Chapter 3 Revenue Forecasting

3.1 Introduction[1]

Over the years, a key issue in the design of fiscal policy rules has been the accuracy of government budget forecasts, particularly those of tax revenues. During the early years of the Reagan administration, "supply-side" forecasts of budget surpluses gave way to the reality of large deficits that persisted into the 1990s, despite several policy changes aimed at deficit reduction (Auerbach, 1994). The persistence of overly optimistic forecasts led to the perception that, as budget forecasts came to occupy a more central role in the policy process, the pressure on forecasters to help policy makers avoid hard choices led to forecasting bias. However, the experience of recent years suggests that prior conclusions may need to be amended.

The remarkable U.S. economic expansion of the 1990s has also been a challenging period for government revenue forecasters, but with different consequences than before. Continual surprises as to the overall strength of the U.S. economy and the share of income going to those in the top income tax brackets have led to a series of what turned out to be overly pessimistic aggregate revenue forecasts. Large predicted deficits gave way to smaller predicted deficits and, ultimately, the realization of budget surpluses and the forecast of larger surpluses to come. These large revisions have had a significant impact on the budget process of a government that has lashed itself to the mast of revenue forecasts to help it withstand the political sirens in its path. Even without such budget rules, though, revenue forecasts would remain an important input to the design of fiscal policy, for they provide a sense of what fiscal actions are sustainable over the longer term.

Some critics of the recent government revenue forecasts have suggested a bias, albeit in the direction opposite to that argued to exist in the earlier period. They point to the size of forecast errors and the consistent sign of these errors in arguing that the forecasters should have been able to do better. But over, short periods, any sequence of individual forecast errors can be given a rational justification, and virtually any statistical error pattern is consistent with an efficient forecasting process that uses all available information in forming its predictions. One needs more data to get a better sense of whether the recent forecasts really are consistent with forecast efficiency, or whether they simply illustrate the continuation of a process of biases

[1] Section 3.1-3.7 draw selectively from Auerbach (1999, pp.767-80).

forecasting, with perhaps a change in bias.

3.2 Data and Methodology

Twice each year (and, on occasion, more frequently), the Congressional Budget Office (hereafter, CBO) and the Office of Management and Budget (hereafter, OMB) produce revenue forecasts for the current and several upcoming fiscal years. One forecast typically occurs around the beginning of February with the presentation of the *Federal Budget* by OMB and the roughly coincident *Economic and Budget Outlook* published by CBO. The second typically occurs in August or September, with OMB's *Midsession Review* and CBO's *Economic and Budget Outlook: An Update*. For most of the sample period, both winter and summer forecasts have been for six fiscal years, including the one ending September 30 of the same year[2].

In revising its revenue forecasts for a particular year, each government agency incorporates changes in its own economic forecasts as well as estimates of the effects of policy changes. For a number of years, both CBO and OMB have followed the practice of dividing each forecast revision into three mutually exclusive categories: *policy*, *economic*, and *technical*. *Policy* revisions are those attributed to changes from "baseline" policy. *Economic* revisions are changes attributed to macroeconomic events. *Technical* revisions are residual, containing changes that the agency attributes neither to policy nor to macroeconomic changes. For example, a technical revision in revenue would result from a change in the rate of tax evasion, a shift in the composition of capital income from dividends to capital gains, or a change in the distribution of income.

We may express this revision process as:

$$x_{i,t} - x_{i+1,t-1} = p_{i,t} + e_{i,t} + r_{i,t} \qquad (1)$$

where $x_{i,t}$ is the *i*-step-ahead forecast at date *t*, and $p_{i,t}$, $e_{i,t}$, and $r_{i,t}$ are the policy, economic, and technical components of the revision in this forecast from period *t*-1. For CBO, these semiannual data on revisions are available continuously from 1984 through the present. While OMB does not publish comparable forecasts, it has produced them internally for revenues over roughly the same period, for the same forecast horizons as are considered by CBO. For the two agencies together, comparable semiannual data are continuously available for revisions during the period from the summer of 1985 to the winter or 1999[3]. Initially, I

---

[2] Recently, CBO has begun to report forecasts over an eleven-year horizon.

[3] Actually, OMB and CBO data on overall revisions $x_{i,t} - x_{i+1,t-1}$ -- but not their breakdown by source – may be constructed for a longer period from successive revenue forecasts, $x_{i,t}$. For an analysis of revisions during this

analyze the sample that begins one forecasting period later, with the revision between forecasts made in the summer of 1985 and the winter of 1986, for this is the first period for which I have been able to obtain comparable Data Resources, Inc. (hereafter, DRI) forecasts[4]. For the corresponding DRI forecasts, though, only the aggregate revisions, $x_{i,t} - x_{i+1,t-1}$, are available.

Because there are two forecasts made each year for revenues in each of six fiscal years, there are twelve forecast horizons and twelve corresponding revisions for the revenues of each fiscal year being predicted. The last such revision, however, differs from the others, in that it reflects changes over the very brief period between the late summer forecast during the fiscal year itself and the September 30 end of that fiscal year[5]. Thus, for the statistical analysis below, I consider only the first eleven forecast revisions for each fiscal year, labeling these revisions 1 (shortest horizon, from winter to summer during the fiscal year itself) through 11 (longest horizon, five years earlier). Odd-numbered revisions are those occurring between winter and summer; even-numbered revisions are those occurring between summer and winter. I refer to the final revision, occurring between the final forecast and the end of the fiscal year, as revision 0.

3.3 Forecast Evaluation

Using the individual revisions underlying the means, one can construct formal tests of forecast efficiency. Consider the relationship between successive forecast revisions for the same fiscal year, $y_{i,t}$ and $y_{i+1,t-1}$. According to theory, if each forecast is unbiased and uses all information available at the time, these revisions should have a zero mean and should be uncorrelated. Letting a_i be the mean forecast revision for horizon *i*, we relate these successive forecasts by the equation:

$$y_{i,t} - a_i = \rho_i (y_{i+1,t-1} - a_{i+1}) + \varepsilon_{i,t} \tag{2}$$

or

$$y_{i,t} = \rho_i y_{i+1,t-1} + \alpha_i + \varepsilon_{i,t} \tag{3}$$

earlier period, see Plesko (1988).

4 DRI makes forecasts at 3-year horizons monthly, but publishes long-range (10-year) forecasts only twice a year – currently in May and November, but with some timing changes over the years. For purposes of comparison with the forecasts of CBO and OMB, I align the May and November forecasts with the CBO and OMB forecasts immediately following.

5 Indeed, CBO does not provide a breakdown of this last forecast revision into components.

where $\alpha_i = a_i - \rho_i a_{i+1}$ and $\varepsilon_{i,t}$ has zero mean and is serially independent[6]. The hypothesis of forecast efficiency implies that $\alpha_i = 0$ (no bias) and $\rho_i = 0$ (no serial correlation).

α_i, in expression (2) is the mean values, Table 1 presents estimates[7] of the coefficients ρ_i for OMB and CBO[8]. The table presents estimates based on two alternative assumptions regarding the sample period. The first is that values of a_i and α_i are constant throughout the period, the second that the means differ between the pre-Clinton and Clinton periods. The table also lists standard errors for each coefficient, ρ_i. At the bottom of each column of estimates are *p*-values corresponding to *F*-tests of three joint hypotheses related to forecast efficiency[9]. The first is that all means are zero for the agency in question ($\alpha_i \equiv 0$); the second, for the case in which separate subsample means are allowed through the use of dummy variables δ_i, is that the differences in means across subsamples are zero ($\delta_i \equiv 0$). The third is that all correlation coefficients are zero ($\rho_i \equiv 0$). Of the first two sets of tests, only the test of different means for CBO fails to be rejected at the .05 level of significance.

Table 1 Serial Correlation of Successive Economic + Technical Forecast Revisions, by Horizon, 1986-99

Treatment of Subsamples:	Common Means				Separate Means			
	OMB		CBO		OMB		CBO	
Forecast Horizon:	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
1	0.52	0.17	0.70	0.24	0.34	0.25	0.57	0.28
2	0.81	0.21	0.56	0.14	0.69	0.23	0.34	0.16
3	0.19	0.19	0.61	0.26	0.03	0.22	0.42	0.24
4	1.28	0.21	0.67	0.16	1.30	0.27	0.37	0.21
5	0.11	0.24	0.59	0.31	-0.03	0.20	0.38	0.29
6	1.28	0.17	0.71	0.14	1.45	0.31	0.37	0.22
7	0.07	0.24	0.50	0.32	-0.01	0.19	0.31	0.31
8	1.21	0.26	0.71	0.14	1.35	0.42	0.34	0.22
9	0.08	0.19	0.39	0.32	0.05	0.15	0.19	0.32
10	1.16	0.39	0.60	0.16	1.37	0.53	0.21	0.20
p-value (*F*-test) $\alpha_i \equiv 0$	.0290		.0213		.0037		.0012	
p-value (*F*-test) $\delta_i \equiv 0$	---		---		.0009		.1165	
p-value (*F*-test) $\rho_i \equiv 0$	.0000		.0011		.0006		.2218	

The coefficients in the first panel of the table, based on common means, show substantial

[6] For revisions at horizon 11, there is no lagged revision, so expressions (2) and (3) become $y_{11,t} = a_{11} + \varepsilon_{11,t} = \alpha_{11} + \varepsilon_{11,t}$.

[7] The estimation is based on the full sample, using lagged values from the summer, 1985 revision.

[8] Formally, these coefficients are correlation coefficients only if the variances of forecasts at successive horizons are the same. However, sample variances typically do not vary much across horizons.

[9] These tests are based on the estimated variance-covariance matrix of each agency's contemporaneous revisions for different horizons, and so reflect the strong correlation among these revisions.

serial correlation for both agencies. The finding of serial correlation in revenue forecasts is not a new one. For example, Campbell and Ghysels (1995) report some evidence of serial correlation in aggregate annual OMB forecasts. As discussed above, one might have expected some of this serial correlation to be due to the presence of the policy component in each forecast. However, it turns out that eliminating the policy components of the successive revisions actually strengthens the results, typically increasing the estimated serial correlation coefficients.

All serial correlation coefficients based on the common means are positive, and the hypothesis that all are zero is strongly rejected. This suggests a partial adjustment mechanism, with not all new information immediately incorporated into forecasts. One can readily imagine institutional reasons for such inertia. For example, it might be perceived as costly to change a forecast and then rescind the change, leading to a tendency to be cautious in the incorporation of new information in forecasts. However, the serial correlation patterns differ between the two agencies.

3.4 The Potential Impact of Taxpayer Behavior

Another implication of forecast efficiency is that forecast revisions should not depend on information available to the forecasters at the time that the initial forecast was made. Failing to take account of such information is one potential source of bias and serial correlation, depending on the nature of the information being ignored. When on considers the behavioral impact of tax changes, "ignoring" information would amount to incorporating systematically incorrect forecasts of taxpayer response; for example, forecasts might systematically understate the strength of taxpayer reaction.

The logic is simple. If revenue estimates overstate the impact of tax increases then, during the period after which taxes increase, there will be subsequent downward revisions in estimated revenue, as estimators in estimated revenue, as estimators realize that they initially had overestimated the impact of the policy change[10]. If this realization occurs over time, it would impart serial correlation to the revisions. If the tax changes being evaluated tend to be in one direction or another during the sample period, this could also impart a bias to the forecast revisions, causing excessive optimism in a period of tax increases and excessive pessimism in a period of tax reductions.

One approach to testing this hypothesis involves regressing the combined economic and technical revisions on lagged policy revisions (thus far excluded from the statistical analysis) for

---

[10] Systematic errors of this sort would not occur simply as a result of the convention of excluding macroeconomic feedback effects from estimates of the impact of policy changes. While such feedback effects are not attributed to individual policies, they are, in principle, incorporated in subsequent macroeconomic forecasts. Thus, if feedback effects were estimated correctly, there would be no need for subsequent forecast revisions.

the same fiscal year, a procedure introduced in Auerbach (1994, 1995). However, estimation using the OMB and CBO data, for a variety of lagged policy revisions, in no cases led to a significant effect, and in most cases led to insignificant effects of the wrong sign (a positive coefficient).

These findings stand in contrast to those reported in Auerbach (1995), where Auerback founds significant effects in an examination of short-horizon OMB revenue forecasts. However, there are at least four differences between the two data sets that can help explain the difference in findings. First, the prior study did not include observations from recent years, during which the large tax increases of 1993 were followed by stronger than predicted revenue growth. Second, the earlier paper considered just technical forecast errors, which Auerback argued there should show more evidence of behavioral response, for they represent precisely the errors that cannot be explained by macroeconomic phenomena. Third, that study found significant effects only for certain disaggregate revenue categories (corporate tax revenues and excise tax revenues), not for the aggregate revenue category being considered here. Finally, as emphasized above, the policy revisions to forecasts do not necessarily measure true changes in policy, but simply changes in the "baseline", which need not reflect actual current policy. Auerback's earlier paper made use of an alternative series that better measured the policy effects of legislative changes, but such a series is available only for OMB, and only at annual frequencies.

Thus, the current findings do not contradict Auerback's earlier ones; we simply lack the data necessary to address the question of taxpayer response in the current context. More generally, these findings in no way rule out the possibility that there could be other types of information available to forecasters at CBO and OMB that are not incorporated in the forecast revisions studied here.

3.5 Implications for Forecasting and Policy

It requires a certain boldness to draw strong implications from the empirical results presented above. We do not really know why the revisions of government forecasts exhibit serial correlation and seasonality, and hence we cannot predict whether this pattern will continue. We cannot rule out the possibility that the seemingly huge and persistent forecast revisions of recent years occurred by chance. But we can safely conclude that the information conveyed by these forecasts and the process by which they are produced is not adequately summarized by the point estimates delivered twice a year to policy makers.

Budget rules currently in effect, and those of earlier periods, do not account for the fact that revisions are persistent. Nor do they make any allowance for the very large standard errors surrounding each forecast, and the fact that a rational policy response to uncertainty might include some fiscal precaution, much as a household would engage in precautionary saving when facing an uncertain future[11]. For example, even if a zero deficit were an appropriate target (and there are good reasons why it probably is not, given the looming fiscal pressures of demographic transition), it might be optimal to structure revenues and expenditures so that an unbiased forecast would predict a surplus. Therefore, in reaction to the fact that budget rules are based only on point estimates of revenue, it might be optimal to build a downward bias into these point estimates. This illustrates the difficulty of producing forecasts intended simultaneously to provide information and to act as inputs to the budget process.

To this state of affairs, one might suggest a number of responses.

First, take whatever measures may be available to improve current forecasting methods. This recommendation undoubtedly falls into the category of "easier said than done", but there must be some explanation for the anomalous pattern of forecast revisions discussed above. Perhaps the explanation lies in the use of mechanical rules, even for the economic and technical components of forecasts, in accordance with certain requirements of the budget process. Alternatively, the pattern may reflect the various incentives present when budget forecasts play such a central role in the policy process. If either of these explanations applies, then the problem may also be addressed by some of the remaining suggestions.

Second, as it is probably unrealistic (and perhaps also unwise) to consider incorporating greater sophistication into budget rules, reduce the mechanical reliance of policy on such rules. As a vast literature elucidates, there are trade-offs of costs and benefits in adopting rules. As to the benefits, many believe that the rules provide credibility to fiscal discipline that would be lacking otherwise. This may or may not be so. But rules also impose costs, by restricting the flexibility of policy responses. While such restriction is inevitable when rules are imposed, being bound by budget rules that so fully ignore available information seems to present very significant costs as well.

Third, do not ask even more of the forecasting process than we presently do, at least until the previous two recommendations are accepted. In particular, do not require "dynamic scoring" for official purposes, or other projections likely to be based on limited information. In brief, dynamic scoring" for official purposes or other projections likely to be based on limited information. In brief, dynamic scoring involves incorporating macroeconomic feedback into each individual revenue estimate, as opposed to the current practice simply of updating the baseline over time to take all changes, including those induced by legislation, into account[12]. In principle, dynamic scoring is a good idea, for it permits the legislative process to be based on

---

[11] For a discussion of the impact of uncertainty on optimal fiscal policy, see Auerbach and Hassett (1998).

[12] Auerbach (1996) discusses dynamic scoring and the associated issues in more detail.

all available information. But it would require the use of more speculative forecasting procedures, to the extent that reasonable forecasts easily might differ not only in magnitude but also in the sign of estimated policy feedback effects.

Attempting to carry out dynamic scoring in an environment in which forecasts already have statistical difficulties, are produced under political pressure, and are relied on without sufficient caution seems ill-advised, a point that has been recognized for some time. For example, Penner (1982) advocates the use of very mechanical rules for constructing official forecasts, not because they produce the most accurate forecasts, but because there will be little disagreement about how the forecasts should be constructed, and hence little bias in the process. The Appendix 1 below presents a simple model that formalizes this trade-off, confirming that the use of more ambitious, and less easily monitored, forecasting methods should hinge on how uncertain these methods are and how much additional information they have the potential to impart.

Fourth, given the current environment in which forecasts are produced, an attractive evolution of the government forecasting process may be the further development of a parallel, and more ambitious, "unofficial" forecasting approach. An illustration is the long-term budget forecasts produced in recent years by CBO (1997), incorporating macroeconomic feedback, long-term projections, and, to some extent, uncertainty. These forecasts have arisen because they serve an important purpose, helping us to understand the long-run fiscal effects of factors such as population aging and the growth of medical expenditures. But they are even less suited than short-run forecasts to a budget process that ignores uncertainty and, inevitably, applies political pressure. If they can remain unhindered by the constraints of budget rules of the type presently in effect, the development of such forecasts actually might provide information of use to thoughtful policy design.

Ultimately, we must confront the fact that budget forecasts currently serve two distinct purposes that are inconsistent, as summary statistics of available information and inputs to the policy process. If we are not able to alter the nature of this second function, then we face a challenge in performing the first. To do so, perhaps it is time to apply to fiscal policy what we have learned about the benefits of an independent monetary authority, and provide some additional autonomy and protection to those in government charged with providing the budget forecasts.

3.6 Empirical Evidence from Japan

(to be added)

Appendix 1[13]

This appendix presents a simple, static model that may be used to illustrate the trade-off that may exist in asking more from the forecasting process, as in the case of "dynamic scoring".

Suppose that there is a basic information set, say Ω, which is commonly observed by all. On the basis of this information, the expected value of revenue is x, is $\bar{x}_{\Omega} = E(x/\Omega)$. One can think of $\bar{x}_{\Omega}$ as the prediction of a relatively simple, commonly understood forecasting methodology. Let us also assume that the forecasting agency has access to a more comprehensive information set, say Π (fo which Ω is a subset), that allows more precise forecasts. The additional information included in Π may be viewed as the greater accuracy of a more sophisticated forecasting process that is not transparent or easily verified, such as the incorporation of dynamic feedback effects. This greater accuracy means that, if the true value of x equals the prediction $\bar{x}_{\Pi} = E(x/\Pi)$ plus a zero-mean stochastic error term, ε, then there is an additional, independent, error term, v, involved when forecasting x with the information set Ω, equal to the error in forecasting x_{Π}. That is, $x = x_{\Pi} + \varepsilon = \bar{x}_{\Omega} + v + \varepsilon$.

Imagine that the government (as distinct from the agency) wishes to ensure that the agency's estimates are as accurate as possible, as represented by minimizing the value of a loss function of its expected squared deviation, $L = E[(\hat{x} - x)^2 | \Omega]$, of actual revenue, x, from that predicted by the agency to use all its own available information, Π, in formulating $\hat{x}$. However, if the agency's forecasting process is biased, its use of this superior information will not result in a forecast equal to the expected value, $\bar{x}_{\Pi}$.

To make this point concrete, suppose that the agency desires to minimize its own loss function, $\Lambda = E[\gamma(\hat{x} - x - \theta)^2 | \Pi]$, where θ represents the bias in its forecasting process. This would lead to a forecast of $\bar{x}_{\Pi} + \theta$. If θ were observable, the bias would present no problem for the government, which could then make the appropriate adjustment to the agency's biased forecast to recover $\bar{x}_{\Pi}$. But, as it may be difficult to know what the inherent forecasting bias is, it makes sense to treat θ as a random variable from the governmnet's viewpoint. For simplicity, we also let the mean of θ equal 0, for, as just shown, the deterministic part of θ is unimportant.

The government faces a difficult choice in deciding whether to let the agency use its "superior" forecasting process, for this will then also open the door to the inclusion of bias. To see how different factors affect this trade-off, suppose that the government may influence the extent to which the agency bases its forecast on Ω, rather than Π, by imposing a penalty on the agency, $P = \beta(\hat{x} - \bar{x}_{\Omega})^2$, determined by the deviation of the agency's forecast from that

---

[13] This appendix draws from Auerbach (1999, pp.781-2).

based on common information. Setting $\beta = 0$ will lead the agency to use Π to minimize its own loss function, Λ, while setting $\beta = \infty$ will cause the agency simply to report the common forecast, $\bar{x}_\Omega$. More generally, its choice of $\hat{x}$ to minimize the sum of its own loss function and the additional penalty, $\Lambda + P$, will be the weighted average, $\beta' \bar{x}_\Omega + (1-\beta')(\bar{x}_\Pi + \theta)$, where $\beta' = \beta/(\beta+\gamma)$ ranges from 0 to 1 as β ranges from 0 to ∞. It is straightforward to show that the value of the relative penalty β' that minimizes the government's expected loss function, L, is $V(\theta)/[V(\theta)+V(v)]$, the ratio of the variance of θ to the sum of this variance and the variance of v.

Thus, the agency should be encouraged to use its superior information, the greater this informational advantage is (i.e., the larger $V(v)$ is), and the less unpredictable the influence of bias on its unobservable forecasting process (i.e., the smaller $V(\theta)$ is).

Appendix 2 Box – Jenkins Forecasting Methods[14]

Provided procedures for the fitting of autoregressive integrated moving average (ARIMA) models, and their seasonal variants, to a particular time series were known. It will be recalled that the fitting procedures consisted of an iterative cycle of identification, estimation, and diagnostic checking. Essentially a particular model is chosen from the general ARIMA class, its coefficients estimated, and its adequacy of representation checked, possibly leading to the choice of an alternative form and a repeat of the model building cycle. In this section, it will be shown how forecasts can be generated from the fitted models. The whole process of constructing an ARIMA model and the generation of forecasts from that model will be referred to as the Box – Jenkins forecasting method since, although a number of elements in the methodology were well known before these authors wrote, it is their contribution that has allowed an integrated and well-defined approach to time series forecasting via model building stimulating a good deal of practical application over a wide range of actual time series.

As a first step, a number of known results are summarized.

(i) Let X_t follow the stationary, invertible ARIMA (*p*,*q*) process

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}, \quad b_0 = 1 \tag{4}$$

Standing at time *n*, let $f_{n,h}$ be the forecast of X_{n+h} which has smallest expected squared error among the set of all possible forecasts which are linear in $X_{n-j}, j \geq 0$. Now write

$$X_{n+h} = \sum_{j=1}^{p} a_j X_{n+h-j} + \sum_{j=0}^{q} b_j \varepsilon_{n+h-j}, \quad b_0 = 1 \tag{5}$$

Then a recurrence relation for the forecasts $f_{n,h}$ is obtained by replacing each element in (2) by its "forecast" at time *n*, as follows:

(a) replace the unknown values X_{n+k} by their forecast $f_{n,k}$ for $k > 0$;
(b) "forecasts" of $X_{n+k}, k \leq 0$, are simply the known values x_{n+k};
(c) since ε_t is white noise, the optimal forecast of $\varepsilon_{n+k}, k > 0$, is simply zero;
(d) "forecasts" of $\varepsilon_{n+k}, k \leq 0$, are just the known values ε_{n+k}.

(ii) The ARIMA (*p*,*q*) process $a(B)X_t = b(B)\varepsilon_t$ can be written as an infinite moving average $X_t = c(B)\varepsilon_t$ where the elements of $c(B) = c_0 + c_1 B + c_2 B^2 + \cdots$ can be obtained by

[14] This section draws selectively from Granger and Newbold (1986, Chapter5).

equating coefficients of B^j, $j = 1,2,\ldots,$ in $a(B)c(B) = b(B)$. Then the forecast errors are given by

$$e_{n,h} = X_{n+h} - f_{n,h} = \sum_{j=0}^{h-1} c_j \varepsilon_{n+h-j} \tag{6}$$

and hence the variances of the forecast errors are given by

$$V(h) = E\{e_{n,h}^2\} = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} c_j^2 \tag{7}$$

(iii) An "updating" formula for the forecasts is given by

$$f_{n,h} = f_{n-1,h+1} + c_h (X_n - f_{n-1,1}) \tag{8}$$

These results hold for stationary time series. However, it is very often the case in dealing with economic time series that the integrated model is of importance. That is, for a particular process X_t, one may need to difference d times to produce a stationary series $Y_t = (1-B)^d X_t$. In the case of seasonal time series, it is often the case that a multiplicative difference filter is required to produce stationarity. However, for the present purposes, this involves no new principle, and so for simplicity of exposition, attention will be restricted to the nonseasonal case. A sensible procedure, then, is to derive forecasts of the series *X* from those of the stationary series *Y*. $f_{n,h}^x$ for forecasts of *X* and $f_{n,h}^y$ for forecasts of *Y*, an obvious formula for generating forecasts of *X* is then

$$f_{n,h}^y = (1-B)^d f_{n,h}^x \tag{9}$$

where here *B* operates on the index *h*, so that, for example, in the case *d*=1,

$$f_{n,h}^x = f_{n,h-1}^x + f_{n,h}^y \tag{10}$$

Thus, forecasts could be obtained by a two-step procedure, where the stationary series *Y* is first forecast and then forecasts of *X* are obtained from (6). However, a moment's reflection should indicate that this is unnecessary. Write

$$(1 - A_1 B - A_2 B^2 - \cdots - A_P B^P) = a(B)(1-B)^d \tag{11}$$

Then, corresponding to (2), one can write

$$X_{n+h} = \sum_{j=1}^{P} A_j X_{n+h-j} + \sum_{j=0}^{q} b_j \varepsilon_{n+h-j}, \quad b_0 = 1 \tag{12}$$

and forecasts may be derived from this equation in the same way as in the stationary case.

Now, define $C(B) = 1 + C_1 B + C_2 B + \cdots$ where

$$A(B)C(B) = b(B) \tag{13}$$

Then, it is clear that, corresponding to (3) and (4), the forecast errors in the integrated case are given by $e_{n,h} = \sum_{j=0}^{h-1} C_j \varepsilon_{n+h-j}$, and hence the error variance is

$$V(h) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} C_j^2, \quad \mathrm{C}_0 = 1 \tag{14}$$

Further, the updating formula (5) is now given by

$$f_{n,h} = f_{n-1,h+1} + C_h (X_n - f_{n-1,1}) \tag{15}$$

All the necessary equipment for the efficient computation of point forecasts from a fitted ARIMA model, or its seasonal variant, is now at hand. Further, since an expression for the variance of forecast error has been derived, it is possible, provided distributional assumptions are made, to derive interval forecasts. In the remainder of this section, practical procedures for the computation of these forecasts are outlined and the methods involved illustrated with a few specific examples.

Initial Calculation of Point Forecasts

Suppose now that one has a set of observations $x_1, x_2 \ldots, x_n$ on a process *X*, an that an ARIMA model has been fitted. Then forecasts of future values of the series can be obtained from (5), substituting in that equation forecasts of each individual term. The forecasting formula can then be written as

$$f_{n,h} = \sum_{j=1}^{P} A_j f_{n,h-j} + \sum_{j=0}^{q} b_j \hat{\varepsilon}_{n+h-j}, \quad b_0 = 1 \tag{16}$$

where

$$\begin{aligned} f_{n,h} &= x_{n+k}, \quad k \le 0 \\ \hat{\varepsilon}_{n+k} &= 0, \quad k > 0 \\ &= \text{estimate of } \varepsilon_{n+k}, \quad k \le 0 \end{aligned} \tag{17}$$

Equations (17) require some further explanation. The theoretical development concerned the forecasting of X_{n+h} given $X_{n-j}, j \ge 0$. That is, it was assumed that an infinite past record of the series to be forecast was available, in which case $\varepsilon_{n-j}, j \ge 0$ would also be known. However in the practical situation, where only a finite run of data is available, the ε_{n-j} will not be known, but must be estimated.

In practice, we employ as estimates the residuals $\hat{\varepsilon}_{n-j}$ from the fitted model. These residuals, are routinely produced by model estimation programs.

Once a forecasting model has been estimated, the procedure for deriving point forecasts is then quite straightforward. Equation (16) is employed one step at a time for *h*=1,2,3,…, substituting appropriate values from (17). Hence, for *h*=1, the forecast $f_{n,1}$ is obtained immediately from (16). Next, setting *h*=2, in (16), the two-steps-ahead forecast $f_{n,2}$ is obtained using $f_{n,1}$, which has already been calculated. Forecasts can then be obtained as far ahead as is required. To illustrate these calculations, some of the examples will be further considered.

Calculation of Interval Forecasts

While it is almost certainly the case that more attention is paid to point forecasts than to any others, it is generally worthwhile to calculate wherever possible confidence intervals associated with these forecasts, if only to provide an indication of their likely reliability. Now, the variance of the error of the point forecast is given by (14), where the c_j are defined in (13). In fact this is an *underestimate* of the true variance since it assumes that the coefficients of the forecasting model are known, whereas in fact they must be estimated leading to a corresponding decrease in accuracy in the resulting forecasts. However, for moderately long time series, this factor will be of relatively small importance.

In fact, our proposed procedure for estimating forecast error variance is a somewhat crude approximation for estimating forecast error variance is a somewhat crude approximation for two reasons. We do not in fact know the infinite past of the time series, and the parameters of the model must be estimated. The quality of these estimates is examined in some detail by Ansley and Newbold [1981] who provide a modified estimate that tends to be somewhat more reliable, especially for relatively short seasonal time series. The details of this procedure are a little cumbersome and will not be discussed further here. For most general purposes it should prove adequate to substitute the parameter estimates in (14).

If, in addition, one is prepared to assume that the forecast errors come from a normal distribution, it is possible to derive, in an obvious way, confidence intervals for the forecasts. Thus an approximate 95% interval is given by

$$f_{n,h} \pm 1.96\hat{\sigma}_{\varepsilon}\sqrt{\sum_{j=0}^{h-1} c_j^2} \tag{18}$$

where $\hat{\sigma}_{\varepsilon}$ is the estimated standard deviation of ε_t, obtained in estimating the coefficients of the fitted model. Similarly, approximate 75% intervals are given by

$$f_{n,h} \pm 1.15\hat{\sigma}_{\varepsilon}\sqrt{\sum_{j=0}^{h-1} c_j^2} \tag{19}$$

Updating the Forecasts

Once a forecasting equation has been built, it is generally not necessary to refit a model when a new piece of data becomes available. Neither is it necessary to employ the rather lengthy procedures just described to recomputed forecasts. A convenient algorithm, based on (15), is available for the updating of previously computed forecasts. Writing n+1 for n in that expression, produces

$$f_{n+1,h} = f_{n,h+1} + c_h(X_{n+1} - f_{n,1}) \tag{20}$$

where $X_{n+1} - f_{n,1} = \varepsilon_{n+1}$ is the error made in forecasting X_{n+1} at time *n*. In words, the forecast of X_{n+1+h} made at time *n*+1 can be obtained by adding to the forecast of the same quantity, made at time *n*, a multiple of the error made in Forecasting X_{n+1} at time *n*. Further, the weights C_h required in this expression will already be known since they will have been calculated for the derivation of interval forecasts. The forecast generating mechanism can thus be viewed as an "error learning process". That is to say, forecasts of future values of the series are modified in the light of past forecasting errors. It is not, of course, necessary to recomputed estimates of the widths of appropriate confidence intervals since the estimated variance for an *h*-step ahead forecast does not change when another observation becomes available.

To illustrate, consider once again the series on construction begun in England and Wales. The forecast of X_{n+1} made at time *n* was $f_{n,1} = 91747$. In fact, the actual value turned out to be $X_{n+1} = 86332$. Thus the point forecast turned out to be an overestimate. The one-step ahead forecast error is

$$X_{n+1} - f_{n,1} = 86332 - 91747 = -5415 \qquad (21)$$

Thus, using (20), the forecasts can be updated by the formula

$$f_{n+1,h} = f_{n,h+1} - 5415C_h \qquad (22)$$

where the *C* weights are given. Thus, for example, the revised forecast of X_{n+6} is

$$f_{n+1,5} = f_{n,6} - 5415C_5 = 123814 - (5415)(0.83) = 119320 \qquad (23)$$

The updated forecasts for this series are given in Table 1.

Table 2 Updated Forecasts of Construction Begun Series

h:	1	2	3	4	5	6	7	8	9	10	11
$f_{n+1,h}$:	112555	106413	102911	94452	119320	113179	109676	101217	126086	119945	116502

Forecast Errors as a Check for Change in Model Structure

If an ARIMA model has been fitted to a moderately long series of data, it is not necessary to go to the trouble of refitting the model each time a new piece of data becomes available. Rather, the originally estimated model can be retained and forecasts updated in the manner just described. However, this procedure would not be appropriate if the model structure were to change. If such a change is suspected, a check can be based on the forecast errors, following a proposal of Bhattacharyya and Andersen [1974] and Box and Tiao [1976].

Let $f_{n,h}$, $h = 1,2,\ldots,H$ be forecasts of X_{n+h}, $h = 1,2,\ldots,H$, all made at time *n*, with errors

$$e_{n,h} = X_{n+h} - f_{n,h}, \qquad h = 1,2,\ldots,H$$

Then, assuming the model structure has remained unchanged,

$$E[e_{n,h}] = 0, \qquad h = 1,2,\ldots,H$$

and,

$$E(e_{n,h}e_{n,h+k}) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} C_j C_{j+k}, \qquad k \geq 0 \tag{24}$$

Now, let

$$E[e_{n,i}e_{n,j}] = \sigma_\varepsilon^2 v_{ij}, \qquad i = 1,2,\ldots,H, \qquad j = 1,2,\ldots,H$$

and **V** be the $H \times H$ matrix whose $(i,j)th$ element is v_{ij}. Then, assuming normality of the forecast errors and an unchanged model structure, the quantity

$$Q = \sigma_\varepsilon^{-2} \mathbf{e}'_H \mathbf{V}^{-1} \mathbf{e}_H \tag{25}$$

is distributed as x^2 with *H* degree of freedom, where

$$\mathbf{e}'_H = (e_{n,1}, e_{n,2} \ldots, e_{n,H})$$

Box and Tiao show that a computationally simpler, but equivalent, version of (22) is

$$Q = \sigma_{\varepsilon}^{-2} \sum_{j=0}^{H-1} e_{n+j,1}^{2} \tag{26}$$

where the $e_{n+j,1}$ are the one-step errors made at time $n + j(j = 0,1,\ldots,H-1)$.

Thus, the hypothesis of no change in model structure can be checked by computing (26), with the estimated error variance $\hat{\sigma}_{\varepsilon}^{2}$ in place of the unknown σ_{ε}^{2}, and comparing with tabulated values of x^2.

Stepwise Autoregression

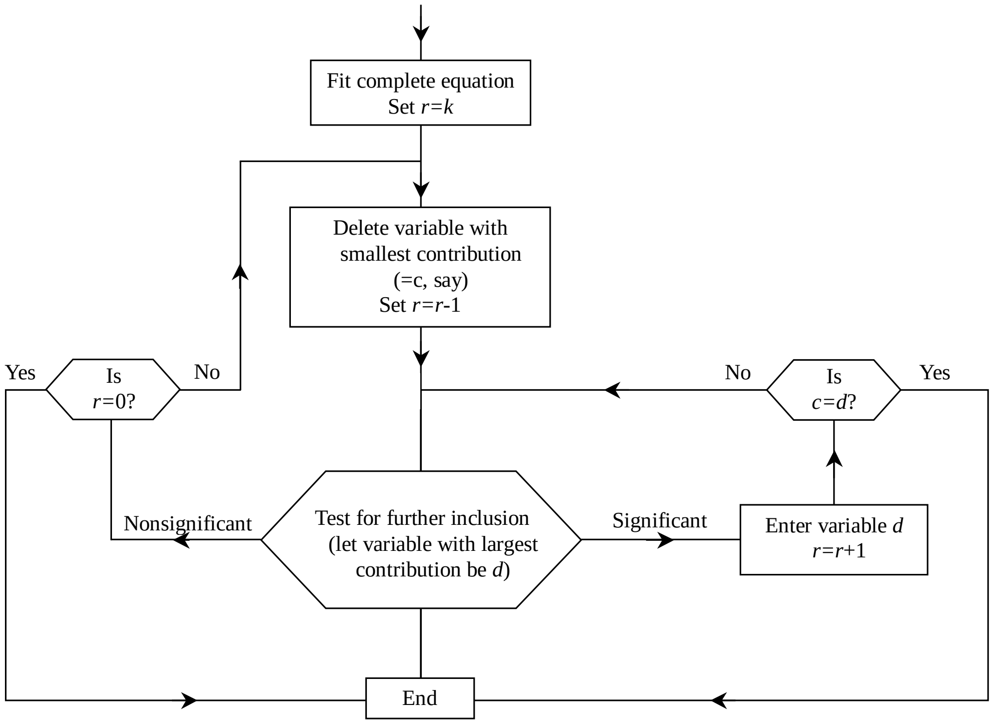
The basic exponential smoothing procedures discussed in the previous section generally postulate a single model from which forecasts are to be generated, and thus do not possess the great virtue of the Box-Jenkins approach, whereby the eventual form of the forecast function is dictated ,through the processes of identification and diagnostic checking, by the data itself. Of course, there is some room for experimentation within Brown’s generalized exponential smoothing framework, but even here there does not exist any clear-cut identification procedure (In addition, the restriction to a single parameter renders this approach overly parsimonious in many situations). The identification and diagnostic checking phases of the Box-Jenkins cycle require manual intervention, however, and it would be desirable for some routine forecasting purposes to eliminate such a requirement. A compromise might be achieved through the design of a forecasting procedure which, while remaining fully automatic, contained a mechanism for discriminating among various possible forms of forecast function. That is, one would like a system to contain an identification procedure which was itself fully automatic. One method for achieving this, briefly introduced by Newbold and Granger [1974], is via stepwise autoregression. The objective is to construct autoregressive models to describe the behavior of given time series. However, for economic data, it is preferable to work with changes $Y_t = X_t - X_{t-1}$ rather than with levels of the series. Consider, now, the general *k*th order autoregressive model

$$Y_t = \sum_{j=1}^{k} a_j Y_{t-j} + \varepsilon_t \tag{27}$$

Typically, models of the form (27) can, as has been seen, easily be fitted to a given set of data. However, unless *k* is taken to be quite small, it is likely that the resulting model will be overparametrized. One way out of this dilemma is to employ the technique of stepwise regression. This has been studied in great detail by Payne [1973], and the treatment given here depends heavily on Payne's work.

One way to proceed is to first select the value Y_{t-j} which, one the criterion of residual sum of squares, contributes most toward "explaining" Y_t. At the second step, the lagged value that most improves the fit of the regression equation obtained at step one is added, and so on until addition of further variables produces no significant improvement in the fit of the regression. Variables, entered at an earlier stage, which cease to contribute significantly, can be dropped. An alternative, favored by Payne on the basis of his experience with a number of simulation experiments, is to proceed in the reverse direction, having initially fitted the complete model (27). At the first step, the lagged-value contributing least to overall explanation of Y_t is dropped from the regression. At the second step the lagged value that contributes least in the model so achieved is dropped, and so on until deletion of further terms significantly worsens the fit of the regression equation. Variables, dropped at an earlier stage, can later be added if doing so would produce a significant improvement in the fit of the achieved regression. The procedure is set out schematically in Figure 1. One might also include a constant term in the formulation (27). The constant could be treated as any other variable within the stepwise framework, or alternatively a decision as to its inclusion or exclusion could be made on purely subjective grounds.

Figure 1 Payne's scheme for stepwise regression procedure



Having decided to undertake the fitting of an autoregressive model by stepwise regression methods, three decisions must be taken:

(i) A value k for the maximum permitted lag in (27) must be chosen.

(ii) A significance level for testing for inclusion or exclusion of further variables must be decided upon.

(iii) An appropriate hypothesis test to determine suitable stopping rules needs to be determined.

Choice of the maximum contemplated lag *k* is likely to be dictated in part by the nature of the time series under study and by the amount of data available. Our experience indicates that for all nonseasonal series, for quarterly seasonal series and shorter monthly seasonal series, a value *k*=13 is generally adequate. For longer monthly seasonal time series, a value *k*=25 is preferable. The question of choosing a suitable significance level for testing variable inclusion or exclusion is by no means a trivial one form a theoretical viewpoint. Looked at in this light, it might be desirable to reflect in one's choice preconceived notions as to how simple a model (in terms of number of parameters) is likely to provide reasonable forecasts. It is possible that one would like the significance level to vary according to how many lagged values have already been included in the model. Nothwithstanding these considerations, however, we have found use of a constant 5% level to be adequate (in terms of forecasting accuracy of the resulting model) for most general purposes.

Testing of hypotheses presents one critical difficulty. Suppose that a stage has been reached where *r* terms are included in the regression. Clearly if each of the remaining *k*-*r* terms was tested for further inclusion at the 5% level, the probability of finding at least one term that apparently significantly improved the fit would be greater than 0.05 even if the true associated coefficient valued were also zero. Payne suggests a number of procedures for overcoming this problem, and prefers use of the statistic

$$F = \frac{m-r}{k-r}\frac{V_r}{V_k} - \frac{m-k}{k-r} \tag{28}$$

where

$$V_j = \frac{\text{residual sum of squares when } j \text{ terms are included in teh regression}}{m-j} \tag{30}$$

and *m* is the effective number of observations for regression, so that if the original sample is $x_1, \ldots, x_n$, one observation is "lost" by differencing and a further *k* by formulation of the autoregressive model (24), so that $m = n - k - 1$. Under the null hypothesis that the coefficients on the excluded variables are all zero, the statistic *F'* is distributed approximately as Fisher's *F* with *k*-*r* and *m*-*k* degree of freedom (One is justified asymptotically in employing the usual normal theory regression tests in the context of autoregressive models as a result of Mann and Wald [1943]).

Stepwise autoregression, then, would appear to provide a reasonable alternative to exponential smoothing as a fully automatic forecasting technique. Its great advantage lies in the wide class of models contemplated, together with a built-in identification structure. Calsulation of forecasts from the fitted model is straightforward along the lines described above.

References

Auerbach, Alan J. (1994) “The U.S. Fiscal Problem: Where We Are, How We Got Here, and Where We’re Going.” *NBER Macroeconomics Annual*, 9: 141-75.

Auerbach, Alan J. (1995) “Tax Projections and the Budget: Lessons from the 1980s.” *American Economic Review*, 85, No.2 (May): 165-9.

Auerbach, Alan J. (1996) “Dynamic Revenue Estimation.” *Journal of Economic Perspectives*, 10, No.1 (Winter): 141-57.

Auerbach, Alan J. (1999) “On the Performance and Use of Government Revenue Forecasts.” *National Tax Journal*, 211(4), pp.767-82.

Auerbach, Alan J. and Kevin A. Hassett (1998) “Uncertainty and the Design of Long-Run Fiscal Policy.” UC Berkeley Burch Center working paper, No.B98-01 (November)

Campbell, Bryan, and Eric Ghysels (1995) “Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency.” *Review of Economics and Statistics*, 77, No.1: 17-31.

Congressional Budget Office (1997) A*n Economic Model for Long-Run Budget Simulations*. Washington, D.C.,: Congressional Budget Office, July.

Ehrbeck, Tilman, and Robert Waldmann (1996) “Why are Professional Forecasters Biased? Agency versus Behavioral Explanations.” *Quarterly Journal of Economics*, 111, No.1 (Feb): 21-40.

Granger C.W.J. and P. Newbold (1986) *Forecasting Economic Time Series*, 2nd ed., Academic Press.

Holmström, Bengt (1982) “Moral Hazard in Teams.” *Bell Journal of Economics*, 13, No.2, (Autumn): 324-40.

Penner, Rudolph (1982) “Forecasting Budget Totals: Why Can’t We Get It Right?” In the *Federal Budget: Economics and Politics*, edited by A. Wildavsky and M. Boskin (San Francisco: Institute for Contemporary Studies): 89-110.

Plesko, Gerorge A. (1988) “The Accuracy of Governemnt Forecasts and Budget Projections.” *National Tax Journal*, 41, No.4 (December): 483-501.

Scharfstein, David S., and Jeremy C. Stein (1990) “Herd Behavior and Investment.” *American Economic Review*, 80, No.3 (June): 465-79.

Zweibel, Jeffrey (1995) “Corporate Conservatism and Relative Compensation.” *Journal of Political Economy*, 103, No.1 (Feb): 1-25.