulting paper filings for Forms 1040, 1040A and 1040EZ. In general, the forecasts indicate that future growth in e-file is expected to draw more heavily from the Form 1040 type, than from the simpler Forms 1040A and 1040EZ types, in contrast to historical experience to date.

Projecting Returns by Examination Classes

The final IRS workload category of projections I wish to cover in this paper are our forecasts by IRS "examination classes." Based on IRS studies of reporting compliance done over the years, particularly the old Taxpayer Compliance Measurement Program

(TCMP) studies, the major types of individual and corporation income tax returns can be slotted into unique and mutually exclusive groupings—which IRS refers to as "examination classes." These groupings are listed in Table 3 along with some associated information on volumes.

In the case of individual income tax returns, the examination classes are based on whether a major portion of the taxpayer's income was derived from a Schedule C (i.e., a "nonfarm business") or a Schedule F (i.e., a "farm business"), and the amount of the associated total gross receipts (TGR) from that business. The rest of the individual returns, including many with relatively small amounts of Schedule C and/or Schedule F income, fall into the "nonbusiness" category. These nonbusiness returns are grouped primarily by the amount of their "total positive income" (TPI). The examination classes for corporation income tax returns are basically grouped into dollar ranges based on the amount of their reported assets (or into "no balance sheet"—if one is not attached from which to determine their asset size). However, the corporation returns are first sorted into regular C corporations versus S corporations, before being slotted into respective asset size groupings.

Our projections by examination classes for both individual and corporation returns are typically handled in a similar manner. First trend extrapolation models are built for each respective examination class series. In our most recent attempts, these have generally been damped trend exponential smoothing models with base periods of around 10 years. The results from these initial models by examination class projections are then adjusted to match their corresponding controls, i.e., the total 1040 series, the total Form 1120S series, or the sum of the Forms 1120, 1120A, and selected other Form 1120 types that comprise the regular C corporation category for examination purposes.

Projections that are illustrative of our forecasting approach to individual and corporation tax returns by examination classes are summarized in Table 3. The top part of the table presents the data on the ten examination classes that comprise the individual returns; the middle section contains the information on the nine classes that comprise the regular C corporation returns; and the bottom portion contains the data for the three classes within S corporations. This table is constructed a little differently than the prior tables, and presents the actual filings in calendar year 2001, plus the projected average annual percentage change over the 2002 through 2005 period. We have also included Figure 7, which uses a bar chart to illustrate the differences in the average annual projected change through 2005, by examination class, for the regular C corporations. These data paint largely a dichotomous portrait for regular C corporations, one with declining volumes of corporations with assets under $5 million, and growth among those over $5 million.

Uses of Research Projections in IRS Resource Planning and Analysis

IRS strategic planning and budgeting efforts are heavily "data driven" and the workload projections developed by NHQ Research staff are an important ingredient in planning many of the IRS activities, including the submission processing and examination functions. And as illustrated by the workload projections discussed in this paper, most of the major

IRS return categories have significant, and uniquely varying, trends. These varying trends, in turn, have important implications in terms of IRS operational planning and resource allocation.

Uses in the Examination Activities

For example, the data in Table 3 indicate that while the number of regular C corporations overall is projected to decline through 2005 at an annual rate of 1.3 percent, this pattern is not uniform across all examination classes. Rather the decline is all concentrated in the lower asset examination classes under $5 million. Meanwhile, C corporations in the higher asset ranges are actually projected to grow, with the largest asset grouping (over $250 million) expected to grow the most at an average rate of nearly 4.9 percent per year. This is important information since it takes more highly skilled (and paid) revenue agents to work the larger (asset size) corporations than the smaller corporations. It is also important to IRS' operating divisions since the examination activities for corporations in the lower asset ranges are handled by the IRS Small Business/Self-Employed (SB/SE) Division, while the audit activity in the higher asset examination classes are mostly handled by IRS' Large and Mid-Sized Business (LMSB) Division.

Also, IRS staff use the Research projections by examination classes, particularly in the individual returns area, to allocate technical positions to various types of returns based on certain considerations or "constraints." One of those constraints is to consider the implications on audit "coverage" (i.e., the percentage of returns that will be audited within a given examination class) from various staffing allocation scenarios. While coverage is not expected to be the same in all examination classes due to other considerations such as the extent of expected non-compliance and the direct enforcement revenue generated from audits, it is still an important consideration. And since the process that considers coverage is started in advance of the fiscal year, it requires projections for the unique sub-populations involved. And as Table 3 shows, some individual examination classes are expected to decline (such as simple nonbusiness returns types with total position income under $25,000) while others are projected to rise quite markedly (such as nonbusiness returns with TPI over $100,000). Thus, for example, IRS could plan to reduce the number of examiners devoted to work in an examination class with a declining trend and still potentially maintain the same coverage rate as before.

Uses in Submission Processing Activities

The most extensive use of NHQ Research workload projections over the years has been for resource planning and analysis for submission processing activities. As noted earlier, processing costs vary by form type and medium of filing, so resource planning staff within the operating divisions, in effect, convert Research workload projections into resource needs by applying associated processing cost-per-return figures by workload type. For example, the projected trends illustrated in Table 2, which quantify the expected growth in individual electronic filing, and how its growth will influence the mix among paper Forms 1040, 1040A and 1040EZ, are among the most critical projections needed by senior IRS management and other policymakers. For these projections have major implications in

terms of IRS budget requests, and whether/when IRS can discontinue any of its paper returns processing "pipeline" operations. Also, in recent years these individual return projections by Research have been instrumental in determining the realignment of states within IRS submission processing centers boundaries—so as to implement the IRS modernization vision that calls for specialized sites by operating division.

In terms of the core mechanics of this process, resource planning staff in such IRS divisions as Wage and Investment, and SB/SE use the Research projections to develop their "work plans" and "work schedules" for the 10 centers that process the initial return filings. Based on the projected mix of returns by type, the work plans essentially divvy up the available IRS funding for submission processing across the 10 centers. Later on, as the filing season approaches, the latest projections are used to plan the hiring and training activities at each center, and to set (schedule) their expectations for rates of processing during the season—and, eventually, to track progress during the filing period as it unfolds.

And there are other important uses for Research workload projections. These include matters such as information system capacity planning, the ability to respond to inquires from Congress, the media and the general public, to help estimate the impact of new legislative or administrative proposals, and to identify growing areas needing potential IRS program attention. In short, workload projections are a key component to a data driven approach to strategic planning that enables IRS to fulfill its mission.

Conclusion

Workload projections developed by the National Headquarters Office of Research are intended as objective forecasts of what will happen given existing historical trends and confirmed (or reasonably certain) future developments. They are a critical input to the strategic planning and resource allocation processes for many of the IRS's major operational programs including its submission processing and examination activities. This paper examined in some detail the various methodologies used by the IRS staff to generate those forecasts, including projections by form type, medium of filing and examination classes. Drawing upon projections work done in the individual and corporation income tax return areas, we illustrated our applications of many of the forecasting methods common to the academic literature. Among these were: use of data transformations; the application of regression models based on economic variables; the use of dummy variables to account for interventions; the application of the innovation diffusion ("S") curve; and the application of time series extrapolation techniques such as auto-regressive (AR) models, exponential smoothing models, and other approaches involving moving averages (MA).

References

Armstrong, J. Scott, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, Norwell, MA, 2001

Council for Electronic Revenue Communication Advancement, "Practitioner Survey Report," December 1999.

Internal Revenue Service, *Calendar Year Return Projections for the United States and IRS Centers: Document 6186*, Washington DC, selected annual updates.

Internal Revenue Service, Electronic Tax Administration, *A Strategy for Growth, IRS Publication 3187*, Washington, DC, December 2000.

Internal Revenue Service, *Statistics of Income Bulletin*, Spring 2001, Washington, DC.

Internal Revenue Service, *The IRS Research Bulletin: Publication 1500*, Washington DC, Updates for 1990, 1993/1994 and 1999. Washington, DC.

Makridakis, Spyros, Steven C. Wheelwright, and Rob J. Hyndman, *Forecasting: Methods and Applications*, Third Edition, John Wiley & Sons, Inc. New York, New York, 1998

United States General Accounting Office, *Tax Administration: Electronic Filing's Past and Future Impact on Processing Costs Dependent on Several Factors*, GAO-02-205, (January 2002).

Note

This article was made possible by the excellent analytical support of Karen Masken, Andre Palmer and Javier Framinan from the IRS National Headquarters Office of Research. Still, the views expressed in this article represent strictly the opinions and conclusions of the author. They do not necessarily represent the opinion of the Internal Revenue Service.

Table 1. Illustration of Historical and Projected Tax Return Filings for Selected Major Form Types*

	Calendar Year of Filing	Individual Total Form 1040 Series		Form 1120		Form 1120A		Form 1120S	
		Number	Yr-to-Yr % change	Number	Yr-to-Yr % change	Number	Yr-to-Yr % change	Number	Yr-to-Yr % change
Actual	1980	93,052,300	2.56%	2,030,092		-		527,824	
	1981	94,014,000	1.03%	2,249,745	10.82%	-		547,177	3.67%
	1982	95,420,000	1.50%	2,229,913	-0.88%	-		566,787	3.58%
	1983	95,541,300	0.13%	2,455,688	10.12%	-		616,700	8.81%
	1984	96,496,900	1.00%	2,446,815	-0.36%	-		653,640	5.99%
	1985	99,528,900	3.14%	2,423,018	-0.97%	199,665		736,945	12.74%
	1986	101,750,900	2.23%	2,514,467	3.77%	285,134	42.81%	811,987	10.18%
	1987	103,250,700	1.47%	2,542,261	1.11%	300,760	5.48%	892,376	9.90%
	1988	107,028,900	3.66%	2,462,931	-3.12%	285,777	-4.98%	1,169,736	31.08%
	1989	109,868,300	2.65%	2,424,623	-1.56%	296,726	3.83%	1,351,092	15.50%
	1990	112,305,000	2.22%	2,329,560	-3.92%	332,025	11.90%	1,536,147	13.70%
	1991	113,829,200	1.36%	2,252,935	-3.29%	336,112	1.23%	1,663,777	8.31%
	1992	114,718,800	0.78%	2,248,538	-0.20%	338,312	0.65%	1,805,291	8.51%
	1993	113,754,400	-0.84%	2,127,419	-5.39%	354,370	4.75%	1,905,765	5.57%
	1994	114,683,400	0.82%	2,157,592	1.42%	335,702	-5.27%	2,036,736	6.87%
	1995	116,059,700	1.20%	2,196,969	1.83%	319,146	-4.93%	2,161,015	6.10%
	1996	118,362,500	1.98%	2,240,844	2.00%	328,005	2.78%	2,290,904	6.01%
	1997	120,342,500	1.67%	2,249,894	0.40%	293,652	-10.47%	2,449,928	6.94%
	1998	122,546,900	1.83%	2,207,641	-1.88%	272,482	-7.21%	2,599,837	6.12%
	1999	124,887,100	1.91%	2,202,352	-0.24%	260,807	-4.28%	2,767,034	6.43%
	2000	127,097,200	1.77%	2,161,690	-1.85%	245,477	-5.88%	2,887,103	4.34%
	2001	129,444,900	1.85%	2,128,731	-1.52%	235,798	-3.94%	3,022,589	4.69%
Projected	2002	131,270,800	1.41%	2,106,435	-1.05%	226,492	-3.95%	3,162,778	4.64%
	2003	132,465,600	0.91%	2,075,872	-1.45%	219,137	-3.25%	3,287,978	3.96%
	2004	134,565,600	1.59%	2,052,925	-1.11%	213,521	-2.56%	3,411,640	3.76%
	2005	136,913,100	1.74%	2,033,792	-0.93%	208,985	-2.12%	3,530,632	3.49%

* Projections are for the illustrative purposes of this article only; they should not be interpreted as official IRS forecasts.

Table 2. Underlying Composition of Individual Returns and Alternative Ways of Filings by Form Type *

	Year of Filing	Adjusted Form 1040	Adjusted Form 1040A	Adjusted Form 1040EZ
		Adjusted Level Individual Returns by Form Type **		
Actual	1996	66,849,900	28,029,000	23,483,700
	1997	69,210,400	27,609,300	23,522,800
	1998	70,774,200	28,025,400	23,747,300
	1999	73,095,700	28,349,200	23,442,300
	2000	74,501,600	29,837,200	22,758,400
	2001	76,169,600	30,614,900	22,660,500
Projected	2002	77,530,900	31,100,500	22,639,300
	2003	78,416,800	31,422,200	22,626,700
	2004	79,940,200	32,007,600	22,617,800
	2005	81,646,700	32,659,400	22,606,900
		Alternative Ways of Filing (AWF) by Form Type ***		
	Year of Filing	Form 1040 Type	Form 1040A Type	Form 1040EZ Type
Actual	1996	6,575,600	9,025,500	6,384,000
	1997	8,190,600	10,273,100	9,070,500
	1998	9,427,300	11,663,500	10,992,100
	1999	11,472,800	12,760,100	11,608,800
	2000	13,831,700	15,093,400	11,612,700
	2001	13,719,700	15,402,300	11,084,800
Projected	2002	16,842,600	17,271,700	11,885,800
	2003	20,620,500	19,563,400	12,716,100
	2004	24,273,800	21,532,200	13,576,900
	2005	28,319,700	22,658,100	14,421,200
		Paper Only Individual Returns by Type		
	Year of Filing	Paper Form 1040	Paper Form 1040A	Paper Form 1040EZ
Actual	1996	60,274,300	19,003,500	17,099,600
	1997	61,019,800	17,336,200	14,452,300
	1998	61,346,900	16,361,900	12,755,200
	1999	61,622,900	15,589,100	11,833,500
	2000	60,669,900	14,743,800	11,145,600
	2001	62,449,800	15,212,600	11,575,800
Projected	2002	60,688,400	13,828,900	10,753,500
	2003	57,796,200	11,858,800	9,910,600
	2004	55,666,400	10,475,300	9,040,900
	2005	53,327,000	10,001,400	8,185,700

* Projections are for the illustrative purposes of this article only; they should not be interpreted as official IRS forecasts.
** "Adjusted Level" counts reflect total individual returns by approximate form type had Alternative Ways of Filing not existed.
*** Alternative Ways of Filing includes all electronically filed returns, TeleFile returns and Form 1040PC volumes.

Table 3. Illustration of Projected Individual Income Tax and Corporation Income Tax Return Filings by Traditional Examination Classes *

Workload Category	Actual Filings During Calendar Year 2001	Projected Average Annual Percentage Change During the Period 2002 thru 2005
Total Individual Form 1040 Series **	129,444,947	1.41%
Nonbusiness:		
TPI Under $25,000 - Form 1040A/EZ Type	40,560,604	-3.07%
TPI Under $25,000 - Form 1040 Type	14,106,067	1.51%
TPI $25,000 Under $50,000	30,720,483	1.18%
TPI $50,000 Under $100,000	24,702,589	4.97%
TPI $100,000 or More	10,692,928	10.54%
Nonfarm Business:		
TGR Under $25,000	2,541,662	0.26%
TGR $25,000 Under $100,000	3,425,939	0.60%
TGR $100,000 or More	2,059,115	1.95%
Farm Business:		
TGR Under $100,000	367,420	-4.50%
TGR $100,000 or More	268,140	0.42%
Corporation Series Returns		
Total Forms 1120, 1120-A, and Other ***:	2,389,080	-1.32%
No Balance Sheet	293,983	0.00%
Assets Under $250,000	1,395,497	-2.19%
Assets $250,000 Under $1 Million	417,973	-0.75%
Assets $1 Million Under $5 Million	191,456	-0.17%
Assets $5 Million Under $10 Million	30,570	1.64%
Assets $10 Million Under $50 Million	32,570	1.79%
Assets $50 Million Under $100 Million	8,057	1.12%
Assets $100 Million Under $250 Million	8,067	2.55%
Assets $250 Million or More	10,908	4.89%
Total Form 1120S:	3,022,589	3.96%
Assets Under $200,000	2,247,934	3.98%
Assets $200,000 Under $10 Million	749,384	3.74%
Assets $10 Million or More	25,271	8.62%

* Projections are for the illustrative purposes of this article only; they should not be interpreted as official IRS forecasts.
** "TPI" stands for total positive income; "TGR" stands for total gross reciepts.
*** Other includes the following Forms: 1120L/PC/SF/FSC/REIT/RIC.

Figure 1. Total Form 1040 Series Filings

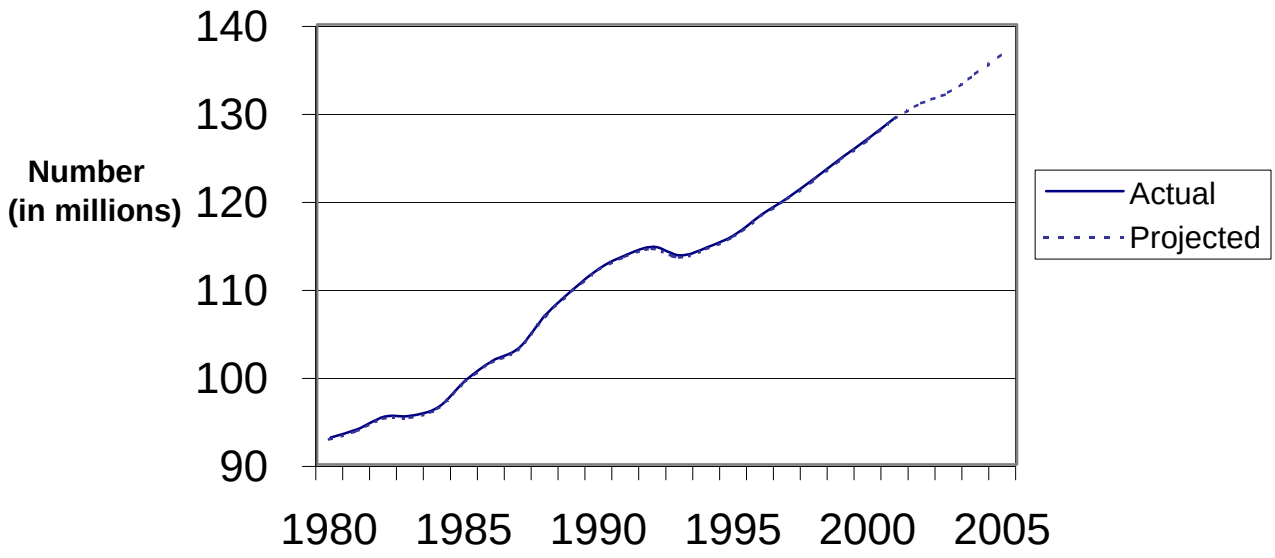

Figure 2. Total Form 1120 and 1120A

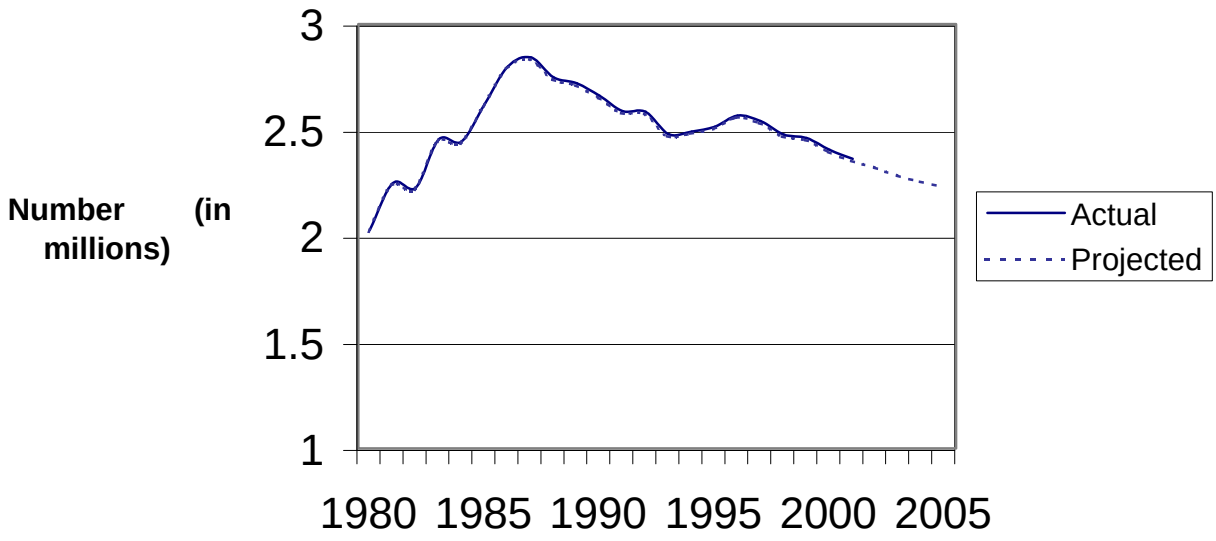

Figure 3. Total Form 1120S

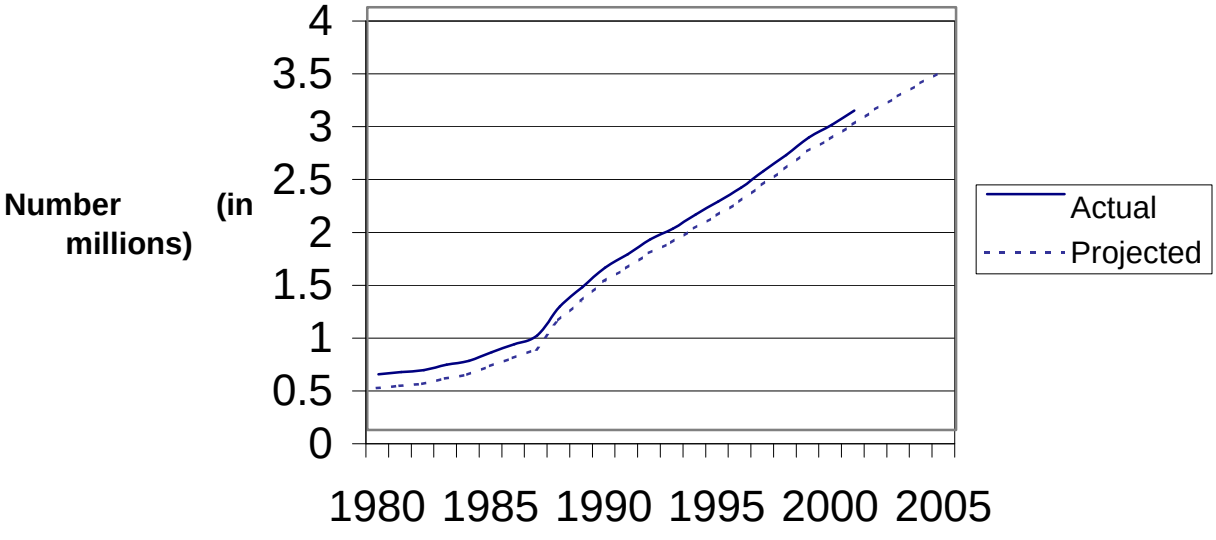

Figure 4. Innovation Diffusion Curve

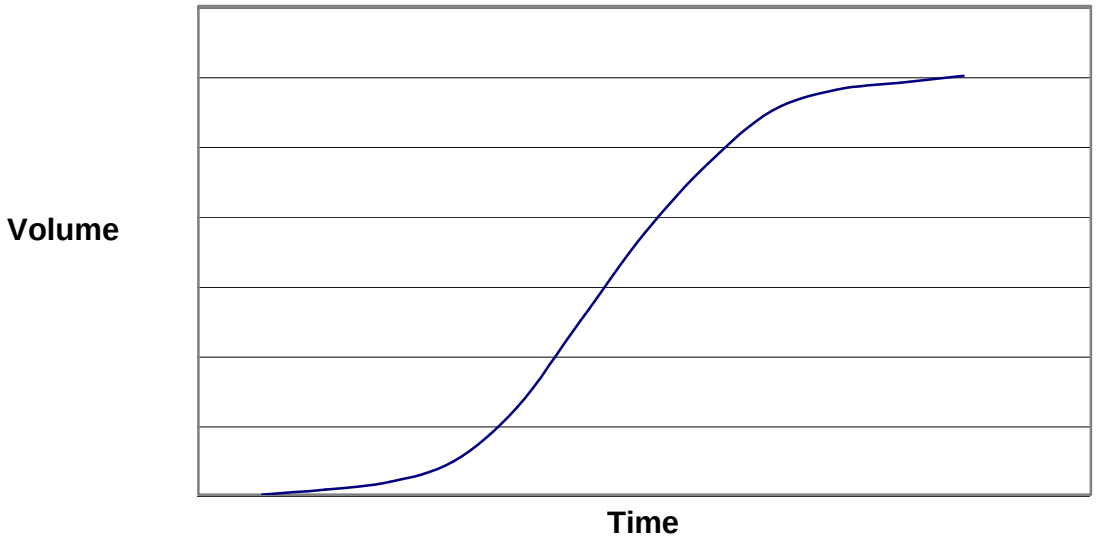

Figure 5. On-line E-file Volumes To Date

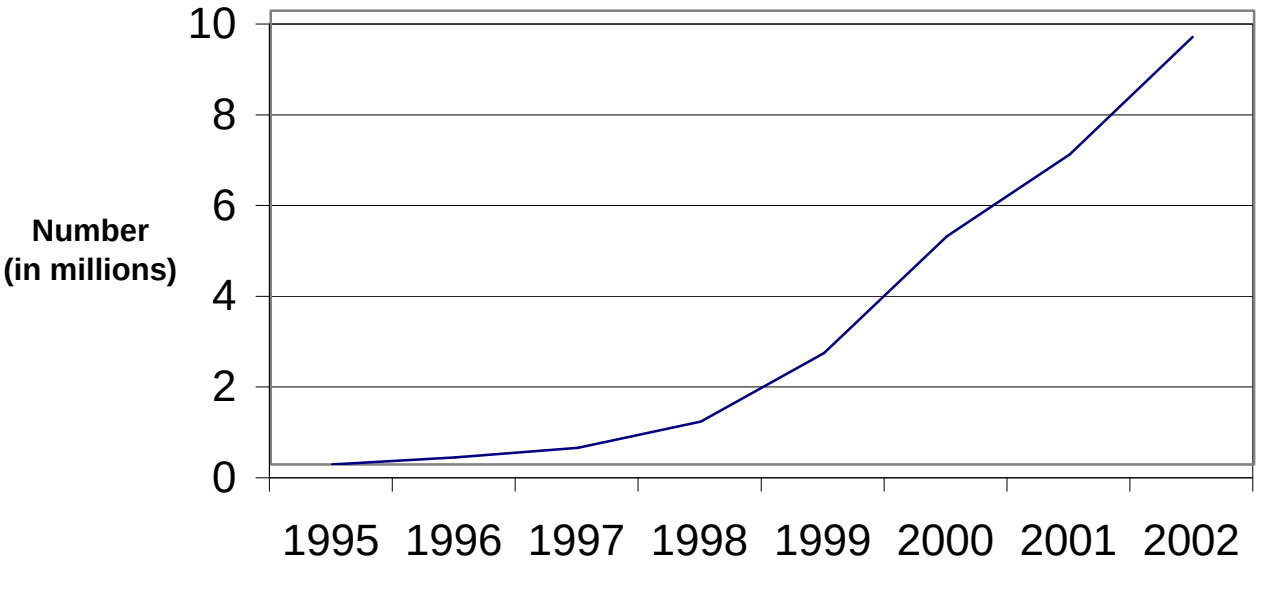

Figure 6. Paper Only vs. Adjusted Level Form 1040A Returns

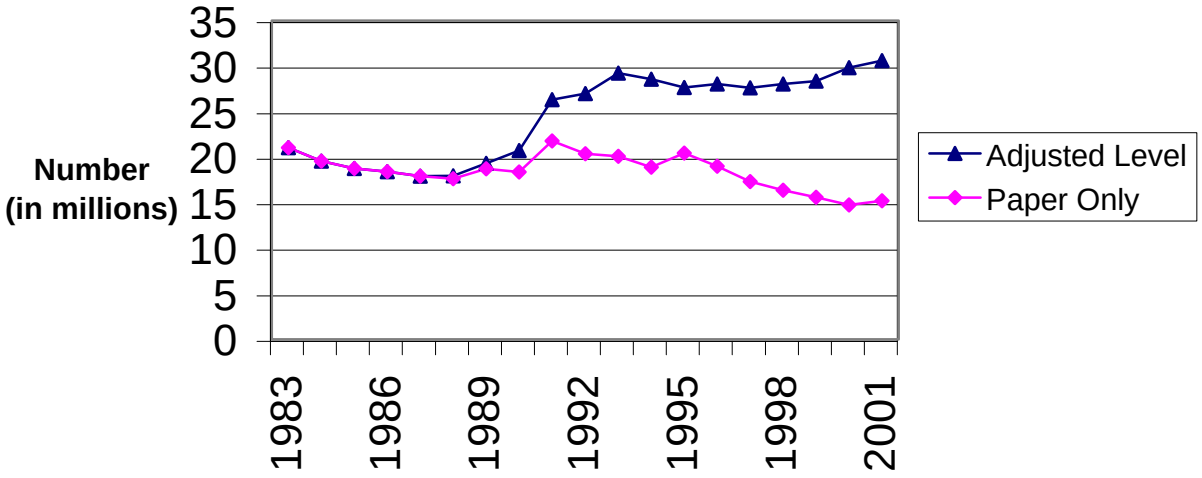

Figure 7. Regular 1120 Exam Class Projected Average Annual Percent Change During the Period 2002 thru 2005