

IMF Working Paper

Tax Revenue Forecasts in IMF-Supported Programs

Mikhail Golosov and John King

IMF Working Paper

Fiscal Affairs Department

Tax Revenue Forecasts in IMF-Supported Programs

Prepared by Mikhail Golosov and John King¹

Authorized for distribution by Michael Keen

December 2002

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

Year-ahead forecasts of tax revenues incorporated into IMF programs for low-income countries, from 1993 to 1999, are compared with the corresponding outturns. The accuracy of these forecasts was low, with a mean absolute percentage error of 16 percent. Forecasts of tax revenues as a percentage of GDP were biased upwards, but there was no significant bias in forecasts of nominal tax revenues. Upward bias in the tax revenue forecasts was associated with subsequent interruptions to the program, and the length of time between the commencement of the program and the beginning of the year for which the forecast was made.

JEL Classification Numbers: H29, O23

Keywords: Forecasting; tax revenues; IMF-supported programs

Authors' E-Mail Addresses: Golosov@econ.umn.edu; Jking@imf.org

¹ This is a revised version of a paper presented at a Fiscal Affairs Department seminar in August 2002. Thanks are due to Jun Il Kim and Howell Zee for very helpful comments, and to Asegedech Woldemariam for assistance with data preparation.

Contents	Page
I. Introduction.....	3
II. Assessing Tax Revenue Forecasts	5
A. The Purpose of Tax Revenue Forecasts	5
B. Tax Revenue Forecasts in IMF Programs	6
C. The Construction of Tax Revenue Forecasts	8
III. Data	9
IV. Characteristics of the Forecast Errors	10
A. Normality and Serial Correlation	11
B. Bias.....	12
C. Accuracy.....	13
V. Exploring the Sources of Bias.....	16
A. Success of the Program	16
B. Conditionality Related to Taxes	19
C. Geographical Differences.....	19
D. Forecast Horizon	19
VI. Conclusion	20
Text Tables	
1. Forecast Errors: Descriptive Statistics	10
2. Regression of Errors in Tax Ratio Forecasts on Errors in GDP Forecasts	13
3. Theil's U Statistics	15
4. Directional Accuracy of the Tax Ratio Forecasts	15
5. Regressions to Account for the Forecast Errors	17
6. Explanatory Variables: Descriptive Statistics.....	17
7. Tax Ratio Forecast Errors: Effects of Measures of Program "Success".....	18
Figures	
1. Errors in Forecast of Tax as a Percent of GDP	10
2. Errors in Forecast of Growth in Revenue	11
References.....	22

I. INTRODUCTION

This paper assesses the tax revenue forecasts² that formed part of financial programs supported by the IMF's concessional lending facilities during the period 1993–99.³ It focuses on forecasts of total tax revenues for the coming fiscal year, such as would typically be incorporated into the government's budget for that year.

The paper addresses three basic questions:

- How accurate were these forecasts of tax revenues?
- Were they biased, and if so, what was the nature of the biases?
- Were the observed forecast errors or biases associated with any particular characteristics of the programs?

Previous research into the IMF's concessional lending programs in the period 1985–95 has established that year-ahead forecasts of total tax revenues as a percentage of GDP exceeded outcomes in those years, on average, by 0.6 percentage points.⁴ This suggests some upward bias in the forecasts. This research did not, however, investigate the statistical significance of this bias, or the variability of the forecast errors. Furthermore, its focus was exclusively on forecasts and outturns of tax revenues **as percentages of GDP**: it did not look at errors in the forecasts of nominal amounts of tax revenues—the numbers that would typically be included in the annual budgets of the program countries.

More recent research comparing projections and outcomes of IMF-supported programs has investigated the issues of accuracy, bias, and efficiency that are the focus of the present paper.⁵ The forecast aggregates that were covered by this research were, however, confined to GDP growth, CPI inflation, the current account balance, net capital inflows, and the change in official reserves: fiscal variables were not included. Furthermore, the programs covered by the study were confined to Stand-By and Extended Arrangements (SBAs and EFFs) during the

² The relevant numbers—denoting tax revenues that are to be collected in future periods—are sometimes referred to in the context of IMF programs as “targets” (e.g., in Abed et al., 1998) or “projections” (e.g., in Musso and Phillips, 2002), rather than “forecasts.” There are subtle differences of meaning between these terms, but they are not precise, and none of the terms is entirely satisfactory for our present purposes. We use the term “tax revenue forecasts” in this paper because that is the term most commonly used in the context of government budgeting and because our purpose is to apply to the numbers some conventional tests of “forecasting accuracy.” The status of the numbers is discussed a little further in Section II C below.

³ All of the financial programs included in the study were programs under the Enhanced Structural Adjustment Facility (ESAF), which was replaced by the Poverty Reduction and Growth Facility (PRGF) in 1999.

⁴ Abed et al. (1998).

⁵ Musso and Phillips (2002).

period 1993–97; ESAF programs, which were the main form of arrangement between the IMF and about 80 of its member countries with the lowest incomes per capita, were excluded.

In the wider literature on economic forecasting, the accuracy of tax revenue forecasts has received much less attention than the accuracy of GDP forecasts. However, a few comparative results have been reported. Mean absolute percentage errors (MAPEs)⁶ of year-ahead forecasts of budgetary revenues in the period 1982–92 were a little under 2 percent in Australia and the Netherlands, around 3 percent in the United States (both for forecasts made by the Office of Management and Budget for the administration, and for those made by the Congressional Budget Office for the legislature), and around 3.7 percent in Canada (Ernst & Young, 1994). For the ten Canadian provinces in the period 1981–96, the average MAPE of budget revenue forecasts was 3.3 percent, with a range from 1.55 percent in Quebec to 7.71 percent in Alberta (Jenness and Arabackyj, 1998). For a sample of 20 U.S. state legislatures in the period 1985–1992, the average MAPE was 4.5 percent (Mocan and Azad, 1995).

Under- or over-prediction of tax revenues in government budgets persisting over a period of years has emerged as a problem in several developed countries in recent years. In the United States, tax revenue forecasts were generally too high in the 1980s—but too low from the mid-1990s onwards (Auerbach, 1999; Penner, 2001). In Canada, an apparent upward bias in revenue forecasts that was identified in the mid-1990s led to radical changes to budget procedures (Ernst & Young, 1994; Finance Canada, 1999). In the United Kingdom, persistent overprediction of revenues was also identified as a problem in the mid-1990s, but the problem was confined to revenues from the value added tax (VAT) (United Kingdom, 1997). In Ireland, the opposite experience—of persistent underprediction of revenues—led to a thorough review of forecasting procedures (Ireland, 2000).

Studies of the possible sources of error, or bias, in tax revenue forecasts have largely been confined to the states of the United States. The effect on forecast errors of a variety of political and institutional factors (mostly specific to the U.S. context), and of the nature of forecasting methods employed, has been investigated. The results obtained have sometimes been in conflict. The most recent study of this kind found no evidence that political factors affect forecasting errors; it did find, however, that state legislatures using purely judgmental methods of forecasting (as opposed to quantitative techniques of various kinds) had significantly larger forecast errors, by as much as 3.8 percentage points (Mocan and Azad, 1995).

The rest of this paper is organized as follows. Section II discusses a variety of features of tax revenue forecasting that are likely to have implications for bias and accuracy. Section III describes the data used in the study, and Section IV analyzes the errors in these forecasts. Section V then investigates relationships between these forecast errors, and a variety of characteristics of the program forecasts. Section VI concludes.

⁶ Mean absolute percentage error (MAPE) is defined as the mean of the absolute differences between each forecast and the corresponding actual value, expressed as a percentage of the actual value.

II. ASSESSING TAX REVENUE FORECASTS

Certain particular characteristics of tax revenue forecasts have important implications for the assessment of forecasting performance.

A. The Purpose of Tax Revenue Forecasts

Tax revenue forecasts are made by national governments in the course of budget preparation. Usually, they will be made at least twice in the annual budget cycle: (1) at an early stage, a tax revenue forecast may be made—on the assumption of no change in government policies—to help establish the “resource envelope” within which budget decisions will be taken; and (2) in the final stage, a forecast incorporating all the budget decisions on tax changes will be made for inclusion in the budget documents presented to the legislature. The second (budget) forecast will often have subsidiary uses within government: in particular, it is commonly used to set performance targets for revenue departments and agencies.

It would seem obvious in the context of budget preparation that a “good” tax revenue forecast would be one that is as accurate as possible (subject, of course, to the costs of making the forecast), and an unbiased estimate of the most likely outturn.⁷ This presumption has, however, been challenged. If the costs of forecast errors are not symmetrical, it is sometimes argued that the appropriate forecast is not the mean, “expected” value of the probability distribution of possible outcomes: instead, it is the value that minimizes the expected cost of being wrong. On the basis that a given underestimate of next year’s fiscal deficit would be much more “costly” than an overestimate by the same amount, it has been suggested that tax revenue forecasts should deliberately err on the conservative side.⁸

But one must ask: how conservative, then, should they be? If a precise answer can be given to this question—for example, “tax revenue forecasts for budget purposes should be x percent lower than expected outcomes”—then the adjustment seems unnecessary: one could achieve exactly the same result by adding that amount to the budget deficit, in some form of contingency reserve, while leaving the tax revenue forecast at the expected value.⁹ On the other hand, if one cannot answer the question, the prescription would seem to be a recipe for confusion in budget formulation.

The secondary use that many countries make of tax revenue forecasts as performance targets for revenue collection agencies has also provided the basis for arguments that they should

⁷ The *Revised Manual of Fiscal Transparency* (IMF, 2001) refers to the importance of “realistic revenue forecasts” (paragraph 152). Although the term “bias” is not used in the manual, the context suggests that a forecast that is known to be biased would, on that account, not be considered a “realistic” one.

⁸ For example, see Auerbach (1999); and, in the context of Canada, Ernst & Young (1994, p. 141).

⁹ In practice, after several years of experimenting with “prudent” revenue and expenditure forecasts (as recommended in the 1994 Ernst & Young report), this is the approach that the government of Canada adopted in 1999. See *Finance Canada* (1999).

incorporate a deliberate ex ante bias. In general, how such performance targets affect collections can be expected to depend critically on the rewards and penalties that are applied to those who exceed or fall short of the targets. It has been suggested that tax revenue forecasts should be set low in order to encourage revenue collectors to exceed them;¹⁰ more commonly in practice, they may be set high in an attempt to encourage additional efforts from collectors. The second of these effects may seem, perhaps, a little more plausible than the first. But it may reasonably be doubted whether tax revenue targets by themselves—in the absence of specific changes to tax administration procedures—can be expected to have significant revenue effects, particularly, if it were to become known that the targets are **not** unbiased forecasts of the revenues that will be collected in the absence of changes to those procedures.

These arguments in favor of deliberate ex ante bias in tax revenue forecasts seem, therefore, rather unpersuasive. Nevertheless, tax revenue forecasts that are unbiased ex ante may turn out to be biased ex post. In particular, when a forecast is made for the next fiscal year, the final outcome for the present fiscal year is not known with certainty. Hence, last year's forecast error is not known. Furthermore, how much of that unknown forecast error represents systematic error in the forecast procedure (which needs to be corrected), and how much is random, cannot be known with certainty. In this situation, if forecasters adopt a Bayesian approach in adjusting their forecast procedures as new information becomes available, serial correlation in forecast errors is likely to occur: ex post, the forecasts could well be biased upwards or downwards for extended periods.¹¹ Some further reasons why forecasts that are unbiased ex ante may show statistical bias in an ex post evaluation are discussed below, in the context of IMF program countries.

B. Tax Revenue Forecasts in IMF Programs

Tax revenue forecasts made in the context of IMF programs may differ from others in several important respects.

First, IMF programs are generally adopted by countries that are in severe financial or macroeconomic difficulty, and the aim of the program is to overcome such difficulties. Almost by definition, therefore, these programs must show an improved macroeconomic performance over the recent past. It does not follow, of course, that the IMF will tend to be overoptimistic about the extent to which the program policies can bring about that improvement; but this is clearly a possible risk.

Second—closely related to this—IMF programs frequently involve a substantial “fiscal adjustment.” In many cases this includes major changes to the tax system, designed both to improve its efficiency and to increase revenues. In these cases, in addition to the normal

¹⁰ Ernst & Young (1994) suggest, in favor of the adoption of conservative tax revenue forecasts, that “In this way, the government would encourage the discipline of achieving or bettering the forecasts” (p. 21).

¹¹ Penner (2001) argues that this was the main reason for the observed pattern in the United States of overoptimistic revenue forecasts at the federal level in the years 1980–91, and overpessimistic forecasts in the years 1992–2000.

problems of forecasting tax revenues from an unchanged tax structure in an uncertain environment, the forecaster faces the more challenging problems of “revenue estimation”—assessing how revenues will **change** as a result of the introduction of new tax measures.¹²

Third, IMF financial programs are not intended to be unconditional predictions of the most likely outcomes for macroeconomic variables during the program period: they are designed to be consistent forecasts of those variables, on the condition that the policies described in the program—whether or not those policies form part of the formal program conditionality—are carried out. Inevitably, not all program policies will be implemented in practice, and some programs will “fail.” As a result one would expect that, ex post, program variables will differ on average from the corresponding forecasts. The direction of bias that is to be expected will not be uniform, however: for example, real GDP growth should, on average, be below forecast, to the extent that not all programs are successful; on the other hand, inflation should, on average, be above the forecast. Since nominal tax revenues depend on both real GDP and inflation, it is not clear whether one would expect any ex post bias in tax revenue forecasts—and if so, in what direction. One **would** expect, however, that “program failures” will lead to an upward bias, ex post, in forecasts of changes in ratios of tax revenues to GDP.

Fourth, the tax revenue forecasts that form part of the program represent a consensus view reached in discussions between the IMF and the authorities of the country in question: in general, they are neither “IMF forecasts” (although IMF staff must be convinced that they are realistic if they are to form part of the program), nor the unmodified forecasts of the authorities themselves. As a result of this consensus nature, the precise assumptions underlying the tax revenue forecasts and their method of construction are rarely recorded.

Fifth, revenue forecasts incorporated into IMF programs do not always correspond to the government’s budget forecasts for the year (or years) in question. In general, when programs are negotiated, the expectation is that the agreed revenue and expenditure forecasts will form the basis for the budget. IMF programs may, however, be negotiated at any time during the year. Frequently, this will be several months before the relevant budget needs to be drawn up. In this situation, IMF staff will normally discuss the details of the budget with the government at the appropriate time, to ensure its overall consistency with the program; at this stage, some adjustments to the program revenue and expenditure forecasts may be agreed.

Finally, IMF programs include a variety of conditions that may determine whether financial resources are made available to the country. Conditions are normally attached to government’s borrowing from the banking system, and often to the government’s deficit as well. In a few cases, conditionality has also been attached to the performance of tax revenues.¹³ It could well be, therefore, that governments facing a deficit higher than that provided for in the program will

¹² The distinction between “revenue forecasting” and “revenue estimation” is commonly drawn in the United States, where responsibility for the two functions is divided (at least at the federal level). For a detailed account of the revenue estimation process, see Sunley and Weiss (1991).

¹³ More generally, “structural” conditionality in ESAF/PRGF programs often includes tax policy or administration measures, such as the introduction of a VAT or the setting up of a large taxpayer unit.

introduce new measures during the program year, including tax increases, with a view to meeting the program conditions. Once again one would expect that this would result in an upward bias, *ex post*, in the forecast of tax revenues.

C. The Construction of Tax Revenue Forecasts

Tax revenues change from one year to the next in every economy primarily for macroeconomic reasons. Hence, the first step in tax revenue forecasting is generally to prepare a macroeconomic forecast. In many members of the Organization for Economic Cooperation and Development (OECD), this macroeconomic forecast will cover aggregates such as wages and salaries, corporate profits, consumers' expenditure, imports, etc., that are closely related to the "bases" on which taxes are levied; in other countries it may cover only GDP. In both cases, however, the results of the macroeconomic forecast will be crucial inputs into the forecast of tax revenues. Tax revenue forecasting can thus be seen as a two-stage process, consisting of: (1) a macroeconomic forecast; and (2) a tax revenue forecast that is **conditional** on the results of that macro forecast.¹⁴ In many countries, these two stages of the tax revenue forecasting process are performed in different government departments.

Hence, tax revenue forecasts can be evaluated in two different ways. First, the forecast of revenues can be evaluated as a simple, unconditional prediction of the most likely outcome. Second, it can be evaluated as a prediction of tax revenues that is conditional on the accuracy of the relevant macroeconomic variables that provide the basis for the forecast. Which of these two approaches should be adopted depends, of course, on the purpose of the evaluation.

In principle, in order to evaluate a tax revenue forecast that is conditional on a set of macroeconomic variables, one needs to know the forecasts and outcomes for the relevant macro variables, and also the precise manner in which those variables were taken into account in making the tax revenue forecast. That forecast can then be revised, substituting actual for forecast values for the macro variables to construct an alternative tax revenue forecast.¹⁵ This procedure requires, however, that the precise manner in which the tax revenue forecast was derived is known. Often this will not be the case. A much simpler procedure is to construct the alternative forecast on the strong assumption that—other things being equal—total tax revenues will be proportional to nominal GDP. The "conditional" forecast evaluation then focuses on the difference between forecast and actual tax revenues, both measured as percentages of GDP.

¹⁴ Since changes in tax revenues have macroeconomic implications, the process of constructing the forecasts may be more complicated than a simple two-stage process: in practice, it may involve several rounds of iteration between the macro forecasters and those who forecast tax revenues on the basis of the macro forecasts. See, for example, CBO (2001) for a description of tax revenue forecasting procedures in the U.S. legislature, and Pike and Savage (1998) for a description of the procedures in the United Kingdom.

¹⁵ For an example of the application of this procedure in the evaluation of forecasts, see Israel (1996).

III. DATA

Data on tax revenue forecasts were obtained from the IMF's MONA (Monitoring of Fund Arrangements) database, which contains extensive information on the economic objectives and outcomes of IMF-supported programs. During the seven-year period from 1993 to 1999, 45 countries received support under ESAF arrangements—25 in Africa, 9 in Asia, 6 in Central and South America (including the Caribbean), and 5 in Europe. The typical ESAF program was designed to last for three years, with the IMF Executive Board approving an annual arrangement for each year. The MONA database provided data for 126 annual arrangements for these 45 countries during the period covered by this study.

Forecasts of tax revenues taken from this database were compared to the first available “actual” data published in IMF country reports.^{16 17} We study only one-year ahead forecasts, i.e., forecasts made in the year of approval of the annual program for the following year. The annual programs typically cover either a calendar year or a fiscal year, and we use data for the corresponding calendar or fiscal year for the actual outcomes.

As discussed in Section II C above, we focus on two forecast measures of tax revenue: forecasts of tax revenues as a ratio to nominal GDP, and forecasts of changes in tax revenues measured in nominal amounts of national currency. These two types of forecast are of course related, but they have some distinctive characteristics. Arguably, the forecast of tax revenues as a ratio to GDP is a “pure” tax revenue forecast because it does not depend directly on the forecasts of other variables such as inflation. By contrast, a forecast of the change in tax revenue in nominal currency is contingent on the correct projection of at least two other variables, real GDP growth and inflation. This forecast may, however, be more relevant for policy purposes since it is the forecast of nominal tax revenues that is needed for inclusion in the budget.

In the paper, we use the following notation. P_t which is the forecast value of the tax variable for the year t , made in year $t-1$. As appropriate, we use superscript r to denote the forecast of tax revenues as a ratio of (forecasted) GDP, and superscript l to denote the forecast of proportional change in nominal tax revenues. Thus, if TAX_t^r is the forecasted level of tax revenue in nominal currency and TAX_{t-1} is the estimated tax revenue at the time of forecast, the proportional change in nominal tax revenues is defined as $P_t^e = (TAX_t^r / TAX_{t-1}) * 100$. Analogously, A_t is the actual outcome of tax revenues. Forecast errors e_t are defined by $e_t = P_t - A_t$, with superscripts r and l (as appropriate) denoting the error in the forecast of the tax ratio or in the percentage change in nominal tax revenues.

¹⁶ For a similar approach in the context of an assessment of the accuracy of GDP forecasts (for a different group of countries), see Loungani (2001).

¹⁷ Comparing a forecast to the first available number, rather than the latest available one, may be more relevant from a policy perspective since it shows to what extent the forecast can be relied upon in the budget process. Later revisions to the macroeconomic aggregates may be quite large for the ESAF/PRGF countries, where estimates of real and nominal GDP are often very imprecise. Tax revenue statistics are less prone to revision.

IV. CHARACTERISTICS OF THE FORECAST ERRORS

In this section, we assess the accuracy and bias of the program forecasts. The procedures are mostly straightforward, and a summary of the results is given in Table 1. Figures 1 and 2 show histograms of the two measures of forecast error that we are concerned with—respectively, errors in the forecasts of tax revenue as a percentage of GDP, and errors in the forecast of percentage growth in nominal tax revenues.

Table 1. Forecast Errors: Descriptive Statistics

	(A) Error in forecast of tax as percent of GDP: e_t^f		(B) Error in forecast of percentage growth in nominal tax revenue: e_t^1	
Mean	1.01**	(0.22)	0.84	(2.08)
Median	0.9**		2.11	
Standard deviation	2.47		22.84	
Mean absolute error	1.86		16.8	
Normality	145.84**		15.98**	
Serial correlation	0.47**	(0.11)	0.61**	(0.09)

Standard errors in parenthesis. * and ** denote significance at 90 percent and 95 percent confidence level.

Figure 1. Errors in Forecast of Tax as a Percent of GDP

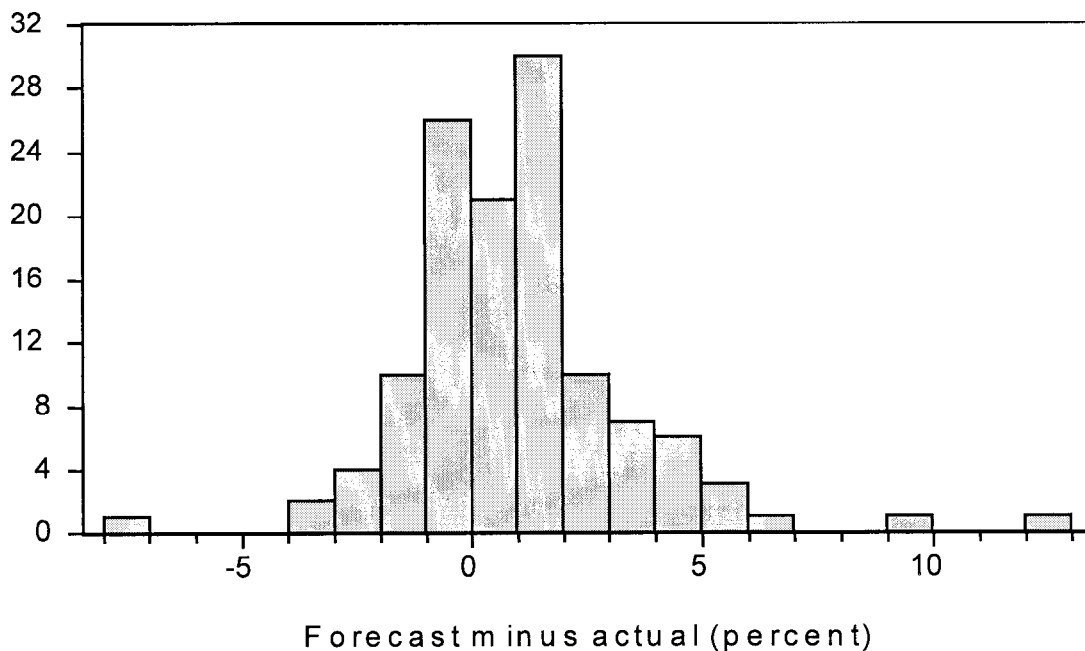
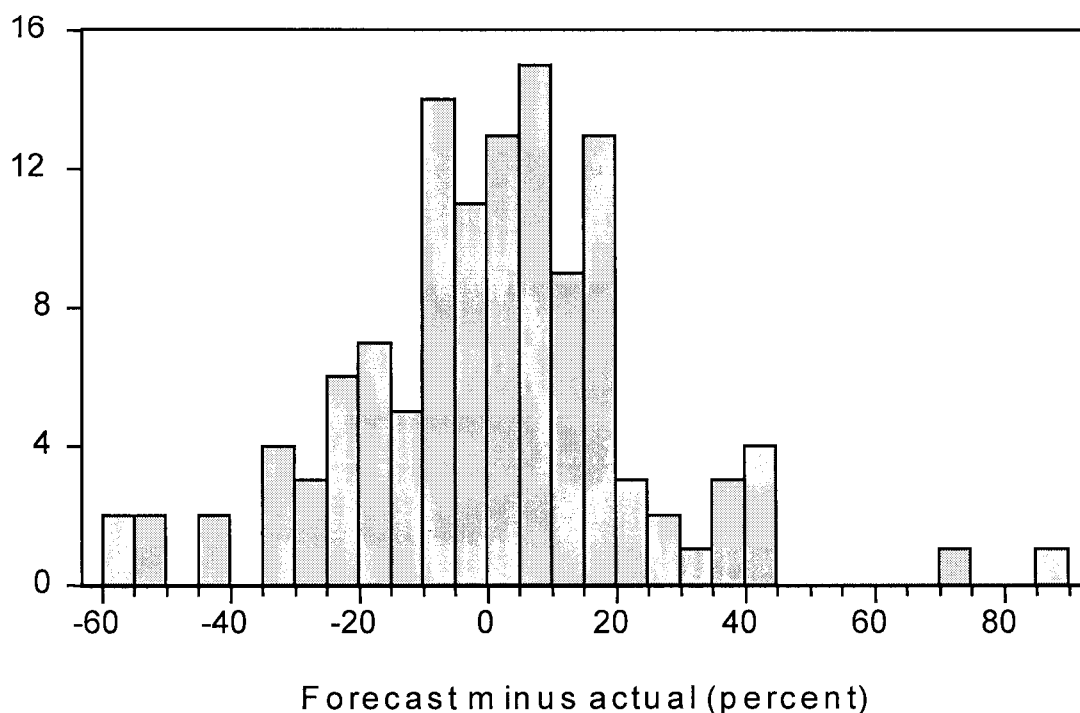


Figure 2. Errors in Forecast of Growth in Revenue



A. Normality and Serial Correlation

Two technical caveats need to be noted at the outset. The first concerns the distribution of the forecast errors. The standard test for normality is the Jarque-Bera test, which measures how significantly the skewness and kurtosis of the density of a series differs from that of a normal distribution. This test decisively rejects the hypotheses that our two series of forecast errors are normally distributed. As suggested by an inspection of Figures 1 and 2, however, this result is heavily influenced by two cases where the tax ratio fell short of the forecast by 12 percent and 9 percent of GDP, respectively. Without those two observations—which occurred during periods of civil conflict—the distribution of forecast errors as a percentage of GDP would not have been significantly different from normal. The errors in the forecasts of growth in nominal tax revenues would still, however, have differed significantly from a normal distribution.

Second, our forecast errors appear to be serially correlated. To detect serial correlation we use a regression of the form:

$$e_t = \alpha + \rho e_{t-1} + \varepsilon_t$$

In the presence of serial correlation, ρ will be significantly different from zero. Since the program years in our data set are often not contiguous, we use only those observations that had a program forecast for the preceding year. There are 57 such observations. Table 1 reports

significant serial correlation for the forecast errors in both the forecast of taxes as a ratio to GDP (e_t^r), and the forecast of change in nominal tax revenue (e_t^l), for this subset of our data.

B. Bias

The standard test of bias is to regress the forecast errors on a constant, and check whether the constant is significantly different from zero. The exact distribution of this test in finite samples is known only for the case of Gaussian errors, and the discussion above implies that this assumption is violated—at least in the case of the forecasts of change in nominal tax revenues. As a complementary test of bias, we follow Campbell and Ghysels (1997) and use the Wilcoxon signed rank test to test the hypothesis that the median is different from zero. This is a nonparametric test which does not require the assumption of the normality.

Both tests reject the hypothesis that the forecast of tax revenues as a percentage of GDP is unbiased, at a very high level of significance. On average, the forecasted tax revenue ratios are higher by about 1 percentage point of GDP than the corresponding actual outcomes. The forecast of the percentage change in nominal revenues, on the other hand, appears to be unbiased: we cannot reject the hypothesis that the mean or median are different from zero at any reasonable level of significance.

The fact that forecasts of the ratio of tax to GDP is biased upwards, while there is no bias in the forecasts of growth in nominal tax revenues, implies an association between the errors in the tax ratio forecasts and errors in the forecasts of nominal GDP. To investigate this association further, we run the following regression:

$$e_t^r = \alpha + \beta e_t^{GDP} + \varepsilon_t$$

where e_t^{GDP} is the percentage forecast error for nominal GDP. The results shown in Table 2 confirm that over (or under) prediction of the tax-to-GDP ratio was associated with under (or over) prediction of nominal GDP. A further regression shows that the forecast of nominal GDP has significant negative bias, so that actual GDP exceeds forecasted GDP by more than 7.5 percent of actual GDP in the program year in our data set.

A variety of macroeconomic mechanisms could account for this association. For example, in the presence of tax collection lags, inflation that is higher than expected—for whatever reason—will tend to depress tax revenues in real terms (Tanzi, 1977), so that the ratio of those revenues to GDP will also be below the expected level. Alternatively, a shortfall in tax revenues will, other things being equal, increase the fiscal deficit, which may have monetary or exchange rate consequences that are reflected in higher inflation.

Table 2. Regression of Errors in Tax Ratio Forecasts on Errors in GDP Forecast

	Value
a	0.72** (0.23)
β	-0.04** (0.01)
R^2	0.072

Standard errors in parenthesis. * and ** denote significance at 90% and 95% confidence level.

C. Accuracy

The accuracy of the program forecasts in the sample period appears quite low. The mean absolute forecast error was around 1.86 percent of GDP for the forecasts of the tax ratio, and 16.8 percent for the forecasts of percentage changes in nominal tax revenues. The mean absolute percentage error (MAPE), which is the statistic most commonly used in comparing tax revenue forecasts in different countries, was 16.0 percent.¹⁸ This figure is much higher than the MAPEs found in studies in OECD countries and the U.S. and Canadian states, as cited in Section I above. This is not surprising, since the economies of ESAF/PRGF countries are likely to be more volatile, which makes any economic series much more difficult to predict. In addition, economic data are in general much less reliable, and statistical and forecasting resources more limited, in these countries.

The accuracy of the forecasts can be judged only in relation to some other forecast in the same environment. We consider the performance of the program forecasts relative to a “naive” forecast. First, we use Theil’s test, which is equal to the ratio of the mean square error (MSE) of the actual forecasts to the MSE of a naive forecast based on the hypothesis that the forecast variable will remain the same in the next period as in the current period.

In the case of the forecasts of tax revenues as a ratio to GDP, this test ratio is 1.07, indicating that the naive forecast actually outperforms the program forecasts, by 7 percent. The difference is, however, not significant.¹⁹

¹⁸ Note that in our measure of error in forecasts of growth in nominal tax revenues, e_t^l , the error is the difference between forecast and outturn, both expressed as percentages of the outturn in the previous year (when the forecast was made). By contrast, in the MAPE measure defined in Section I, the errors are expressed as percentages of the outturn in the year for which the forecasts were made. Hence, the measures are likely to differ slightly.

¹⁹ To assess significance we use the Diebold-Mariano test, as modified by Harvey, Leybourne, and Newbold (1997). The same test was employed in a study by Artis and Marcellino (2001) of published IMF forecasts of budget deficits for a sample of OECD countries from 1976 to 1995.

It is difficult to specify an appropriate naive comparator for the forecasts of change in nominal tax revenues. The most obvious would be an assumption that these revenues grow at the same rate as nominal GDP. This is, however, equivalent to assuming no change in the ratio of tax to GDP, so the result would be exactly the same as that reported above for the forecasts of tax revenues as a ratio to GDP. A possible alternative naive assumption that nominal tax revenues do not grow at all, would not seem to provide a very useful standard of comparison.²⁰

Another measure of the accuracy of forecasts presented by Theil (1966) is the U statistic, which is computed as:

$$U = \frac{\sqrt{N^{-1} \sum e_t^2}}{\sqrt{N^{-1} \sum A_t^2}}$$

This measure allows comparisons to be made of forecasts of different series, taking account of the difference in their variability. Following the discussion in Clements and Hendry (1998), the statistic can be further decomposed into bias (UM), variance (US), and covariance (UC) proportions:

$$\begin{aligned} UM &= \frac{(\bar{P} - \bar{A})^2}{N^{-1} \sum e_t^2}, \\ US &= \frac{(S_p - rS_a)^2}{N^{-1} \sum e_t^2}, \\ UC &= \frac{(1 - r^2)S_a^2}{N^{-1} \sum e_t^2}, \end{aligned}$$

where bars denote means of the series, S_p and S_a are standard deviations of the predicted and actual series respectively, and r is the correlation between P_t and A_t . The three proportions sum to one, since

$$N^{-1} \sum e_t^2 = (\bar{P} - \bar{A})^2 + (S_p - rS_a)^2 + (1 - r^2)S_a^2$$

The forecast that minimizes MSE will have the first two components equal to zero, with all the weight being concentrated on UC . A high value for UM would indicate that bias is responsible for a large part of MSE. Table 3 contains decomposition statistics for the program forecasts together with the statistics for a naive forecast.

²⁰ Theil's MSE ratio for this naive comparison is 0.79, indicating that the actual forecasts of percentage change in nominal tax revenue do indeed outperform naive forecasts that those revenues would remain constant in nominal terms.

Table 3. Theil's U Statistics

		U	UM	US	UC
Forecasts of ratio of tax to GDP	Program forecasts	0.170	0.163	0.066	0.770
	Naive alternative (no change in tax ratio)	0.159	0.045	0.153	0.817
Forecasts of percentage change in nominal tax revenue	Program forecasts	0.795	0.003	0.083	0.913
	Naive alternative (no change in nominal tax revenue)	1.000	0.383	0.000	0.617

By these measures, the program forecasts appear to do reasonably well. Most of the errors are unsystematic, coming from a random component. Although we have established that there is a significant bias in the forecast of tax ratios, this bias accounts for only 16 percent of the MSE.

Finally, we assess the directional accuracy of the forecasts of tax ratios. We test whether, on average, the program forecast correctly predicts the direction of change of the tax ratios. In Table 4 below, "Forecast 'up'" is the number of observations for which the program predicted an increase in tax ratio in the next period compared to the estimated tax ratio in the current period. The other rows and columns are defined analogously.

Table 4. Directional Accuracy of the Tax Ratio Forecasts

	Actual "up"	Actual "down"
Forecast "up"	44	59
Forecast "down"	12	11

Thus, in most cases the program predicted an increase in the tax to GDP ratio (in 103 out of 126 observations), while in practice tax ratios mostly decreased (in 70 cases). A simple chi-squared test cannot reject the hypothesis that the forecasts and realizations are

independent.²¹ Combining this result with Theil's measure of forecast accuracy, we conclude that the accuracy of the program forecasts is indeed very low: they do not outperform even the most simple naive forecasts.

V. EXPLORING THE SOURCES OF BIAS

In this section, we investigate possible sources of the forecast errors. We look for evidence on whether some common explanations of bias are supported by the data.

Table 5 summarizes the results of a series of regressions in which the dependent variable was the error (a) in the forecasts of ratios of tax revenues to GDP, and (b) in the forecasts of percentage changes in nominal tax revenues. These errors are regressed on a series of possible explanatory variables. These variables are described more fully below; a brief description is provided in Table 6, together with summary statistics.

As shown in the Table 5, for each measure of forecast error, our procedure was to start by regressing those errors on all the possible explanatory variables that we are interested in; then, to drop the one with the lowest significance; and so on—until all the remaining variables are significant at a level of at least 10 percent. Columns 4 and 7 of Table 5 contain these final regressions.

A. Success of the Program

As discussed in Section II above, a possible reason for ex post bias in the tax forecasts in IMF programs is that these numbers are conditional forecasts, representing the tax revenues expected if the country concerned follows all the conditions of the program. We investigate whether this explanation holds in the data, using various proxies for the success of the program.²²

One possible indicator of the success of the program is whether it was implemented in the time schedule originally intended. The dummy *INTERRUPTION* takes a value of one for programs which were discontinued, or delayed for more than six months.²³ Our use of this dummy follows Musso and Phillips (2002) and is based on the idea that the final review of the program is completed only when the progress of the program is deemed satisfactory.

It is conceivable that a program could be interrupted merely because negative exogenous shocks led to a failure of the program. In such cases, the dummy *INTERRUPTION* would reflect not

²¹ The use of this simple chi-squared test in evaluating direction-of-change forecasts is described in Diebold and Lopez (1996), p. 256–57. The test uses only the qualitative information contained in Table 4.

²² For a description of these proxies see Ivanova et al. (2003). We are grateful to Alex Mourmouras for providing us with the data used in this section.

²³ Almost all of the programs were subject to some interruption, but in general this was for periods shorter than six months.

Table 5. Regressions to Account for the Forecast Errors

	(A) Error in forecast of tax as percentage of GDP: e_t^r			(B) Error in forecast of percentage growth in nominal tax revenue: e_t^l		
Intercept	-0.05 (0.95)	0.00 (0.94)	-0.46 (0.55)	-11.74 (8.55)	-11.78 (8.60)	-13.95** (5.01)
INTERRUPTION	0.86* (0.51)	0.88* (0.51)	0.97* (0.49)	12.04** (4.60)	12.63** (4.65)	12.83** (4.50)
MONTHS	0.16** (0.07)	0.16** (0.07)	0.17** (0.07)	1.48** (0.63)	1.60** (0.64)	1.56** (0.61)
AFRICA	-0.28 (0.51)	-0.27 (0.51)		-0.16 (4.56)	-0.18 (4.62)	
TAXCONDIT	-0.40 (0.54)	-0.42 (0.54)		2.57 (4.95)	2.45 (5.01)	
MACRO	0.69 (1.17)			16.89 (10.62)		
STRUCTURAL	-0.76 (1.26)			-22.20* (11.43)		
OVERALL		-0.10 (0.84)			-4.28 (7.60)	
R ²	0.10	0.09	0.08	0.15	0.12	0.12
Prob(F-statistic)	0.111	0.076	0.010	0.010	0.020	0.002

Standard errors in parenthesis. * and ** denote significance at 90% and 95% confidence level.

Table 6. Explanatory Variables: Descriptive Statistics

Variable		Mean	Std. dev.	Min.	Max.
INTERRUPTION	Dummy for program interruption	0.37	0.48	0	1
MONTHS	Months before forecast period	6.64	3.46	-1	16
AFRICA	Dummy for Sub-Saharan Africa	0.59	0.49	0	1
TAXCONDIT	Dummy for tax conditionality	0.33	0.47	0	1
MACRO	Index of macro conditions achieved	0.74	0.37	0	1
STRUCTURAL	Index of structural conditions achieved	0.67	0.31	0	1
OVERALL	Combined index of conditions achieved	0.68	0.34	0	1

only deviations from the economic program by the country, but also such shocks as military conflicts, civil wars, droughts, etc.

As a more direct way of measuring the compliance of the program country with the conditions of the program, we used three indices: *MACRO*, *STRUCTURAL*, and *OVERALL*. These indices take a value between 0 (indicating no compliance at all) and 1 (indicating full compliance). The indices are calculated as the average number of program performance criteria that were met (weight 1), partially met (weight 0.5) or not met at all (weight 0) during the program period.

In Table 7, we investigate whether these variables can account for bias in the tax ratio forecasts. The evidence that programs with better compliance will have smaller bias is very weak. All three indices have the wrong sign, and none is significant. Only programs that were interrupted exhibit higher bias, but even that effect is not detectable at the 90 percent confidence level. The intercept is significantly different from zero, which suggests that bias is present even in the programs that were not interrupted.

Table 7. Tax Ratio Forecast Errors: Effects of Measures of Program "Success"

Test	Intercept	Dummy for interruption	Macro index	Structural index	Overall index	R ²	Prob (F-statistic)
1	0.75** (0.30)	0.82 (0.50)				0.024	0.105
2	0.59 (0.42)		0.58 (0.51)			0.011	0.261
3	0.79* (0.41)			0.32 (0.54)		0.003	0.560
4	0.71* (0.42)				0.44 (0.55)	0.006	0.421

Standard errors in parenthesis. * and ** denote significance at 90% and 95% confidence level.

From columns 4 and 7 in Table 5 it can be seen that, in a more general regression with other factors being controlled for, similar results still hold. In general, the compliance indices have no significant impact on the error term for either the tax ratio forecasts or the forecasts of nominal tax changes. The dummy *INTERRUPTION* has, however, significant and positive coefficient in both cases. This result is consistent with the hypothesis that an ex post bias results from the conditional nature of the forecasts, but it does not fully account for the bias: other variables that are not directly related to the program conditionality are also significant.

B. Conditionality Related to Taxes

Since we are interested in the tax revenue forecast, it is of interest to determine the properties of the forecasts for those programs which set explicit performance targets related to taxes. These targets take various forms, from an explicit target for tax revenue collections, to the adoption of particular structural or administrative reforms. In our sample, there were 40 program years for which such targets were set. The forecasts of tax revenues in this subsample might be expected to have different properties from the rest of the programs. For example, when there is a condition that a program country should have a certain level of tax revenues in the program year, that country may undertake additional measures to raise those revenues to meet program criteria. Then, ex post, we could observe negative bias because, when the forecast turns out to be too low, the country will not be making any fiscal changes.

The forecast of tax revenues as a percentage of GDP shows a smaller bias in this subsample, but it is still significant: the mean forecast error is 0.58 and its standard error is 0.27. The forecast of growth in nominal tax revenues remains unbiased, with the mean equal to -0.14 and standard error 3.09. This result is not robust in the whole sample, however. The dummy *TAXCONDIT*, which takes the value 1 if the program had a tax-related condition, is not significant in regressions, as may be seen from Table 5.

C. Geographical Differences

Over half of our sample consists of sub-Saharan African countries, which might have different structural characteristics from those of low-income countries in Asia, Latin America, or transition economies. We use a dummy *AFRICA* to test for possible regional effects on the forecast error. There is, however, no evidence that forecast errors are different for African countries: this dummy is insignificant, and it was dropped from the final results.

D. Forecast Horizon

IMF programs commence at any time during the year. Hence, the program forecasts for tax revenues in the next calendar or fiscal year are agreed anything up to 12 (or more) months before the beginning of the forecast period. In principle, the mean error of an efficient forecast should not depend on the length of this lag: only the variance of the errors would be expected to be higher, the longer is the horizon.²⁴ However, the forecasts exhibit significantly higher bias at the longer horizons. We construct a variable *MONTHS* which is defined as the number of months from the start of the program to the beginning of the year (fiscal or calendar) for which the forecast was made. Thus, in the case of a program which starts in June, and in which forecasts refer to the calendar year, the variable *MONTHS* will have a value of 6. The results indicate that bias is larger for programs with a longer forecast horizon: the bias for the tax ratio forecast is increased by about 0.17 percentage points of GDP for each additional month of the horizon.

²⁴ The different forecast horizons could potentially lead to heteroskedasticity of the errors and require additional correction of the standard errors. In practice, that does not appear to be the case as White's tests do not detect heteroskedasticity in any of the regressions.

Finally, it may be seen from Table 5 that the two variables *MONTHS* and *INTERRUPTION*, in combination, fully account for the bias in the forecasts of the ratio of tax revenue to GDP: the constant in the estimated equation is not significantly different from zero. They account, however, for only 8 percent of the variance in the forecast errors.

VI. CONCLUSION

This paper examined the accuracy of forecasts of total tax revenues prepared in the context of IMF programs supported by the ESAF in the years 1993–99. The focus was on the accuracy of these forecasts, on whether they display any ex post bias, and on what factors could account for any such bias. Two tax forecast measures were analyzed: (a) forecasts of tax revenues as a percentage of GDP; and (b) forecasts of percentage changes in nominal tax revenues.

The overall accuracy of these forecasts is low. The mean absolute errors are 1.86 percentage points for tax revenues as a percentage of GDP, and 16.8 percent for percentage changes in nominal tax revenues. The root mean square error (RMSE) of the forecast of tax revenues as a percentage of GDP is actually higher (though not significantly so) than the RMSE of a naive “no change” forecast. Perhaps most strikingly, the forecast direction of change in tax revenues as a percentage of GDP is unrelated to the direction of the change that actually occurred.

There is statistically significant upward bias in the forecasts of tax revenues as a percentage of GDP, but no evidence of similar bias in the forecasts of changes in nominal tax revenues. These contrasting findings reflect significant positive correlation between the forecast errors in tax revenues as a percentage of GDP, and errors in the program forecasts of nominal GDP: these two errors tend to offset each other in their effects on the forecast of nominal tax revenues.

The size of the upward bias is positively correlated with the length of time between the forecast, and the beginning of the year for which that forecast was made. It is also higher in the case of programs that were discontinued, or where implementation was delayed by more than six months. On the other hand, we have found no evidence that the bias is related to the country’s compliance with the program conditionalities, and no evidence of regional differences.

The object of this paper was to examine the facts about tax revenue forecasts in IMF-supported programs in low-income countries, rather than to explore any implications those facts may have for policy, or for the construction of such programs. The findings could, however, have important implications in two particular areas.

The first of these concerns forecast bias. To the extent that a significant upward bias is associated with characteristics of the program that are known when the forecast is made, it should be possible to remove that bias by systematic adjustments to the forecast. Whether or not the program will in fact be implemented within the time originally scheduled is, of course, not known at the time of the forecast; and we have suggested that program interruptions are most appropriately seen as a source of ex post rather than ex ante bias. But the time lag between the forecast and the start of the forecast year is known: in principle, therefore, an adjustment could be made that would remove the associated bias.

The second area concerns accuracy. Even if the two sources of bias that we have identified could have been removed, the variance of the errors in the forecasts of the tax ratio would have been reduced by only 8 percent. The remaining, unexplained errors in the tax revenue forecasts are very large. The appropriate use of year-ahead revenue forecasts will be rather different in an environment where their mean absolute error is 16 percent, rather than 2 or 3 percent. It is reasonable to suppose that, in many cases, these forecasts could be improved upon—for example, by the use of better forecasting techniques, better organization, better trained personnel, or better data. But the major part of the observed errors seems likely to reflect the underlying forecasting difficulties arising from the economic environment and structure of low income countries, aggravated by the conditions of financial stress in which IMF programs are formulated.

REFERENCES

- Abed, George T. and others, 1998, *Fiscal Reforms in Low-Income Countries: Experience Under IMF-Supported Programs*, IMF Occasional Paper No. 160 (Washington: International Monetary Fund).
- Artis, Michael, and M. Marcellino, 2001, "Fiscal Forecasting: The Track Record of the IMF, OECD and EC," *Econometrics Journal*, Vol. 4, pp. S20–S36.
- Auerbach, Alan J., 1999, "On the Performance and Use of Government Revenue Forecasts," *National Tax Journal*, Vol. 52, No. 4, pp. 767–82.
- Campbell, Bryan and Eric Ghysels, 1997, "An Empirical Analysis of the Canadian Budget Process," *Canadian Journal of Economics*, Vol. 30, Issue 3, pp. 553–76.
- Clements, Michael P., and David F. Hendry, 1998, *Forecasting Economic Time Series* (Cambridge, England: Cambridge University Press).
- Congressional Budget Office, 2001, *Description of CBO's Models and Methods for Projecting Federal Revenues*, Congress of the United States CBO Paper, May.
- Diebold, Francis X., and Jose A. Lopez, 1996, "Forecast Evaluation and Combination" in *Handbook of Statistics*, Vol. 14, ed. by G.S. Maddala and C.R. Rao (Amsterdam: Elsevier).
- Ernst & Young, 1994, *Review of the Forecasting Accuracy and Methods of the Department of Finance: Final Report* (Ottawa).
- Harvey, David, Stephen Leybourne, and Paul Newbold, 1997, "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, Vol. 13, pp. 281–91.
- International Monetary Fund, 2001, *Revised Manual of Fiscal Transparency* (Washington).
- Ireland, Department of Finance, 2000, *Report of the Tax Forecasting Methodology Group*. Available via Internet: <http://www.irlgov.ie/finance/taxfor.htm>
- Israel, Ministry of Finance, 1996, *Annual State Revenue Report for 1996. Chapter I: State Revenue Forecast for 1997*, Israeli Ministry of Finance, Economic Research and State Revenue Division. Available via Internet: <http://www.mof.gov.il/beinle/state/chap1a.htm>
- Ivanova, Anna, Wolfgang Mayer, Alex Mourmouras, and George Anayiotos, 2003, "What Determines the Success or Failure of Fund-Supported Programs?" Forthcoming *IMF Working Paper* (Washington: International Monetary Fund).
- Jenness, R., and Arabackyj, S., 1998, "Budget Forecasting Records of the Federal and Provincial Governments," *Monthly Economic Review*, Vol. 17, No. 1, pp. 1–26.

- Loungani, Prakash, 2001, "How Accurate Are Private Sector Forecasts? Cross-Country Evidence from Consensus Forecasts of Output Growth," *International Journal of Forecasting*, Vol. 17, No. 3, pp. 419–32.
- Mocan, H.N., and Azad, S., 1995, "Accuracy and Rationality of State General Fund Revenue Forecasts: Evidence from Panel Data," *International Journal of Forecasting*, Vol. 11, pp. 417–27.
- Musso, Alberto, and Steven Phillips, 2002, "Comparing Projections and Outcomes of IMF-Supported Programs," *Staff Papers*, International Monetary Fund, Vol. 49 (March), pp. 22–48.
- Penner, Rudolph G., 2001, *Errors in Budget Forecasting* (Washington: The Urban Institute).
- Pike, T., and D. Savage, 1998, "Forecasting the Public Finances in the Treasury," *Fiscal Studies*, Vol. 19, No. 1, pp. 49–62.
- Sunley, Emil M., and R.D. Weiss, 1991, "The Revenue Estimating Process," *Tax Notes*, June.
- Tanzi, Vito, 1977, "Inflation, Lags in Collection, and the Real Value of Tax Revenue," *Staff Papers*, International Monetary Fund, Vol. 24 (March), pp. 154–67.
- Theil, Henri, 1966, *Applied Economic Forecasting* (Amsterdam: North-Holland).
- United Kingdom, HM Treasury, 1997, *The VAT Shortfall: Report of the Working Group on VAT Receipts and Forecasts*, Treasury Occasional Paper No. 9 (London).